DNA polymerase fidelity and the polymerase chain reaction.

K A Eckert and T A Kunkel

References	This article cites 45 articles, 32 of which can be accessed free at: http://genome.cshlp.org/content/1/1/17.refs.html
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

DNA Polymerase Fidelity and the Polymerase Chain Reaction

Kristin A. Eckert and Thomas A. Kunkel

Laboratory of Molecular Genetics, National Institute of Environmental Health
Sciences, Research Triangle Park, NC 27709

The distinctive ability of the polymerase chain reaction (PCR) to produce a substantial quantity of DNA from an initially small amount of genetic material is revolutionizing molecular biology. The variety of highly successful applications of PCR technology and the ease with which PCR can be performed should not be misinterpreted as evidence that the PCR is a simple process. In fact, quite the opposite is true. A single cycle of DNA polymerization is a complex process requiring the precise interaction of several components. The repetitive cycles characteristic of PCR provide an additional layer of complexity to the technique. The final PCR product cannot be considered a unique entity, as even a discrete DNA band on an agarose gel may contain a variety of DNA molecules differing from the original genetic information by one or more nucleotides.

How serious a concern is the fidelity of PCR? The answer depends on the precise nature of the application. In many instances, for example, direct characterization of the amplified population by DNA sequence analysis or nucleic acid hybridization, random errors in nucleotide sequence that may be produced during PCR are of little concern. However, some PCR applications involve the characterization of individual DNA molecules or rare molecules present in a heterogeneous population. Examples include the study of allelic polymorphism in individual mRNA transcripts[1,2] and the characterization of the allelic states of single sperm cells[3] or single DNA molecules.[4,5] In these circumstances, the fidelity (error rate per nucleotide) of PCR is an important consideration, because errors generated during amplification may interfere with the interpretation of data.

During enzymatic DNA amplification, the majority of changes in nucleotide sequence can be attributed to errors made by the DNA polymerase. While PCR is a recent development in the field of molecular biology, the fidelity of DNA polymerases has been studied biochemically for almost 30 years. In this review, we will attempt to relate our current understanding of the mechanisms of polymerase fidelity in vitro to the accuracy of enzymatic DNA amplification. We begin by describing steps in polymerase error discrimination and variables important for determining fidelity. The error rates of several DNA polymerases that have been used for PCR will be described, first as determined in model "one-cycle" fidelity assays and then during actual PCR. Our intent is to focus on the variables known to influence the fidelity of DNA synthesis in vitro that can be controlled experimentally.

ERROR ACCUMULATION DURING PCR

Error rates in PCR vary according to the precise DNA sequence and the in vitro conditions of DNA synthesis. Several laboratories have estimated the error frequency (mutations per nucleotide per cycle) during PCR catalyzed by the thermostable *Thermus aquaticus* (*Taq*) DNA polymerase by cloning individual DNA molecules from the amplified population and determining the number of DNA sequence changes.[6-12] As seen from the data in Table 1, observed error frequencies during PCR vary more than 10-fold, from ~2 x 10^{-4} to <1 x 10^{-5}. For applications of PCR products that utilize the population of amplified DNA molecules, random errors at these frequencies will not interfere with the biochemical analyses. However, polymerase-mediated errors at a frequency of 1 mutation per 10,000 nucleotides

TABLE 1 Estimates of Fidelity during PCR Catalyzed by *Taq* Polymerase

Target sequence	Number of cycles	Error frequency per nucleotide	Error rate per nucleotide per cycle[a]	Reference
Apolipoprotein B	30	22/8,000	1/5,600	6
HLA-DPβ	30	17/6,692	1/5,900	7
α-Antitrypsin	33	7/4,700	1/11,000	8
HPRT	32	16/15,000	1/15,000	9
HLA-A,B	30	20/21,870	1/16,000	10
	20	17/30,710	1/18,000	10
TCR Vδ	25	1/1,500	1/19,000	11
HIV *gag/env*	30	<1/5,411	<1/83,000	12

[a]Error rate = 1/[(observed error frequency/number of cycles) x 2], assuming a constant efficiency per cycle.[7]

Review███████ █ █ █ ███

per cycle are an important consideration for any PCR application that begins with a small amount of starting material (i.e., less than a total of 10,000 nucleotides of target DNA) or that focuses on individual DNA molecules present in the final PCR population.

The proportion of DNA molecules that contain sequence changes is a function of the error rate per nucleotide per cycle, the number of amplification cycles, and the starting population size.[13,14] The population of altered DNA molecules arises during PCR from two sources: (1) new errors at each cycle and (2) the amplification of DNA molecules containing errors from previous cycles. On average, the same number of DNA sequence changes in the final population results from polymerase errors during the last amplification cycle as result from the amplification of errors that occurred during previous cycles. The formula $f = np/2$ describes the average mutation frequency (f) for PCR amplification as a function of the polymerase error rate per nucleotide per cycle (p) and the number of cycles (n), assuming that p is constant at each cycle (for more detailed mathematical analyses, see refs. 13 and 14). Theoretically, the frequency of DNA sequence changes can be controlled by altering the number of cycles (n) and/or the polymerase error rate per nucleotide (p) (see Table 6 in ref. 13). For example, as the number of cycles increases, the error frequency is expected to increase. Keeping p constant at one error per 10,000 nucleotides per cycle, the error frequency when n = 20 cycles is one change per 1000 nucleotides and increases to one change per 400 nucleotides after 50 amplification cycles. Alternatively, the final error frequency can be varied by changing p, the polymerase error rate. Keeping a constant cycle number of n = 50, the final mutation frequency is only 1/2000 nucleotides if p = 1/50,000, but increases to 1/400 nucleotides if p = 1/10,000. Due to the exponential nature of PCR, the occurrence of an early error can increase the final error frequency above the average described by $f = np/2$, because the mutant DNA molecules will be amplified with each cycle, resulting in populations with a larger than average number of mutants.[15] The remainder of this review explores the ways in which the polymerase error rate, p, can be manipulated experimentally as a means of controlling the PCR error frequency.

