Specificity, Efficiency, and Fidelity of PCR

Rita S. Cha and
William G. Thilly

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

The efficacy of PCR is measured by its specificity, efficiency (i.e. yield), and fidelity. A highly specific PCR will generate one and only one amplification product that is the intended target sequence. More efficient amplification will generate more products with fewer cycles. A highly accurate (i.e., high-fidelity) PCR, will contain a negligible amount of DNA polymerase-induced errors in its product. An ideal PCR would be the one with high specificity, yield, and fidelity. Studies indicate that each of these three parameters is influenced by numerous components of PCR, including the buffer conditions, the PCR cycling regime (i.e., temperature and duration of each step), and DNA polymerases. Unfortunately, adjusting conditions for maximum specificity may not be compatible with high yield; likewise, optimizing for the fidelity of PCR may result in reduced efficiency. Thus, when setting up a PCR, one should know which of the three parameters is the most important for its intended application and optimize PCR accordingly. For instance, for direct sequencing analysis of a homogenous population of cells (either by sequencing or by RFLP), the yield and specificity of PCR is more important than the fidelity. On the other hand, for studies of individual DNA molecules, or rare mutants in a heterogeneous population, fidelity of PCR is vital. The purpose of current communication is to focus on the essential components of setting up an effective PCR, and discuss how each of these component may influence the specificity, efficiency, and fidelity of PCR.

SETTING UP PCR

Template

Virtually all forms of DNA and RNA are suitable substrates for PCR. These include genomic (both eukaryotic and prokaryotic), plasmid, and phage DNA and previously amplified DNA, cDNA, and mRNA. Samples prepared via standard molecular methodologies⁽¹⁾ are sufficiently pure for PCR, and usually no extra purification steps are required. Shearing of genomic DNA during DNA extraction does not affect the efficiency of PCR (at least for the fragments that are less than ~2 kb). In some cases, rare restriction enzyme digestion of genomic DNA before PCR is suggested to increase the yield.^(2,3) In general, the efficiency of PCR is greater for smaller size template DNA (i.e., previously amplified fragment, plasmid, or phage DNA), than high molecular (i.e., undigested eukaryotic genomic) DNA.

Typically, 0.1–1 µg of mammalian genomic DNA is utilized per PCR.^(1,3,4–6) Assuming that a haploid mammalian genome (3×10^9 bp) weighs $\sim 3.4 \times 10^{-12}$ grams, 1 µg of genomic DNA corresponds to $\sim 3 \times 10^5$ copies of autosomal genes. For bacterial genomic DNA or a plasmid DNA that represent much less complex genome, picogram (10^{-12} grams) to nanogram (10^{-9} grams) quantities are used per reaction.^(1,3) Previously amplified DNA fragments have also been utilized as PCR templates. In general, gel purification of the amplified fragment is recommended before the second round of PCR. Purification of the amplified product is highly recommended if the initial PCR generated a number of unspecific bands or if a different set of primers (i.e., internal primers) is to be utilized for the subsequent PCR. On the other hand, if the amplification reaction contains only the intended target product, and the purpose of the subsequent PCR is simply to increase the overall yield utilizing the same set of primers, no further purification is required. One could simply take out a small aliquot of the original PCR mixture and subject it to a second round of PCR. In addition to the purified form of DNA, PCR from cells has also been demonstrated. In this laboratory, direct amplification of hprt exon 3 fragment from 1×10^5 human cells (following proteinase treatment to open up the cells) had been carried out routinely (P. Keohavong, unpubl.).

Primer Design

For many applications of PCR, primers are designed to be exactly complementary to the template. On the other hand, for other applications as in the allele-specific PCR, the engineering of mutation or new restriction endonuclease sites into a specific region of the genome, and cloning of homologous genes where sequence information is lacking, base pair mismatches are introduced either intentionally or unavoidably.⁽³⁾ In either case, an ideal set of primers should hybridize efficiently to the target sequence with negligible hybridization to other related sequences that are present in the sample. Primers are typically 15–30 bases long. Assuming that the nucleotide sequences of the genome is randomly distributed, the probability of a 20-base-long primer finding a perfect match is $(1/4)^{20} = 9 \times 10^{-13}$. Because there are 3×10^9 bp per haploid mammalian genome, it is highly unlikely that this primer will find another perfectly matched template in the genome. However, amplification of unspecific products in PCR, utilizing a specific 20-base primer, is not an unusual one. This is likely attributable to the fact that primers containing a number of mismatches are still amplified under most PCR conditions. For example, the likelihood of a particular region of the genome having the 12 out of 20 bases homologous to the primer is $(1/4)^{12} = 6 \times 10^{-8}$. Theoretically, there would be ~ 180 places in the haploid mammalian genome where this will occur.¹ To optimize the specificity of the genes suspected to be duplicated in the genome, primer sequences should be selected from intronic regions of the gene because they are divergent even in members of tandemly repeated gene families.

