Nucleic Acids Research, 1999, Vol. 27, No. 8 **1819–1827**

Nucleotide analogs and new buffers improve a generalized method to enrich for low abundance mutations

Joseph P. Day, Robert P. Hammer[1], Donald Bergstrom[2] and Francis Barany*

Department of Microbiology, Box 62, Hearst Microbiology Research Center, Strang Cancer Prevention Center, Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, USA, [1]Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803, USA and [2]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 46208, USA

Received as resubmission January 19, 1999; Revised and Accepted February 18, 1999

ABSTRACT

A high sensitivity method for detecting low level mutations is under development. A PCR reaction is performed in which a restriction site is introduced in wild-type DNA by alteration of specific bases. Digestion of wild-type DNA by the cognate restriction endonuclease (RE) enriches for products with mutations within the recognition site. After reamplification, mutations are identified by a ligation detection reaction (LDR). This PCR/RE/LDR assay was initially used to detect PCR error in known wild-type samples. PCR error was measured in low $|\Delta pK_a|$ buffers containing tricine, EPPS and citrate, as well as otherwise identical buffers containing Tris. PCR conditions were optimized to minimize PCR error using perfect match primers at the *Msp*I site in the p53 tumor suppressor gene at codon 248. However, since mutations do not always occur within pre-existing restriction sites, a generalized PCR/RE/LDR method requires the introduction of a new restriction site. In principle, PCR with mismatch primers can alter specific bases in a sequence and generate a new restriction site. However, extension from 3′ mismatch primers may generate misextension products. We tested conversion of the *Msp*I (CCGG) site to a *Taq*I site (TCGA). Conversion was unsuccessful using a natural base T mismatch primer set. Conversion was successful when modified primers containing the 6*H*,8*H*-3,4-dihydropyrimido[4,5-*c*][1,2]oxazine-7-one (Q_6) base at 3′-ends were used in three cycles of preconversion PCR prior to conversion PCR using the 3′ natural base T primers. The ability of the pyrimidine analog Q_6 to access both a T-like and C-like tautomer appears to greatly facilitate the conversion.

INTRODUCTION

High sensitivity medical diagnostic assays depend on accurate DNA amplification by DNA polymerases (1–4). Such DNA-based diagnostic methods are needed, for example, to improve cancer staging and aid clinical decisions through molecular characterization of the disease. Low level mutations may be detected by cloning PCR-amplified fragments and accurately quantified by probing for mutant DNA using allele-specific oligonucleotides (ASOs), however, this process requires several days to complete (5,6). Alternatively, allele-specific PCR (AS-PCR) can be used to detect low abundance mutations. By designing primers with one or more mismatches, mutant DNA template can be efficiently extended, while poor extension is achieved on wild-type DNA template. However, once these primers extend with or without a mismatch, the products thereafter are perfect matches for the primer in subsequent PCR cycles. Thus, false positive signals are amplified in subsequent cycles. Moreover, PCR error can generate a base change in the template which perfectly matches the primer. AS-PCR can detect pyrimidine↔purine transversions at sensitivities of 1 in 10^5 (7,8). Nevertheless, the majority of cancer-associated mutations are C↔T and A↔G transitions, for example, >80% of p53 point mutations (9). A DNA diagnostic method is required to accurately quantify this type of low abundance mutation.

The ligation detection reaction (LDR) uses two adjacent primers and a thermostable ligase to distinguish all four bases potentially found at any position in a DNA sequence (10–13). Thermostable ligase demonstrates the highest fidelity when the discriminating base is located at the 3′-end of the upstream primer (14). PCR/LDR (PCR of a sequence from genomic DNA followed by LDR) can detect mutations with a sensitivity of approximately one mutant allele in 4000 normal alleles (13). Sensitivity of approximately 1 in 10^6 has been achieved by combining PCR with restriction endonuclease (RE) digestion of wild-type DNA (15,16). Mutations occurring within the restriction site prevent cleavage of the mutant allele, while wild-type alleles bearing canonical restriction site sequence are depleted. As a result, subsequent PCR cycles preferentially amplify mutant DNA. If a mutation site is not within an endonuclease recognition site present in wild-type DNA, a restriction site must be introduced. This is typically done by PCR using a primer or primers with mismatched bases. Mutations cannot be detected in any portion of the restriction site spanned by the primers, since those bases are introduced directly through the primers. In a random DNA sequence, >20% of bases are contained within a pre-existing four base restriction site and 60% of bases are within a four base subsequence that can be converted into a restriction site by a single base change. In these small sites, 3′-terminal base mismatch primers must frequently be used. While conceptually