MOLECULAR NATURE OF MUTATIONS

A variety of changes in DNA sequence can occur during enzymatic DNA amplification. These mutations vary from single base-substitution mutations to deletions and insertions.[6-12] The most familiar type of mutation is the base substitution that results from the misincorporation of an incorrect dNTP during DNA synthesis. Twelve distinct base mispairs are possible, varying in composition and in the template base versus the incoming dNTP (e.g., T·dGTP versus G·dTTP). Together with the effects of neighboring template sequence, this creates a variety of possible molecular structures among which any polymerase must discriminate. All DNA polymerases examined to date also generate mutations during DNA polymerization in vitro wherein one or more nucleotides is lost or added. One factor that will influence the types of DNA sequence changes (base substitution versus deletion mutation) likely to occur during enzymatic amplification is the sequence of the target DNA.

The frequency of deletion mutations is sequence dependent and is increased in repetitive DNA sequences (for reviews, see refs. 16 and 17) Many examples exist for the repetitive nature of DNA, from simple reiteration of a single base, to alternating dinucleotide tracts in potential Z-DNA structure, to short direct repeats. The generation of deletion mutations is likely to be determined by the stability of the mutational intermediates.[16,17] The most common polymerase-mediated deletion error is the loss of one nucleotide. As first proposed by Streisinger,[18] one mechanism for the gain or loss of a single nucleotide is the misalignment within a repetitive homopolymeric sequence of the DNA template-primer during synthesis. Consistent with this mechanism, the error rates for some minus-one-base deletion mutations produced by purified polymerases in vitro increase as the length of the repeated sequence increases.[16] Alternative models have also been proposed, including deletions initiated by base misinsertion.[19] Substantial evidence from model assays demonstrates that polymerases also can generate deletions of a few to hundreds of nucleotides during polymerization in vitro. These deletions often involve directly repeated sequences and palindromic or quasi-palindromic sequences that have the potential to form stem-and-loop DNA structures.[17]

In addition to intramolecular errors during DNA synthesis, deletions and other types of rearrangements can occur between DNA molecules. Such intermolecular events are thought to arise when the polymerase terminates synthesis on one DNA strand and continues synthesis after priming occurs on a complementary DNA strand (strand switching or jumping PCR).[20] Both the presence of a high degree of sequence homology between regions of DNA molecules and a high DNA concentration increase the incidence of this type of error in PCR.[10,20]

ERROR DISCRIMINATION BY DNA POLYMERASES AND FIDELITY IN MODEL SYSTEMS

Since the majority of DNA sequence changes introduced during PCR may be polymerase-mediated, we will review what is known about DNA polymerase error rates from model in vitro systems. Much of our understanding of fidelity is derived from assays that quantitate errors made by DNA polymerases during synthesis equivalent to a single PCR cycle.[21-24] Among these, the broadest description of fidelity comes from the M13mp2 forward mutation assay,[24] which detects all possible base substitution errors along with a variety of base deletion and addition errors at a large number of sites. In this assay, single-stranded M13mp2 DNA containing the α-complementation region of the *Escherichia coli lacZ* gene is used as the template for a single cycle of DNA synthesis. Upon transfection of an appropriate *E. coli* strain with the products of the reaction, accurate DNA synthesis can be detected as dark blue M13mp2 plaques. However, polymer-

ase errors during synthesis of the *lacZ* gene result in lighter blue or colorless plaques, due to decreased α-complementation of β-galactosidase activity in the infected host cell. The error rates for specific classes of polymerase errors during a single synthesis cycle can be calculated from the frequency of light blue and colorless (mutant) plaques among the total plaques scored and DNA sequence analysis of a collection of the mutants. A more sensitive assay for specific types of mutations can be performed by using an M13mp2 template DNA which contains a base substitution or deletion mutation, resulting in a colorless phenotype. In these assays, polymerase errors are scored as DNA sequence changes that revert the mutant back to a wild-type or pseudo–wild-type phenotype. This approach is especially useful for highly accurate polymerases. Unlike the forward mutation assay, however, M13mp2 reversion assays are focused on a limited subset of errors occurring at only a few sites.