Reaction Mixture

The “standard” buffer for *Taq* polymerase-mediated PCR contains 50 mM KCl, 10 mM Tris (pH 8.3 at room temperature), and 1.5 mM MgCl_2 .⁽³⁾ The standard buffers for other DNA polymerases, including modified T7 or Sequenase,⁽⁷⁾ T4,⁽⁸⁾ *Vent*,⁽⁹⁾ and *Pfu*⁽¹⁰⁾ are also available. Although the standard buffer works well for a wide range of templates and oligonucleotide primers, the “optimal” buffer for a particular PCR may vary, depending on the target and the primer sequences, and the concentrations of other component in the reaction (i.e., dNTP and primers). Therefore, these so-called standard conditions should be regarded as a point of departure to explore modifications and potential improvements. In particular, the concentration of Mg^{2+} should be optimized whenever a new combination of target and primers is first used or when the concentration of dNTPs or primers is altered. dNTPs are the major source of phosphate groups in the reaction, and any change in their concentration affects the concentration of available Mg^{2+} . The presence of divalent cations is critical, and it has been shown that magnesium ions are superior to manganese, and that calcium ions are ineffective.⁽¹¹⁾ In addition to the standard components of the PCR buffer mentioned above, some researchers routinely use additional components such as gelatin, Triton X-100, or bovine serum albumin for stabilizing enzymes, glycerol⁽¹²⁾ or formamide⁽¹³⁾ to enhance specificity, and mineral oil to prevent evaporation of water in the reaction mixture.

Primers and dNTP

To ensure that the target DNA is efficiently amplified, one must ensure that the reaction mixture contains nonlimiting (i.e., excess) amounts of primers and dNTPs. Typically, in a 100- μl reaction mixture, between 0.3 (1.8×10^{13}

¹ $(3 \times 10^9 \text{ bp}) \times (6 \times 10^{-8}) = 180$.

molecules) and $3\ \mu\text{M}$ (1.8×10^{14} molecules) of each primer, and between $37\ \mu\text{M}$ (2.2×10^{15} molecules) and $1.5\ \text{mM}$ (9×10^{16} molecules) of each dNTP are utilized. Thus, for a genomic DNA PCR containing $1\ \mu\text{g}$ of template DNA (3×10^5 copies of autosomal genes), the initial molar ratio between the primers and the genomic target sequence is at least $\sim 10^8:1$. Having such a large excess of primers ensures that once template DNA becomes denatured, it will anneal to primers rather than to each other. Because the maximum copy number of amplified target sequence is $\sim 10^{12}$ copies (see Fig. 2 and Exponential Phase of PCR, below), each primer is always in at least 10-fold excess of the target sequence (assuming that primers are not consumed by generating unspecific amplification products). The ratio of the primer to template is also important regarding the specificity of PCR. If the ratio is too high, PCR is more prone to generate unspecific amplification products, and also primer dimers are formed. However, if the ratio is too low (i.e., $<0.1\%$ of the standard condition for the genomic DNA PCR), the efficiency of PCR is greatly compromised.

For primers, the fraction of free (i.e., unincorporated) primers is strictly dependent on how many target sequences are generated. Unlike primers, however, the fraction of free dNTPs not only depends on the number of target sequences generated but also on the size of the target sequence. For example, generating 10^{12} copies of a 100-bp target sequence will consume $10^{12} \times 100 = 10^{14}$ dNTP molecules. On the other hand, generating 10^{12} copies of a 2-kb fragment will consume 2×10^{15} dNTP molecules and effectively decrease the concentration of free dNTP. This, in turn, will have a deteriorative effect on the overall efficiency of PCR. Thus, for amplifying a large target sequence, a higher concentration of dNTP is recommended.⁽⁷⁾