*To whom correspondence should be addressed. Tel: +1 212 746 4288; Fax: +1 212 746 8588; Email: barany@mail.med.cornell.edu

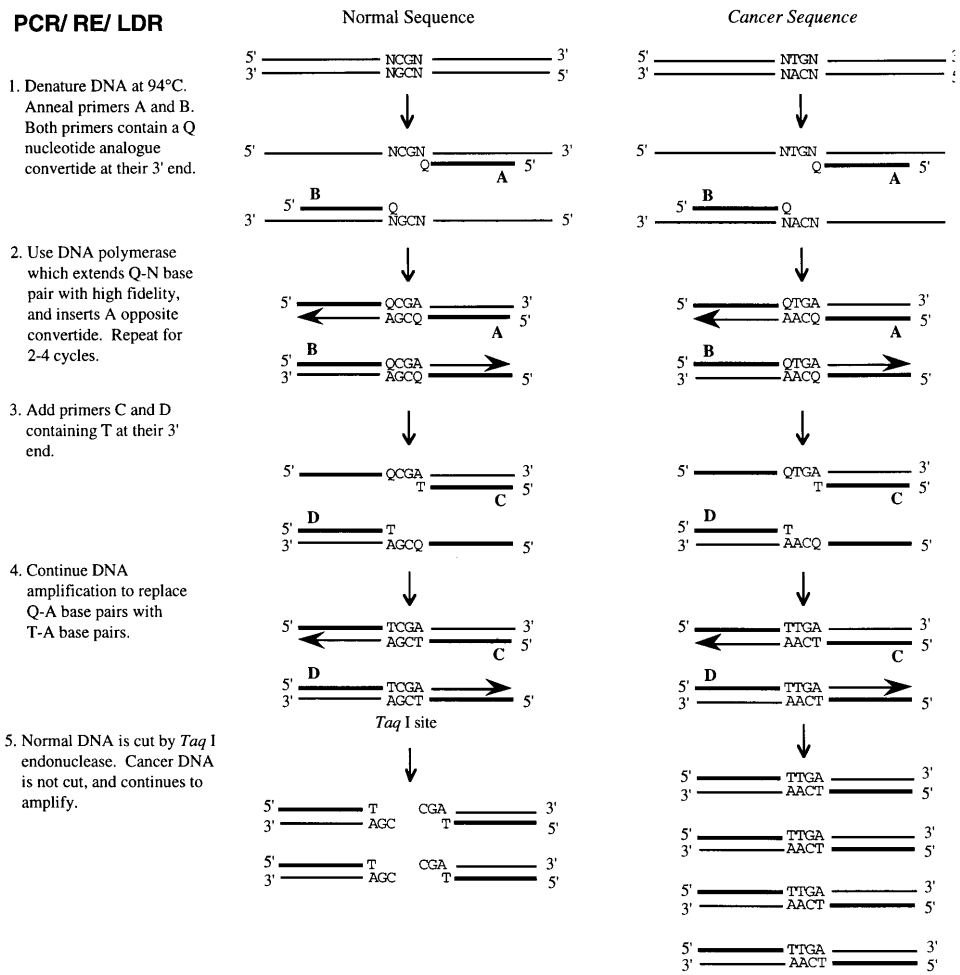

Figure 1. Preferential amplification of mutant DNA by PCR/RE/LDR. Preconversion (steps 1 and 2) using a nucleotide analog (Q) in mutant and wild-type DNA allows more efficient introduction of a restriction site at a CpG dinucleotide. The addition of natural base primers (step 3) completes conversion, replacing the nucleotide analog with the desired normal base. Finally, wild-type DNA is digested with the RE for which the site was created (step 5). Mutant DNA is subsequently reamplified.

straightforward, 3′ mismatch extension has proven to be difficult (7,17–19). The introduction of interrupted palindromic restriction sites has been more successful using internal mismatch primers spanning one half-site through the intervening bases up to the other half-site (20,21). Several perfectly matched bases stabilize the 3′-end of the mismatch primer. This approach may be used only if the second half-site is present naturally in wild-type DNA.