Polymerase-mediated discrimination against errors during synthesis occurs at several steps in the biochemical reaction (Fig. 1) (for review, see ref. 25). The first control point is the binding of the dNTP substrate to the polymerase (step A). The rate of binding of both the correct and incorrect incoming deoxynucleoside triphosphate is assumed to be the same. However, the correct dNTP substrate may remain preferentially bound to the enzyme-DNA complex due to hydrogen bonding and base-stacking interactions. Alternatively, an incorrect dNTP may dissociate from the E·DNA·dNTP complex more readily than a correct dNTP.[21,22] At this stage, polymerase error rates can be varied experimentally by adjusting the ratio of correct and incorrect dNTP substrates, reflecting the relative probability that a polymerase will bind an incorrect versus a correct dNTP. An example of such a reaction catalyzed by *Taq* polymerase is shown in Table 2 (part I), wherein base substitution fidelity is decreased eightfold by decreasing the concentration of a single correct nucleotide. Imbalances in the dNTP pools can be either mutagenic or antimutagenic, depending on the error being considered. A second point for

error discrimination is the more rapid rate of phosphodiester bond formation for correct rather than for incorrect E·DNA·dNTP complexes (step B). Observations from several laboratories suggest that DNA polymerases use selectivity at both steps A and B to improve base substitution fidelity, but that the relative contribution of the two steps varies from one polymerase to the next.[26–28]

Once the phosphodiester bond is formed between the incoming dNTP and the DNA primer, the rate of pyrophosphate release (step C) may occur more slowly in the case of incorrect incorporations.[21] This third level of discrimination, together with the observation that some DNA polymerases catalyze pyrophosphorolysis and pyrophosphate exchange reactions,[29] provides the possibility that the presence of pyrophosphate in a polymerization reaction may influence fidelity. Both mutagenic[30] and antimutagenic[31] pyrophosphate effects on the fidelity of *E. coli* DNA polymerase I have been observed.

A fourth level of error discrimination is the selective ability of a polymerase to continue DNA synthesis on a correct primer–template structure rather than on a mutational intermedi-

ate (step D). The probability of preserving any mispaired intermediate as a mutation is increased by reaction conditions that increase the rate of extension synthesis, such as high dNTP concentrations, high enzyme concentrations, or long incubation times. The rate of mispair extension varies for both the type of mispair and the polymerase.[32,33] In general, purine–purine mispairs are the most difficult to extend (10^{-5} to 10^{-6}), while purine–pyrimidine and pyrimidine–pyrimidine mispairs are extended more efficiently (10^{-2} to 10^{-4}) as compared to the corresponding correct base pairs (M. Goodman, pers. comm.). For the *Taq* polymerase, the failure to extend mispairs is not due to differential binding of the polymerase to correct versus incorrect terminal base pairs, but rather is due to a kinetic block to the addition of the next nucleotide onto the mispair (M. Goodman, pers. comm.). As for other DNA polymerases, the intrinsic extension efficiency for *Taq* polymerase measured under steady-state conditions differs for each type of mispair (M. Goodman, pers. comm.).

Finally, one of the most important control points for the production of mutations is the $3' \rightarrow 5'$ exonucleolytic removal (proofreading) of $3'$ ter-
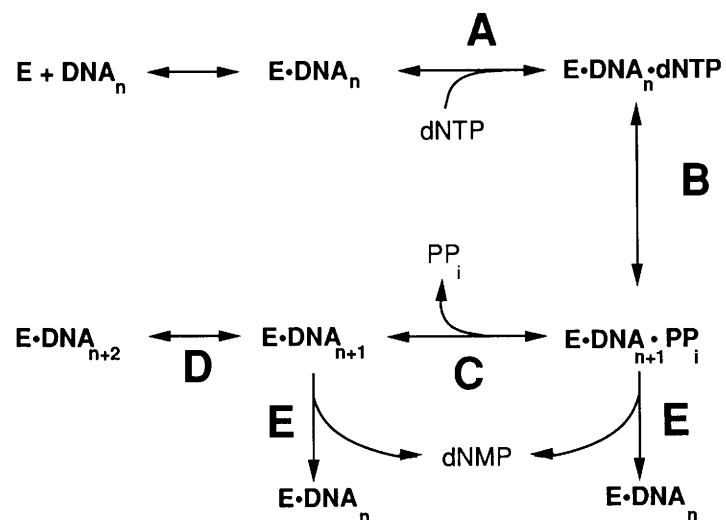

FIGURE 1 Steps in polymerase error discrimination. This reaction scheme for error discrimination by DNA polymerases in vitro has been generalized from data for a variety of enzymes,[25] but relies heavily on data generated using the Klenow fragment of *E. coli* DNA polymerase I.[21] (E) Polymerase; (DNA_n) starting DNA primer–template of length *n*; (DNA_{n+1}) DNA_n primer–template after incorporation of one nucleotide; (DNA_{n+2}) DNA_{n+1} primer–template after extension synthesis; (dNTP) deoxynucleoside triphosphate; (PP_i) pyrophosphate; (dNMP) deoxynucleoside monophosphate. Letters A–E denote control points for error discrimination (see text for details).