PCR Cycle

A typical PCR cycle consist of three steps: (1) a denaturation step (1–2 min of incubation at $\geq 94^\circ\text{C}$); (2) a primer annealing (or hybridization) step (1–2 min of incubation at $50\text{--}55^\circ\text{C}$); and (3) an extension step (1–2 min of incubation at 72°C). Previously, it was believed that each of the three steps in the cycle requires a minimal amount of time to be effective while too much time at each step can be both wasteful (time wise) and deleterious to the DNA polymerase.⁽³⁾ Recently, however, we have demonstrated that PCR consisting of two steps (e.g., a denaturation step— 94°C incubation step for 1 min followed by a primer hybridization/extension step at $50\text{--}57^\circ\text{C}$ for 1 min) can generate as much product as three step PCR.⁽⁹⁾ This has been the case for at least four different sets of primers tested on two different genes.^(12,14) This finding appears somewhat inconsistent with the previous contention that *Taq* polymerase is most active at 72°C . However, it is also possible that because of the high processivity of *Taq*, primers that annealed to the template will become fully extended during the short time period during which the reaction mixture reaches the optimal temperature for *Taq* polymerase ($70^\circ\text{C}\text{--}75^\circ\text{C}$) between the $50^\circ\text{C}\text{--}94^\circ\text{C}$ transition. This notion is also consistent with the results of “rapid PCR.”⁽¹⁵⁾ In an attempt to increase the speed of temperature cycling (i.e., reduce “ramp times”), researchers have utilized capillary tubes as containers and air as the heat-transfer medium for PCR. This study reports that for a $100\text{-}\mu\text{l}$ sample in a standard heat block, it takes ~ 6 sec to go from 56°C to 55°C . On the other hand, for a $10\text{-}\mu\text{l}$ sample in a commercially available “rapid air” cycler, it takes just ~ 1 sec to go from 60°C to 55°C . Utilizing the rapid cycler, the investigators completed 35 cycles of three-step PCR in 15 min.⁽¹⁵⁾ In this rapid PCR, each cycle consisted of a 0-sec denaturation step at 94°C , a 0-sec annealing step at 45°C , and a 10-sec elongation step at 72°C . In addition to improving cycle times, the rapid cycle PCR amplification was

more specific than three-step PCR utilizing a conventional thermal cycler. One possible limitation of the currently available rapid PCR technique is its small reaction volume (10 μ l). Because of these volume constraints, only 50 ng of mammalian genomic DNA was used as PCR template. Because it represents $\sim 1.5 \times 10^4$ copies, it would not be useful for detecting rare mutations. Nevertheless, rapid PCR could be utilized effectively for the analysis of less complex genomes (i.e., bacteria, plasmid, or phages) and/or homogeneous populations. Finally, as in the case of the standard PCR buffer, the standard three-step PCR regime should also be viewed as a point of departure from which further improvement could be made. In general, higher annealing temperature and shorter time allowed for annealing and extension steps improved specificity of PCR. It should also be pointed out that it is necessary to increase the duration of each step for efficient amplification in the amplification of large fragments (i.e., >1 kb).^(3,16)

EXPONENTIAL PHASE OF PCR

To set up an informative and analytical PCR, one must understand the kinetics of specific product accumulation during PCR. A schematic representation of different products accumulating as a function of cycle is depicted in Figure 1. The desired blunt-ended duplex fragments appear for the first time during the third cycle of the PCR, and from this point on, this product accumulates exponentially according to the formula $N_f = N_o (1 + Y)^{n-1}$, where N_f is the final copy number of the double-stranded target sequence, N_o is the initial copy number, Y is the efficiency of primer extension per cycle, and n is the number of PCR cycles under conditions of exponential amplification.⁽⁴⁾ As depicted in Figure 2, in most cases, once the final copy number of the desired fragment (N_f) reaches $10^{12} \sim 10^{13}$, its efficiency per cycle (Y) drops dramatically, and at the same time, the product stops accumulating exponentially. The exponential phase of a PCR refers to the early cycle period during which the products accumulate in a manner that is consistent with the equation above. Continuing PCR beyond this point often results in amplification of unspecific bands, and, in certain instances, disappearance of the specific product (G. Hu, unpubl.). These undesired effects of overamplification are presumably attributable to the fact that as the number of cycle increases and more products are generated, some components of the PCR becomes limiting. Consistent with this notion, taking a small aliquot of the reaction mixture that has already undergone 10^6 – 10^7 doublings and placing it in a fresh reaction mixture results in exponential amplification.

For many applications of PCR, especially the ones that are quantitative in nature, it is critical that amplification is carried out in the exponential phase of PCR (Fig. 2). Numerous laboratories have studied efficiencies of different DNA polymerases that are utilized in PCR. As a result, we now have a fairly good idea regarding how efficient different DNA polymerases are in a typical PCR.^(17,18) Thus, utilizing the equation above, knowledge of the initial copy number will permit one to estimate how many cycles it will take for the final copy number to reach $\sim 10^{12}$, and at which point one should stop the PCR. For *Taq* PCR (assuming efficiency per cycle of 70%) starting with 1 μ g of genomic DNA (e.g., 3×10^5 copies of mammalian genome), the equation becomes, $10^{12} = 3 \times 10^5 (1 + 0.7)^n$. Solving for n gives 28.6, indicating that in this hypothetical case, the desired product will accumulate exponentially up to about cycle number 29. Thus, if one were to carry out an analysis that is quantitative in nature, one must do so on the samples that are taken out at or before cycle 29. Because 10^{12} copies of a particular sequence is sufficient for most application in molecular biology, there is no apparent reason to carry out additional cycles.