REs recognizing interrupted palindromes are less abundant than endonucleases recognizing contiguous four and six base sites. Multiple base changes would often be required to introduce an interrupted palindrome restriction site to identify mutations at any base. In order to develop a general approach, this work introduces contiguous four base Type II restriction sites in wild-type sequences containing a central CpG dinucleotide by altering one base on each side of the CpG. CpG dinucleotides are frequent sites of mutation; for example, ~40% of the mutations observed in the p53 tumor suppressor gene fall into this category (9). We tested conversion of the *Msp*I site (CCGG) to a *Taq*I site (TCGA) in p53 exon 7 at codon 248 by PCR using a set of 3′ mismatch primers which alter the outer bases of the *Msp*I site (Fig. 1). Preconversion using 3′ nucleotide analog (Q) primers may be performed prior to adding natural base primers to avoid extension from primers with 3′ natural base mismatches. Conversion was performed with and without preconversion to determine whether preconversion facilitates conversion. This process introduces a restriction site that allows digestion of wild-type DNA with an RE and leads to preferential amplification of the undigested mutant DNA. Mutant products are quantified by LDR. This PCR/RE/LDR assay is a high sensitivity variant of PCR/LDR.

Techniques, such as PCR/RE/LDR, that rely on mutant enrichment require optimization of reaction conditions in order to minimize PCR errors. These errors would be indistinguishable from mutations originally present in clinical samples. Standard PCR buffers contain Tris, however its pK_a is strongly dependent on temperature. A PCR reaction containing Tris pH 8.3 (measured at 23°C) is ~pH 7 near 65°C (the extension temperature) and drops to ~pH 6 near 95°C (the template melting temperature). PCR error can result from template degradation and polymerase misincorporation. Template degradation occurs during periods of high temperature and low pH in each PCR cycle and limits product size in 'long' PCR (22–24). Raising the buffer pH in long PCR (using Tris 9.1) reduces the amount of template cleavage and increases PCR efficiency (22). Although the efficiency of long PCR increases with higher pH, the level of mutations within these PCR products may also increase since high pH decreases the fidelity of *Taq* and *Pfu* polymerases (25–27). Use of alternative PCR buffers

with lower $|\Delta pK_a|$ can improve polymerase fidelity and still reduce template damage by maintaining more neutral pH over a wider temperature range (26,28). The addition of glycerol or formamide may reduce mutations arising from template damage during PCR cycling and may help avoid misextension from mispaired primers (2,29).

We tested proofreading and non-proofreading thermostable DNA polymerases in several PCR buffers formulated on the basis of an analysis of known sources of PCR error. Our test PCR buffers contained low $|\Delta pK_a|$ buffering compounds and different salts. We show that mismatch extension is prone to error far beyond that produced by polymerase error or template degradation during PCR. Directly probing PCR/RE products using LDR allows identification of specific mutations and quantification of each mutation produced. PCR fidelities using buffers with and without formamide were compared. The use of low $|\Delta pK_a|$ buffers with formamide greatly reduces background PCR error. Preconversion with 3′ nucleotide analog primers significantly improved the fidelity of base conversion to introduce a new restriction site.

MATERIALS AND METHODS

Oligonucleotide synthesis

Oligonucleotides were synthesized at the 0.2 µmol scale by cyanoethyl phosphoramidite chemistry on an Applied Biosystems 394 DNA synthesizer. Standard 500 Å CPG columns and reagents (Applied Biosystems) were used with the following exceptions. Oligonucleotides 50 bases in length were synthesized using wide pore 1000 Å CPG columns (Applied Biosystems). Oligonucleotides with fluorescent dye FAM at the 5′-terminus were synthesized using FAM phosphoramidite (Applied Biosystems) with a 15 min coupling step. Oligonucleotides with 5′ phosphate were synthesized using phosphorylation reagent (Glen Research) with a 15 min coupling step. Oligonucleotides with 3′ blocking group were synthesized using 3′-Spacer CPG columns (Glen Research). Oligonucleotides with the 3′ nucleotide analog 6-(2′-deoxy-β-D-ribofuranosyl)-6*H*,8*H*-3,4-dihydro-pyrimido-[4,5-*c*]-[1,2]oxazine-7-one (Q_6) were synthesized using dP-CPG (Glen Research).