*Review*IIII■■ I ■ I ■

TABLE 2 Comparative DNA Polymerase Fidelity in the M13mp2 Mutation Assays

Polymerase	$3' \rightarrow 5'$ Exonuclease	Base substitution error rate per nucleotide[a]	Reference
I. *Taq*	–	$1/2,400^b$	13
Taq (100 μM dNTP)	–	$1/290^b$	13
II. Klenow (1 μM dATP)	+	1/110,000	35
Klenow (1 μM dNTP)	–	1/15,000	35
III. T4	+	1/300,000	T.A.K. (unpubl.)
		$1,80,000^c$	13
Native T7	+	1/290,000	T.A.K. (unpubl.)
		$1/53,000^c$	13
Klenow	+	1/45,000	50
		$1/27,000^c$	13
T. litoralis	+	1/31,000	P. Mattila (pers. comm.)
IV. AMV (500 μM dNTPs)	–	1/24,000	51
Taq	–	1/5,600	13
Sequenase 2.0	–	$1/5,000^c$	T.A.K. (unpubl.)
V. *Taq* (4 mM Mg^{2+})	–	1/300,000	37
Taq (pH 5.7)	–	1/180,000	37

All fidelity measurements were done using 1 mM each of four dNTPs and 10 mM $MgCl_2$ unless otherwise indicated by parenthetical notations. The T4, native T7, Sequenase 2.0, AMV, and Klenow polymerase reactions were incubated at 37°C; the *Taq* polymerase reactions at 70°C, and the *T. litoralis* polymerase reactions at 72°C.

[a]Except where noted, error rates per detectable nucleotide were determined from the mutant frequency observed in an opal codon reversion assay as previously described.[52]

[b]The base substitution error rate was calculated for changes at the first position (T·G) of the TGA codon only. As determined by DNA sequence analysis, first position changes account for 85% of the total errors produced by *Taq* polymerase at equal dNTP concentrations.

[c]The base substitution error rate was calculated as described[50] from DNA sequence analysis of mutants produced in the M13mp2 forward mutation assay.

minal mispairs to regenerate the starting DNA configuration (step E) (for review, see ref. 34). Slow pyrophosphate release after an incorrect incorporation (step C) and slow extension from a mispaired terminus (step D) increase the opportunity for selective removal of misincorporated nucleotides by the exonuclease activity. A direct comparison of the $3' \rightarrow 5'$ exonuclease-proficient versus exo-nuclease-deficient form of Klenow polymerase (Table 2, part II) shows a sevenfold difference in the base substitution error rates of these enzymes, reflecting the contribution of proofreading to the fidelity of this polymerase.[35]

The efficiency of proofreading as an error discrimination mechanism reflects a balance between the competing processes of polymerization of the next correct nucleotide (step D) and

excision (step E). This balance is not the same for all polymerases, as demonstrated by the observation that the T4 and native T7 DNA polymerases are more accurate than is the Klenow polymerase under identical reaction conditions (Table 2, part III). Furthermore, this balance can be modulated by changing the conditions of in vitro polymerization. For example, proofreading by the Klenow polymerase is less effective at 1 mM dNTPs than at 1 μM dNTPs (Table 2, parts II and III); the high substrate concentration increases the rate of extension from mispaired termini (step D), lessening the time available for excision of the error (step E). Table 2 (part III) also illustrates that the overall fidelity of a polymerase reaction is controlled by the least accurate reaction that occurs. Although the proofreading-proficient T4 and na-

tive T7 polymerases can have an exquisitely high fidelity for particular base-substitution errors, the fidelities are significantly lower when a larger number of sites and classes of errors are monitored in a forward mutation assay.

Given that thermostable DNA polymerases are preferred for PCR, it is noteworthy that preparations of the thermostable DNA polymerase from *Thermococcus litoralis* (Vent polymerase) contain a $3' \rightarrow 5'$ exonuclease activity (P. Mattila and J. Ronka, pers. comm.). These investigators have demonstrated that this enzyme has a base-substitution error rate of 1/31,000 in reactions containing 1 mM dNTPs, similar to that observed for the Klenow polymerase (Table 2, part III). Consistent with the possibility that exonucleolytic removal of base-substitution errors is occurring during polymerization, the fidelity of this polymerase is fivefold higher than the fidelity of the exonuclease-deficient *Taq* polymerase under identical reaction conditions (Table 2, parts III and IV), and the $3' \rightarrow 5'$ exonuclease activity has a slight preference for excision of mismatched rather than matched base pairs (P. Mattila and J. Ronka, pers. comm.).

The majority of commercially available thermostable DNA polymerases do not contain $3' \rightarrow 5'$ exonuclease activities. When the fidelities of various polymerases are compared using similar reaction conditions, the *Taq*, *Thermus thermophilus* (*Tth*), and *Thermus flavis* (Replinase) DNA polymerases are approximately 10-fold less accurate than are the T4 or native T7 polymerases, consistent with these enzymes being proofreading-deficient (Table 2)[13] (P. Mattila and J. Ronka, pers. comm.). Even among exonuclease-deficient DNA polymerases, error rates are not constant. The fidelity of the AMV reverse transcriptase is significantly higher than the fidelities of the *Taq* polymerase or the exonuclease-deficient form of T7 polymerase (Sequenase 2.0) (Table 2, part IV). Fluctuations in base-substitution error rates of more than 10-fold have been seen with several DNA polymerases, reflecting differences in polymerase error discrimination at steps A–D (Fig. 1), the molecular nature of the error, and the

surrounding DNA sequence (for a more extensive discussion, see ref. 36).