FIGURE 1 Schematic representation of PCR. N_0 copies of duplex template DNA are subjected to n cycles of PCR. During each cycle, duplex DNA is denatured by heating, which then allowed primers (arrows) to anneal to the targeted sequence (hatched square). In the presence of DNA polymerase and dNTPs, primer extension takes place. The desired blunt-ended duplex product (thick bars with arrows) appears during the third cycle and accumulates exponentially during subsequent cycles. Following n cycles of exponential PCR, there will be $N_0 (1 + Y)^{n-1}$ copies of the duplex target sequence.

It must be pointed out that the efficiency of the same polymerase can vary significantly depending on the nature of target sequence, the primer sequences, and the reaction conditions.⁽¹⁹⁾ Therefore, the efficiencies listed in Table 1 may not reflect the efficiency of a different PCR that is carried out under different conditions. One could utilize the reported values to have a reasonable estimate. Nevertheless, to carry out an accurate quantitative analysis, one should determine the efficiency of the particular PCR (see Fig. 2).

DNA POLYMERASES AND PCR

In vitro DNA replication has been accomplished by DNA polymerases from many different sources.^(4,7-9,21,22) The initial PCR procedure described by Saiki et al.⁽⁴⁾ utilized the Klenow fragment of *Escherichia coli* DNA polymerase I. This enzyme was heat labile; and as a result, fresh enzyme had to be added

FIGURE 2 Accumulation of target sequence during PCR as a function of number of cycles. Approximately 10^5 (N_0) copies of rat *H-ras* gene exon 1 are subjected to 60 cycles of PCR under a standard *Taq* PCR condition.⁽¹²⁾ A 2.5- μ l aliquot is taken at 20, 25, 30, 35, 40, 45, 50, 55, and 60 cycles (n) and analyzed on a polyacrylamide gel. The number of target sequence generated at each stage (N_f) is estimated based on the intensity of the band following ethidium bromide staining. *Taq* (2.5 units) is added following 30 cycles of PCR.

during each cycle following the denaturation and primer hybridization steps. Introduction of the thermostable *Taq* polymerase in PCR⁽²⁰⁾ subsequently alleviated this tedium and facilitated the automation of the thermal cycling portion of the procedure. For PCR, thermostable DNA polymerases (e.g., *Taq*, *Vent*, and *Pfu*) are preferred over heat-labile polymerases (e.g., T4, T7, and Klenow) simply because they are much easier to handle and, most importantly, amenable to automation.

Studies have shown that different DNA polymerases have distinct characteristics, which affects the efficacy of PCR. For example, *Taq* polymerase does not have the 3'→5' exonuclease "proofreading" function; and as a result, it has a relatively high error rate in PCR (Table 1). On the other hand, its inability to edit the mispaired 3' end has been an asset for researchers who developed the allele-specific PCR based on the concept that primers contain-

TABLE 1 Fidelity and Efficiency of DNA Polymerases Used in PCR

Enzyme	Error rate (errors/base)	PCR-induced mutant fraction ^a (%)	Efficiency per cycle (%)	Number of cycles required ^b	Reference
<i>Taq</i>	2×10^{-4}	56	88	22	17, 20
<i>Taq</i>	7.2×10^{-5}	25	36	45	18
Klenow	1.3×10^{-4}	41	80	24	6, 17
T7	3.4×10^{-5}	13	90	22	17
T4	3×10^{-6}	2	56	32	17
<i>Vent</i>	4.5×10^{-5}	16	70	26	18

^aFraction of PCR-induced noise following 10^6 -fold amplification of 200-bp target sequence given the error rate.

^bNumber of cycles required to obtain 10^6 -fold amplification given the efficiency per cycle.

ing mismatches at the 3' end were not extended as efficiently as the perfectly matched primers.^(12,16,23,24) This concept would not have worked for enzymes with exonuclease activities, because once the 3' mismatch is recognized by the polymerase, it would be first repaired and then extended, thus abolishing the specificity conferred by the 3' mismatches. As applications of the PCR become increasingly sophisticated and specific, distinctive properties of polymerases should be utilized to meet specific needs.