Not unexpectedly, the fidelity of a polymerization reaction in vitro depends upon the synthesis conditions chosen. The influence of variations in reaction conditions on the fidelity of *Taq* polymerase has revealed that even this exonuclease-deficient DNA polymerase can be highly accurate. Significant improvements in fidelity resulted from decreasing the $MgCl_2$ concentration or the pH of the reaction.[37] Under these conditions, the error rate per nucleotide polymerized at 70°C can be as low as 10^{-5} for base-substitution errors, similar to the observed error frequencies of proof-reading-proficient enzymes (Table 2, part V). In these studies, fidelity was highest at a $MgCl_2$ concentration that was equimolar to the total concentration of dNTP substrates present in the reaction, and fidelity decreased as the concentration of free divalent cation increased.[37] Low rates for both base-substitution and minus-one base deletion errors are also obtained when *Taq* polymerase reactions were carried out using Tris, 4-morpholineethane-sulfonic acid (MES), and 1,4-piperazinebis (ethane-sulfonic acid) (PIPES) buffers at pH 5–6 (70°C). The base-substitution error rate increased 60-fold as the pH was raised approximately three units.[37] The fidelity of the *Taq* polymerase is moderately temperature dependent.[13] Two- to threefold increases in both base-substitution and minus-one base deletion error rates were observed as the reaction temperature was increased from 23–30°C to 70–80°C. Finally, the overall error rate of *Taq* polymerase was observed to increase as the dNTP concentration was increased from 1 μM to 1 mM.[13,37] This suggests that the fidelity of PCR amplification can be improved using low dNTP concentrations, since unextended mispairs will be lost during PCR because they do not yield full-length DNA products for further amplification.

DNA DAMAGE IN PCR

Incubation of DNA at high temperatures during PCR produces DNA damage, thus increasing the potential for mutations. One of the most frequent forms of DNA damage is deamination of cytosine to produce uracil.[13,38] Since uracil has the same coding potential as thymine, faithful replication of a template uracil by a DNA polymerase will result in a $C \cdot G \rightarrow T \cdot A$ transition mutation. In PCR, the greatest risk of cytosine deamination is during the denaturation step. The rate constant for deamination of cytosine in single-stranded DNA at 80°C is $\sim 1 \times 10^{-8}$ events/sec, increasing to $\sim 2 \times 10^{-7}$ events/sec at 95°C. A second common type of DNA damage is spontaneous base release resulting from hydrolysis of the *N*-glycosylic bond.[39] The rate of this hydrolytic reaction is significantly increased at high temperatures and low pH. Depurination of native DNA at pH 7.4 and 70°C occurs at a rate of $\sim 4 \times 10^{-9}$ events per sec. The release of pyrimidines, while approximately 20 times slower than purines, increases significantly as the temperature is increased to 95°C. The rate of spontaneous purine loss in native DNA is markedly increased at low pH. Following spontaneous base hydrolysis, the resulting apurinic/apyrimidinic (AP) sites in the DNA can inhibit synthesis by DNA polymerases.[39] In some cases, DNA polymerases are capable of replicating past AP sites. As AP sites are noncoding lesions, such bypass replication is error-prone and results in base-substitution mutations, most frequently transversions.

The presence of DNA damage also increases the occurrence of jumping PCR. The amount of intermolecular DNA rearrangements relative to intramolecular amplification during PCR was found to be greater when the starting DNA contained double-stranded breaks, AP sites, or ultraviolet light photoproducts.[20] Strand-switching becomes an important consideration to fidelity when ancient DNA or DNA samples from forensic analyses are used as the original DNA for amplification, because these samples often contain various forms of DNA damage.[40]

ERROR DISCRIMINATION DURING PCR

Generally, the rate of polymerase extension from mispaired or misaligned primer-termini (Fig. 1, step D) is dependent upon the concentration of the next correct dNTP substrate.[32] In the absence of proofreading, high dNTP concentrations decrease polymerase fidelity by decreasing the amount of error discrimination at the extension step, while low dNTP concentrations increase fidelity by reducing the rate of mispair extension. Such error discrimination at the extension step can be utilized for primer-specific PCR amplification.[41,42] Thus, template-primers can be differentially extended during PCR, depending upon the precise nature of the 3' terminal mispair.[41,43] While primers creating a $T \cdot G$ (primer–template), $G \cdot T$, $C \cdot A$, or $A \cdot C$ mismatch provide efficient amplification,[41] the presence of an $A \cdot A$ mismatch reduces synthesis approximately 20-fold, and the presence of $A \cdot G$, $G \cdot A$, or $C \cdot C$ mismatches reduces overall PCR product yield about 100-fold relative to amplification from correctly base-paired primers.[43] In addition, when the dNTP concentration is lowered from 800 μM to 6 μM, only $T \cdot G$ mismatches or perfect base pairs are extended.[43]

Recently, Keohavong and Thilly have described a method to quantitate errors produced during PCR using denaturing gradient gel electrophoresis (DGGE).[44] In this technique, a 224-bp DNA fragment containing the HPRT exon 3 sequence is amplified by PCR. After purification of the desired DNA band by polyacrylamide gel electrophoresis, the DNA strands are denatured and allowed to reanneal in solution, generating heteroduplex molecules between DNA strands containing errors (mutant) and wild-type DNA strands. A linear gradient of denaturant in a polyacrylamide gel is then used to separate the mutant/mutant or mutant/wild-type heteroduplexes from the wild-type homoduplex molecules. The proportion of mutant DNA molecules in the population is estimated from either the heteroduplex fraction[44,46] or from the wild-type homoduplex fraction.[45] Finally, the polymerase error frequency per base pair per cycle is calculated from the mutant fraction, the number of duplications, and the DNA target size. If desired, individual heteroduplex bands can be excised to determine the exact nature of the mutation. Unlike the M13mp2 fidelity assay, which measures errors produced during a single round of DNA synthesis,[24] the DGGE/PCR

technique detects the polymerase errors that accumulate during repeated cycles of amplification. In this system, only those errors that allow full-length synthesis are recovered in the final population, whereas errors that inhibit extension synthesis will be lost. DGGE has been used successfully to quantitate the error frequency of PCR catalyzed by various polymerases under differing conditions at the HPRT exon 3 target locus (Table 3).[44–46]

The fidelity of PCR catalyzed by the *Taq* polymerase as measured by DGGE is ~1–2 × 10^{-4} (Table 3), in close agreement with the estimates from cloning PCR products (Table 1) and with the fidelity of *Taq* polymerase in M13mp2 in vitro assays (Table 2). Higher-fidelity PCR was obtained using the proofreading-proficient T4 DNA polymerase (~1/100,000) or the thermostable *Thermococcus litoralis* DNA polymerase (Table 3). PCR catalyzed by the chemically modified form of T7 DNA polymerase (Sequenase) yielded an error frequency of approximately one error per 25,000 nucleotides per cycle (Table 3). These observations do not conflict with low-fidelity DNA synthesis obtained with Sequenase 2.0 in the M13mp2 assay (Table 2). The difference presumably reflects the fact that the chemically modified T7 polymerase retains a low but detectable 3′→5′ exonuclease activity, while Sequenase 2.0, a 28-amino-acid deletion mutant, is completely devoid of exonuclease activity.[47] The disparity may also reflect the selective loss during PCR amplification and/or DGGE analysis of some types of mutations that can be scored

in the M13mp2 system. Unexpectedly, the high fidelity observed during PCR by Sequenase requires high concentrations of deoxynucleoside triphosphates, and the error rate increases up to 1/5000 when <0.5 mM dNTPs are used.[46] The reason for this dependence is as yet unknown.

The influence of PCR conditions on the fidelity of *Taq* polymerase was also examined using the DGGE/PCR technique.[46] In these studies, less extreme synthesis conditions could be used as compared to the M13mp2 fidelity assays,[37] since suboptimal reaction conditions resulted in a decreased efficiency or absence of amplification. Nevertheless, *Taq* polymerase fidelity was observed to decrease as the reaction pH was increased from pH 8 (25°C) to pH 9 (25°C), and to decrease as the dNTP concentration was increased from 50 μM to 200 μM.[46] Surprisingly, the highest fidelity was observed at still higher dNTP concentrations. At 500 μM dNTPs and 5 mM Mg^{2+}, the observed error rate was 1/14,000; however, amplification under these conditions was inefficient and generated several products other than the desired 224-bp HPRT fragment.[46] The fidelity of the *Thermococcus litoralis* polymerase is constant over a broad range of dNTP concentrations (10 μM to 200 μM).[45] However, when 1 mM dNTPs were used for synthesis, the error rate of *T. litoralis* increased to 1/10,000,[46] consistent with possible inhibition of proofreading activity at high dNTP concentration.[34]

DNA target sequences that are

repetitive in nature may pose special problems for PCR amplification. Strand-switching events between homologous DNA segments have been observed during PCR amplification of *HLA* genes, and become more prevalent during the final cycles of amplification when the concentration of DNA becomes high.[10] An increased frequency of spurious DNA products with cycle number was also observed during amplification of tandemly repeated minisatellite DNA.[48] Attempts to amplify a segment of the human 18S rRNA gene using the *Taq* polymerase generated a 54-bp deletion artifact consistent with the formation of a stable intrastrand hairpin loop.[49] Interestingly, this deletion was not observed when the PCR amplifications were performed using the modified T7 polymerase, illustrating again that DNA polymerases can differ significantly in their biochemical properties.

SUMMARY

High-fidelity DNA synthesis conditions are those that exploit the inherent ability of polymerases to discriminate against errors. This review has described several experimental approaches for controlling the fidelity of enzymatic DNA amplification. One of the most important parameters to consider is the choice of which polymerase to use in PCR. As demonstrated by the data in Tables 2 and 3, high-fidelity DNA amplification will be best achieved by using a polymerase with an active 3′→5′ proofreading exonuclease activity (Fig. 1E). For those enzymes that are proofreading-deficient, the in vitro reaction conditions can significantly influence the polymerase error rates. To maximize fidelity at the dNTP insertion step (Fig. 1A,B), any type of deoxynucleoside triphosphate pool imbalance should be avoided. Similarly, stabilization of errors by polymerase extension from mispaired or misaligned primer-termini (Fig. 1D) can be minimized by reactions using short synthesis times, low dNTP concentrations, and low enzyme concentrations. Additional improvements in fidelity can be made by further manipulating the reaction conditions. To perform high-fidelity PCR with *Taq* polymerase, reactions should contain a low $MgCl_2$ concentration,