Fidelity of in vitro DNA polymerization is perhaps one of the most intensively studied subjects in PCR. For many applications of PCR, where a relatively homogeneous DNA population is analyzed (i.e., direct sequencing or restriction endonuclease digestion), the polymerase-induced mutations during PCR are of little concern. In general, polymerase-induced mutations are distributed randomly over the sequence of interest, and an accurate consensus sequence is usually obtained. However, PCR is also utilized for studies of rare molecules in a heterogeneous population. Examples include the study of allelic polymorphism in individual mRNA transcripts,^(25,26) the characterization of the allelic stages of single sperm cells⁽²⁷⁾ or single DNA molecules^(28,29), and characterization of rare mutations in a tissue⁽¹²⁾ or a population of cells in culture. For these applications, it is vital that the polymerase-induced mutant sequences do not mask the rare DNA sequences. Each polymerase-induced error, once introduced, will be amplified exponentially along with the original wild-type sequences during subsequent cycles. This will result in the overall increase in the fraction of polymerase-induced mutant sequences as a function of number of amplification cycles. Analyses that utilize small amounts of template DNA are especially prone to PCR-induced artifacts. For example, if one were to carry out PCR with 10 copies of template DNA, any polymerase-induced mutation during the first few cycles would appear as a major mutant population in the final PCR products. Because the number of template DNAs is small, and the error rate of *Taq* polymerase is $\sim 10^{-4}$, the probability of this event occurring is low (i.e., 10^{-3}). However, if such event should occur, the particular mutation induced by polymerase will comprise as much as 10% of the final PCR products. One can prevent this "jackpot" artifact by starting with a large amount of template DNA (i.e., $\geq 10^5$ copies). In this case, ~ 10 mutations are introduced on the average during the first few cycles of the PCR; however, all of these mutations will constitute only ~ 1 in 10^5 of the final products.

Under low-fidelity conditions (i.e., *Taq* or Klenow PCR), the polymerase induced mutant fraction can become significant. For example, following 1 million-fold amplification by a DNA polymerase with an error rate of 10^{-4} , PCR-induced error will constitute as much as 33% of the 200-bp long amplified products.² Assuming that polymerase errors are uniformly distributed, the error frequency per base, on average, is 1.7×10^{-3} ($0.33 \times 200 = 0.0017$). This level of PCR-induced noise will certainly render attempts to characterize rare mutations in tissue culture or in animals and humans where the expected mutant frequency of a particular mutation could be as low as 10^{-7} or 10^{-8} .

The fidelity of PCR varies depending on reaction conditions and the nature of target sequences. In the past, several groups have found conditions that permitted for more accurate PCR by modifying reaction buffer conditions. For instance, Ling et al.⁽¹⁸⁾ were able to reduce the error rate of *Taq* PCR by a factor of 2.8 (from 2×10^{-4} to 7.2×10^{-5}) by modifying reaction conditions. One may assess the significance of this 2.8-fold improvement on *Taq* PCR

²Fraction of PCR-induced mutants is calculated according to a formula $F(>1) = 1 - e^{-bfd}$, where b is the length of target sequence, f is the error rate, and d is the number of doublings.⁽¹⁷⁾ Thus, following a 10^6 -fold amplification (e.g., 20 doublings) of a 200-bp fragment at an error rate of 10^{-4} /bp incorporated will lead to an estimated PCR-induced mutant fraction of 33% ($1 - e^{-(200)(10^{-4})(20)} = 0.33$).

fidelity by comparing the fractions of PCR-induced noise before and after the improvement. According to the formula $F(>1) = 1 - e^{-bfd}$,⁽¹⁸⁾ 56% of the PCR product amplified under low-fidelity condition will be *Taq* polymerase-induced noise (Table 1). On the other hand, only 25% of the PCR product generated under the high-fidelity condition will be polymerase-induced noise. In this case, a 2.8-fold reduction in the *Taq* polymerase error rate reduced the overall PCR-induced mutant fraction by more than half (Table 1). Thus, it is possible to substantially improve the overall fidelity of PCR by adjusting reaction conditions. Nevertheless, it must be pointed out that despite much effort to optimize *Taq*, T7, and *Vent* PCR in regard to fidelity by altering reaction conditions, their improved fidelity has never reached the level of T4 polymerase,^(17,18) suggesting that some intrinsic properties of the polymerase also contribute to its overall error rate. Regarding the error rates of exo^+ polymerases, one should realize that the measured error rate reflects the average value of a heterogeneous population of DNA polymerases, with this heterogeneity presumably arising as a result of errors during transcription of the gene. It is possible that some of the transcription errors are introduced in the region of the gene that is critical for fidelity of the polymerase (i.e., the proofreading function) and, thus, increase the average error rate. If this is the case, one may be able to enhance the fidelity of exo^+ polymerase PCR by devising a means to physically separate, or biologically inactivate these rare exo^- mutant polymerases (W. Thilly, unpubl.).