TABLE 3 PCR Error Rates as Measured by DGGE

Polymerase	Error rate per nucleotide per cycle	Reference
T4	~1/100,000	44
Modified T7	1/29,000	44
	1/23,000	45
	1/18,000	46
T. litoralis	1/42,000	45
	1/15,000	46
Klenow	1/7,700	44
Taq	1/11,000	45
	1/5,000	46
	1/4,800	44

not in large excess over the total concentration of dNTP substrates, and be buffered to ~pH 6 (70°C) using Bis-Tris Propane or PIPES (Table 2). These buffers have a pK_a between pH 6 and pH 7 and a small temperature coefficient (ΔpK_a/°C), allowing the pH to be maintained stably throughout the PCR cycle.

For amplifications in which fidelity is the critical issue, one should avoid the concept that conditions generating more DNA product are the better conditions. Reactions that optimize fidelity may be relatively inefficient for amplification. The total number of cycles should be kept to the minimum necessary to produce a feasible amount of product DNA for the desired use. Thus, although using low pH, low Mg^{2+}, and/or a low concentration of dNTPs may not produce an impressive amount of amplification product as visualized on an ethidium bromide stained-agarose gel, these conditions may provide an adequate amount of DNA for cloning purposes. Using a small number of cycles will also minimize the contribution of mutations resulting from DNA damage at high temperature. The DNA sequence of several independent clones from the final PCR population should be determined to verify that the original DNA sequence has been maintained during the amplification process. Ideally, the clones should be derived from separate PCR experiments to minimize the risk of overrepresentation of early polymerase errors.

The fidelity of DNA synthesis in vitro is limited by the least accurate reaction that occurs. For example, the T·G mispair is one of the most frequently produced and most easily extended base substitution error.[35] Even under high-fidelity reaction conditions, most of the *Taq* polymerase errors were the result of T·G mispairs that led to A·T→G·C transitions.[13] Therefore, the upper limit for *Taq* polymerase fidelity will be the frequency with which the enzyme creates this mispair. In general, this limitation is not restricted to a particular mispair. The sequence of the target DNA will have a profound influence on the generation of mutations. Thus, the least accurate reaction may be the result of an exceptionally high error rate at a

particular nucleotide sequence, since the phenomenon of "hot spots" for polymerization errors has been clearly established.[16]

ACKNOWLEDGMENTS

REFERENCES

1. Lacy, M.J., L.K. McNeil, M.E. Roth, and D.M. Kranz. 1989. T-Cell receptor δ-chain diversity in peripheral lymphocytes. *Proc. Natl. Acad. Sci.* **86:** 1023–1026.
2. Frohman. M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85:** 8998–9002.
3. Li. H., X. Cui, and N. Arnheim. 1990. Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. *Proc. Natl. Acad. Sci.* **87:** 4580–4584.
4. Jeffreys, A.J., R. Neumann, and V. Wilson. 1990. Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60:** 473–485.
5. Ruano, G., K.K. Kidd, and J.C. Stephens. 1990. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc. Natl. Acad. Sci.* **87:** 6296–6300.
6. Dunning, A.M., P. Talmud, and S.E. Humphries. 1988. Errors in the polymerase chain reaction *Nucleic Acids Res.* **16:** 10393.
7. 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.
8. Newton, C.R., N. Kalsheker, A. Graham, S. Powell, A. Gammack, J.

Riley, and A.F. Markham. 1988. Diagnosis of α_1-antitrypsin deficiency by enzymatic amplification of human genomic DNA and direct sequencing of polymerase chain reaction products. *Nucleic Acids Res.* **16:** 8233–8243.
9. Vrieling, H., M.L. Van Roooijen, N.A. Groen, M.Z. Zdzienicka, J.W. Simons, P.H. Lohman, and A.A. Van Zeeland. 1989. DNA strand specificity for UV-induced mutations in mammalian cells. *Mol. Cell. Biol.* **9:** 1277–1283.
10. Ennis, P.D., J. Zemmour, R.D. Salter, and P. Parham. 1990. Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: Frequency and nature of errors produced in amplification. *Proc. Natl. Acad. Sci.* **87:** 2833–2837.
11. Loh, E.Y., J.F. Elliott, S. Cwirla, L.L. Lanier, and M.M. Davis. 1989. Polymerase chain reaction with single-sided specificity: Analysis of T-cell receptor δ chain. *Science* **243:** 217–220.
12. Goodenow, M., T. Huet, W. Saurin, S. Kwok, J. Sninsky, and S. Wain-Hobson. 1989. HIV-1 isolates are rapidly evolving quasispecies: Evidence for viral mixtures and preferred nucleotide substitutions. *J. Acquired Immune Defic. Syndr.* **2:** 344–352.
13. Eckert, K.A. and T.A. Kunkel. 1991. The fidelity of DNA polymerases used in the PCR. In *Polymerase chain reaction I: A practrical approach* (eds. M.J. McPherson, P. Quirke, and G.R. Taylor). IRL Press at Oxford University Press, Oxford. (In press.)
14. Reiss, J., M. Krawozak, M. Schloesser, M. Wagner, and D.N. Cooper. 1990. The effect of replication errors on the mismatch analysis of PCR-amplified DNA. *Nucleic Acids Res.* **18:** 973–978.
15. Luria, S.E. and M. Delbruck. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28:** 491–511.
16. Kunkel, T.A. 1990. Misalignment-mediated DNA synthesis errors. *Biochemistry* **29:** 8003–8011.
17. Ripley, L.S. 1990. Frameshift mutation: Determinants of specificity. *Annu. Rev. Genet.* **24:** 189–213.
18. Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inoye. 1966. Frameshift mutations and the genetic code. *Cold Spring*

Review

Harbor Symp. Quant. Biol. **31:** 77–84.