In addition to the error rate during PCR, the kinds of mutations that are introduced during PCR are also dependent on DNA polymerases. Whereas GC→AT transitions were the predominant mutations for T4 and T7 polymerases, AT→GC transitions are observed most frequently with *Taq* polymerase.⁽¹⁷⁾ *Taq* polymerase is also highly prone to generate deletion mutations if the template DNA has the potential to form secondary structures.⁽¹⁰⁾ Klenow fragment induces possible transitions, and deletions of 2 and 4 bp. These observations again suggest that each polymerase has a distinctive mode of operation with regard to fidelity in *in vitro* replication.

The findings that different polymerases induce different types of mutations in PCR also have a very practical value in designing PCR-based experiments. For example, if one were to look for a rare allele that had undergone a GC→AT transition, it would be best to use *Taq* polymerase for PCR. Because *Taq* predominantly induces AT→GC transitions,⁽¹⁷⁾ utilizing *Taq* will minimize false-positive cases that may arise as a result of a *Taq* polymerase-induced artifact. In another hypothetical case, assume that *Taq* PCR followed by sequencing analysis (either by cloning and sequencing, or by DGGE type analysis followed by sequencing) reveals that in the population of cells analyzed, a rare AT→GC mutant allele exists at a frequency of 10^{-3} . However, because this mutation is the type expected from *Taq* amplification, one cannot be certain whether this is a true variant in the original sample or a PCR artifact. To distinguish between these two possibilities, one may carry out the same analysis again using a T7 or T4 polymerase to determine whether the AT→GC mutation appears again. If this AT→GC mutation appears following PCR mediated by two different enzymes with different mutational specificities, it is fair to say that the mutation existed in the original sample.

Because of its thermostability, reliability, and durability, *Taq* DNA polymerase has been utilized most widely in PCR. However, as summarized in Table 1, the fidelity of *Taq* (2×10^{-4} error/bp per duplication) is the lowest among DNA polymerases with fidelity that has been measured. This, in turn, effectively prevents utilization of *Taq* polymerase in PCR where the fidelity is of concern. Thus far, the most accurate DNA polymerase utilized in PCR is T4 polymerase. Its error rate is estimated to be 3×10^{-6} errors/bp per duplication (Table 1). The fraction of PCR induced noise in a 200-bp target sequence

following a 10^6 -fold amplification in this case is 2% (56% for *Taq* PCR; see Table 1). Unfortunately, T4 polymerase is not thermostable; thus, not many laboratories will be willing to utilize this enzyme, especially if the number of samples to be analyzed is large. Recently, a number of additional thermostable enzymes have been isolated. Unlike *Taq*, which does not have the 3'→5' exonuclease proofreading function, these newly isolated enzymes (e.g., *Vent* and *Pfu*) do have the editing function. And as expected, they turned out to be more accurate than *Taq* polymerase. As summarized in Table 1, *Vent* has a error rate and efficiency that are comparable to that of the heat-labile T7 DNA polymerase. Although *Vent* polymerase is not as accurate as T4 polymerase, the fact that it can be automated makes it a much more attractive choice of enzyme. Thus, for analyzing samples where fidelity of the product is important, *Vent* PCR appears to be the best choice regarding its error rate, efficiency, and ease of the procedure. Note that *Vent* PCR product will have less than one-third (56% vs. 16%) of the noise induced by *Taq* polymerase. Our laboratory is currently in the process of optimizing *Pfu* in regard to its fidelity.

ANALYZING PCR RESULTS

PCR results may be analyzed in regard to its specificity, yield (i.e., efficiency), or fidelity. Specificity and yield can be readily determined by running a gel that separates DNA molecules according to their sizes (e.g., polyacrylamide or agarose gels). A highly specific PCR would generate one and only one product of the correct size. However, it is not unusual to observe a series of bands, especially when a new target sequence and/or primers are utilized for the first time. Appearance of unspecific amplification products can be attributed to a number of factors. First, primers may be annealing to unspecific sites in template DNA. In this case, one may be able to increase the specificity of PCR by changing reaction mixture that would make it more difficult for primers to anneal to unspecific sites in the sample. These include addition of glycerol,⁽¹²⁾ or formamide,⁽¹³⁾ reduced pH, or lowering concentrations of primers, dNTPs and $MgCl_2$.^(16,23,30) One may also try altering the annealing temperature and/or the duration of the annealing and extension steps. In general, higher temperature, and shorter annealing and extension periods confer higher specificity.^(12,15) Alternatively, the unspecific bands may have resulted from overamplification (Fig. 2; see Exponential Phase of PCR, below). In this case, one can simply reduce the number of cycles. As mentioned earlier, amplified products should be analyzed while they are still in the exponential phase of PCR. This is not only crucial for extracting quantitative information (e.g., calculating efficiencies, estimating the initial copy numbers), but also for generating the specific target sequence. Undesired consequences of overamplification include generation of small deletion mutants, appearance of unspecific bands, and in some cases disappearance of the specific product (G. Hu, unpublished observations). If none of these have a significant effect on the specificity of amplification, it may be necessary to change the primer; unfortunately, some primers simply do not work.

For many applications of PCR where rare variants are involved, the fidelity of PCR is an important concern. A number of laboratories have studied the fidelity of PCR, and the error rates of commonly utilized DNA polymerases are known (Table 1). However, because the fidelity of a polymerase varies significantly depending on the reaction conditions and the nature of target sequences, it must be determined on a sequence by sequence and/or a reaction condition by reaction condition basis. There are at least three independent methods of measuring the fidelity of PCR: (1) the forward mutation assay; (2) the reversion mutation assay; and (3) the DGGE (denaturant gradient gel electrophoresis)-type analysis.

The forward mutation assay consists of cloning individual DNA molecules from the amplified population and determining the number of DNA sequence changes by what fraction of the cloned population displays a particular phenotype.^(6,13,31) For example, one can assess the error rate during synthesis of the *lacZ* gene by the frequency of light blue and colorless (mutant) plaques among the total plaques scored. The nature of mutations can also be determined by DNA sequence analysis of a collection of the mutants. The second method is a reversion mutation assay using a phage template DNA that contains specific mutations that result in a measurable phenotype (i.e., *lacZ*⁻, or colorless phenotype). In these assays, polymerase-induced errors are scored as DNA sequence changes that revert the mutant back to a wild-type or pseudo-wild-type phenotype.⁽³²⁾ This approach has been especially useful for highly accurate polymerases.⁽¹⁹⁾ However, unlike the forward mutation assay, reversion assays are focused on a limited subset of errors occurring at only a few sites. Although in general, polymerase-induced mutations are randomly distributed throughout a target sequence, there are a number of locations in the target sequence that are more prone to polymerase-induced errors.⁽²³⁾ Thus, error rates measured by the reversion assay may vary significantly depending on the nature of initial mutations placed in the phage template. The third method of assessing fidelity of PCR utilizes the DGGE analysis. DGGE is a system that separates DNA fragments harboring small changes (i.e., single-base substitutions, small additions, or deletions) based on their sequences. In this case, the DGGE is utilized to separate polymerase-induced mutant sequences from the correctly amplified sequences. By measuring the fraction of signals coming from the portion of the gel corresponding to the polymerase-induced mutant sequences (heteroduplex fraction), one could calculate the fidelity of the enzyme according to a formula, $f = HeF/(b \times d)$, where f is the error rate (errors/base pair incorporated per duplication), HeF is the heteroduplex fraction, b is the length of the single-strand low melting domain in which mutants can be detected, and d is the number of DNA duplications.^(17,18) Unlike the other two assays in which only the changes that result in phenotypic changes are scored as PCR-induced mutations, DGGE allows for the visualization and detection of all the mutations introduced in the target sequence. This feature makes DGGE the most comprehensive and sensitive means of measuring fidelity of PCR among the currently available techniques.

CONCLUSIONS

PCR is utilized for rapid in vitro amplification of a specific fragment of (genomic) DNA or RNA. The ideal PCR is the one with high specificity, yield (i.e., efficiency), and fidelity. The specificity, yield, and fidelity of PCR are influenced by the nature of target sequence, as well as by each component of PCR. Often, the conditions that would permit maximum yield are not compatible with high fidelity or specificity, and conditions optimized in regard to fidelity may adversely affect the efficiency. Thus, in setting up a PCR, it is important to plan beforehand to attain the specificity, yield, and the fidelity of PCR that is required for the intended application. As mentioned above, DNA polymerases that are utilized in PCR have different characteristics that affect the overall PCR efficiency as well as the fidelity. Understanding the purpose of PCR will also allow one to choose the appropriate polymerase for PCR. Unfortunately, the most accurate enzyme studied to date, T4 polymerase, and the most efficient polymerase, T7, are not thermostable, and thus will not be become widely utilized in PCR. Among the remaining thermostable polymerases, *Vent* or *Pfu* would be the enzyme of choice if fidelity is of concern, whereas *Taq* will be sufficient (in regard to its fidelity) if the purpose of PCR

is simply to generate a large quantity of a specific target sequence. With additional thermostable enzymes being isolated, and our understanding of polymerase error discrimination rapidly increasing, it is possible that we may eventually achieve a PCR utilizing thermostable enzyme with fidelity comparable with that of T4 polymerase.