19. Bebenek, K. and T.A. Kunkel. 1990. Frameshift errors initiated by nucleotide misincorporation. *Proc. Natl. Acad. Sci.* **87:** 4946–4950.

20. Paabo, S., D.M. Irwin, and A.C. Wilson. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* **265:** 4718–4721.

21. Kuchta, R.D., P. Benkovic, and S.J. Benkovic. 1988. Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry* **27:** 6716–6725.

22. Goodman, M.F. 1988. DNA replication fidelity: Kinetics and thermodynamics. *Mutat. Res.* **200:** 11–20.

23. Loeb, L.A. and T.A. Kunkel. 1982. Fidelity of DNA synthesis. *Annu. Rev. Biochem.* **52:** 429–457.

24. Kunkel, T.A. 1985. The mutational specificity of DNA polymerase-β during in vitro DNA synthesis. *J. Biol. Chem.* **260:** 5787–5796.

25. Echols, H. and M.F. Goodman. 1991. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem* (in press).

26. 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.

27. Preston, B.D., B.J. Poiesz, and L.A. Loeb. 1988. Fidelity of HIV-1 reverse transcriptase. *Science* **242:** 1168–1171.

28. Carroll, S.S., M. Cowart, and S.J. Benkovic. 1991. A mutant DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* **30:** 804–813.

29. Kornberg, A. 1980. *DNA replication.* W.H. Freeman and Company, San Francisco, CA.

30. Kunkel, T.A., R.A. Beckman, and L.A. Loeb. 1986. On the fidelity of DNA synthesis. Pyrophosphate-induced misincorporation allows detection of two proofreading mechanisms. *J. Biol. Chem.* **261:** 13610–13616.

31. Doubleday, O.P., P.J. Lecomte, and M. Radman. 1983. A mechanism for nucleotide selection associated with the pyrophosphate exchange activity of DNA polymerases. In *Cellular responses to DNA damage* (eds. E.C. Friedberg and B.A. Bridges), pp. 489–499. Alan R. Liss, New York.

32. Mendelman, L.V., J. Petruska, and M.F. Goodman. 1990. Base mispair extension kinetics. Comparison of DNA polymerase a and reverse transcriptase. *J. Biol. Chem.* **265:** 2338–2346.

33. Perrino, F.W. and L.A. Loeb. 1989. Differential extension of 3′ mispairs is a major contribution to the high fidelity of calf thymus DNA polymerase-α. *J. Biol. Chem.* **264:** 2898–2905.

34. Kunkel, T.A. 1988. Exonucleolytic proofreading. *Cell* **53:** 837–840.

35. Bebenek, K., C.M. Joyce, M.P. Fitzgerald, and T.A. Kunkel. 1990. The fidelity of DNA synthesis catalyzed by derivatives of *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* **265:** 13878–13887.

36. Kunkel, T.A. and K. Bebenek. 1988. Recent studies of the fidelity of DNA synthesis. *Biochim. Biophy. Acta* **951:** 1–15.

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

38. Lindahl, T. 1979. DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base-excision repair. *Prog. Nucleic Acid Res. Mol. Biol.* **22:** 135–189.

39. Loeb, L.A. and B.D. Preston. 1986. Mutagenesis by apurinic/apyrimidinic sites. *Annu. Rev. Genet.* **20:** 201–230.

40. Paabo, S., R.G. Higuchi, and A.C. Wilson. 1989. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J. Biol. Chem.* **264:** 9709–9712.

41. Newton, C.R., A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, and A.F. Markham. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17:** 2503–2516.

42. Wu, D.Y., L. Ugozzoli, B.K. Pal, and R.B. Wallace. 1989. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci.* **86:** 2757–2760.

43. Kwok, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, and J.J. Sninsky. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human

immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18:** 999–1005.

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

45. 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.* (in press).

46. 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 and Applications* (in press).

47. Tabor, S. and C.C. Richardson. 1989. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J. Biol. Chem.* **264:** 6447–6458.

48. Jeffreys, A.J., V. Wilson, R. Neumann, and J. Keyte. 1988. Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16:** 10953–10971.

49. Cariello, N.F., W.G. Thilly, J.A. Swenberg, and T.R. Skopek. 1991. Deletion mutagenesis during polymerase chain reaction: Dependence on DNA polymerase. *Gene* **99:** (in press).

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

51. Roberts, J.D., K. Bebenek, and T.A.; Kunkel. 1988. The accuracy of reverse transcriptase from HIV-1. *Science* **242:** 1171–1173.

52. Kunkel, T.A. and A. Soni. 1988. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chick embryo DNA polymerase-γ. *J. Biol. Chem.* **263:** 4450–4459.