REFERENCES

1. Sambrook, J., E.F., Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
2. Keohavong, P., C.C. Wang, R.S. Cha, and W.G. Thilly. 1988. Enzymatic amplification and characterization of large DNA fragments from genomic DNA. *Gene* **71**: 211–216.
3. D.M. Coen. 1991. The polymerase chain reaction. *Current protocols in molecular biology*. Greene Publishing/Wiley Interscience, New York.
4. Saiki, R.K., S. Scharf, F. Faloona, K. B. Mullis, G.T. Horn, H.A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.
5. Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods Enzymol.* **155**: 335–350.
6. Scharf, S.J., G.T. Horn, and H.A. Erlich. 1986. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* **233**: 1076–1078.
7. Keohavong, P., C.C. Wang, R.S. Cha, and W.G. Thilly. 1988. Enzymatic amplification and characterization of large DNA fragments from genomic DNA. *Gene* **71**: 211–216.
8. Keohavong, P., A.G., Kat, N.F. Cariello, and W.G. Thilly. 1988. Laboratory Methods: DNA amplification in vitro using T4 DNA polymerase. *DNA* **7**: 63–70.
9. VentTM DNA polymerase technical bulletin from New England Biolabs. 1990,1991.
10. 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**: 105–108.
11. Chien, A., D.B. Edgar, and J.M. Trela. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127**: 1550.
12. Cha, R.S., H. Zarbl, P. Keohavong, and W.G. Thilly. 1992. Mismatch amplification mutation assay (MAMA): Application to the c-H-ras gene. *PCR Methods Applic.* **2**: 14–20.
13. Sarkar, G., S. Kapelner, and S.S. Sommer. 1990. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* **18**: 7465.
14. Jin, Z., R. Cha, and H. Zarbl. (in prep.).
15. Witter, C.T., B.C. Marshall, G.H. Reed, and J.L. Cherry. 1993. Rapid cycle allele-specific amplification: studies with the cystic fibrosis ΔF_{508} locus. *Clin. Chem.* **39**: 804–809; Witter, C.T. and D.J. Garling. 1991. Rapid cycle DNA amplification: Time and temperature optimization. *BioTechniques* **10**: 76–83.
16. Kowk, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, 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.* **17**: 2503–2516.
17. Keohavong, P. and W.G. Thilly. 1989. Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci.* **86**: 9253–9257.
18. 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.
19. Eckert, K.A. and T.A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Applic.* **1**: 17–24.
20. Dunning, A.M., P. Talmud, and S.E. Humphries. 1988. Errors in the polymerase chain reaction. *Nucleic Acids Res.* **16**: 10393.
21. Saiki, R.K., D.H. Gelfand, S. Stoffe, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* **239**: 487–491.
22. Stratagene Product Catalog, 1992.
23. Wu, D.Y., L. Ugozzoli, B.J. 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; 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. *Nucleic Acids. Res.* **17**: 2503–3516.
24. Bottema, C.D.K. and S.S. Sommer. 1993. PCR amplification of specific alleles: Rapid detection of known mutation and polymorphisms. *Mutat. Res.* **288**: 93–102.
25. 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.
26. 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.
27. Li, H., X. Cui, and N. Arnheim. 1990. Direct electrophoretic detection of the allelic state of a single DNA molecules in human sperm by using the polymerase chain reaction. *Proc. Natl. Acad. Sci.* **87**: 4580–4584.
 28. 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.
 29. 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.
 30. Loeb, L.A. and T.A. Kunkel. 1982. Fidelity of DNA synthesis. *Annu. Rev. Biochem.* **52**: 429–457; Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acid Res.* **18**: 3739–3744.
 31. Goodenow, M. T., W. 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. AIDS* **2**: 344–352.
 32. Kunkel, T. A. 1985. The mutational specificity of DNA polymerase- β during in vitro DNA synthesis. *J. Biol. Chem.* **260**: 5787–5796.

Specificity, efficiency, and fidelity of PCR.

R S Cha and W G Thilly

Genome Res. 1993 3: S18-S29

References

This article cites 27 articles, 10 of which can be accessed free at:
<http://genome.cshlp.org/content/3/3/S18.full.html#ref-list-1>

License

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](#).

Affordable, Accurate
Sequencing.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>