PCR polymerases and buffers

The polymerases used were Ampli*Taq* (Applied Biosystems), Vent and Vent(exo–) (New England Biolabs) and Expand polymerase mix (*Taq* and *Pwo*I polymerase mixture, in the Expand High Fidelity kit; Boehringer Mannheim). The commercially available PCR buffers used were supplied in the Ampli*Taq* and Expand kits. Tris pH 9.1 (pH values were measured using 1 M stock solutions at 23°C), tricine pH 8.7, EPPS [*N*-(2-hydroxyethyl)piperazine-*N*′-3-propanesulfonic acid] pH 8.4 and citrate pH 7.6 (Sigma) were used for alternative PCR buffers. Unless otherwise noted, each 20 µl reaction contained 20 mM Tris, tricine or citrate, 200 µg/ml bovine serum albumin, 2.5 mM $MgCl_2$, 200 µM dNTP (each) and either 16 mM $(NH_4)_2SO_4$ or 50 mM potassium acetate. Formamide at 10% concentration was used as indicated (see Enzyme/buffer notation). PCR buffers were made as 10× stocks requiring the addition of formamide as needed, dNTPs and the oligonucleotide primers and template DNA.

Enzyme/buffer notation

Test PCR buffers are named to indicate the presence of one or more components: Tris/potassium acetate, buffer A; Tris/ammonium sulfate, buffer B; tricine/ammonium sulfate, buffer D; EPPS/potassium sulfate, buffer E; EPPS/ammonium sulfate, buffer F; citrate/ammonium sulfate, buffer G. Component concentrations are described above.

Amplification of p53 exon 7 from genomic DNA

Part of p53 exon 7 surrounding codon 248 was amplified. The upstream primer (5′-GCCTCATCTTGGGCCTGTGTTATC-3′) hybridized within the preceding intron and the downstream primer (5′-GTGGATGGGTAGTAGTATGGAAGAAATC-3′) hybridized within exon 7. All PCR, RE digestion and ligation steps described throughout were performed using a GeneAmp PCR System 2400 (Perkin Elmer). Several buffers and enzymes were used (see PCR polymerases and buffers) as indicated. The p53 exon 7 amplification from genomic DNA was performed starting with a 20 µl reaction mixture containing 50 ng of DNA, 2.5 mM each dNTP and 12.5 pmol of each primer in 1× buffer without polymerase. The reaction mixture was covered with paraffin oil and preincubated for at least 1.5 min at 94°C in order to perform hot start by adding 1 µl of polymerase diluted in 1× buffer to introduce the required units of polymerase. The exon 7 segment was amplified for 40 cycles of 94°C for 15 s, 65°C for 2 min, with an additional 5 min at 65°C at the end of the last cycle. PCR amplifications departing from this procedure were performed as indicated.

PCR/RE/LDR

Fidelity assay. Templates were amplified with conversion primer pairs bracketing the central two base pairs of the *Msp*I site (CCGG) at codon 248 (Fig. 2B). Tubes were prepared containing 10 fmol/reaction of either PCR-amplified p53 exon 7 or wild-type synthetic duplex template, PCR buffer and primers. In parallel reactions, a synthetic 50 bp duplex marker template (MK), with the sequence CGGG replacing the *Msp*I site at codon 248, was added at 10^{-3}, 10^{-4}, 10^{-5} and 0 molar ratio to wild-type template. Reactions were preincubated for at least 1.5 min at 94°C with all components present in CiNF buffer except Vent(exo–) polymerase. A ‘hot start’ was performed by adding 1 µl of polymerase at 94°C. When preconversion was performed, two cycles of 94°C for 15 s, 55°C for 1 min, 60°C for 1 min were executed with 500 fmol each of the primers p53-248Q_6 and p53-248Q_6R. Afterwards, 1 pmol of p53Taq248T and p53Taq248TR primers were added. When preconversion was not performed, the reactions contained 1 pmol each of the primers p53Taq248T and p53Taq248TR or the control primers p53Msp248C and p53Msp248CR. After reactions with and without preconversion were performed, conversion PCR was carried out as follows: 5 cycles of 94°C for 15 s, 55 + 1°C/cycle for 1 min (temperature ramp), 60°C for 1 min; then 20 cycles of 94°C for 15 s, 60°C for 2 min; a final 60°C for 5 min extension. After three cycles of the temperature ramp 10 pmol of long zipcode conversion primers (p53zip248T and p53zip248TR or p53zip248C and p53zip248CR) were added. After conversion, the wild-type DNA was digested periodically during 20 cycles of ‘zipcode’ PCR (described below). Polymerase was inactivated by freezing and thawing twice. Finally, LDR was performed to detect the

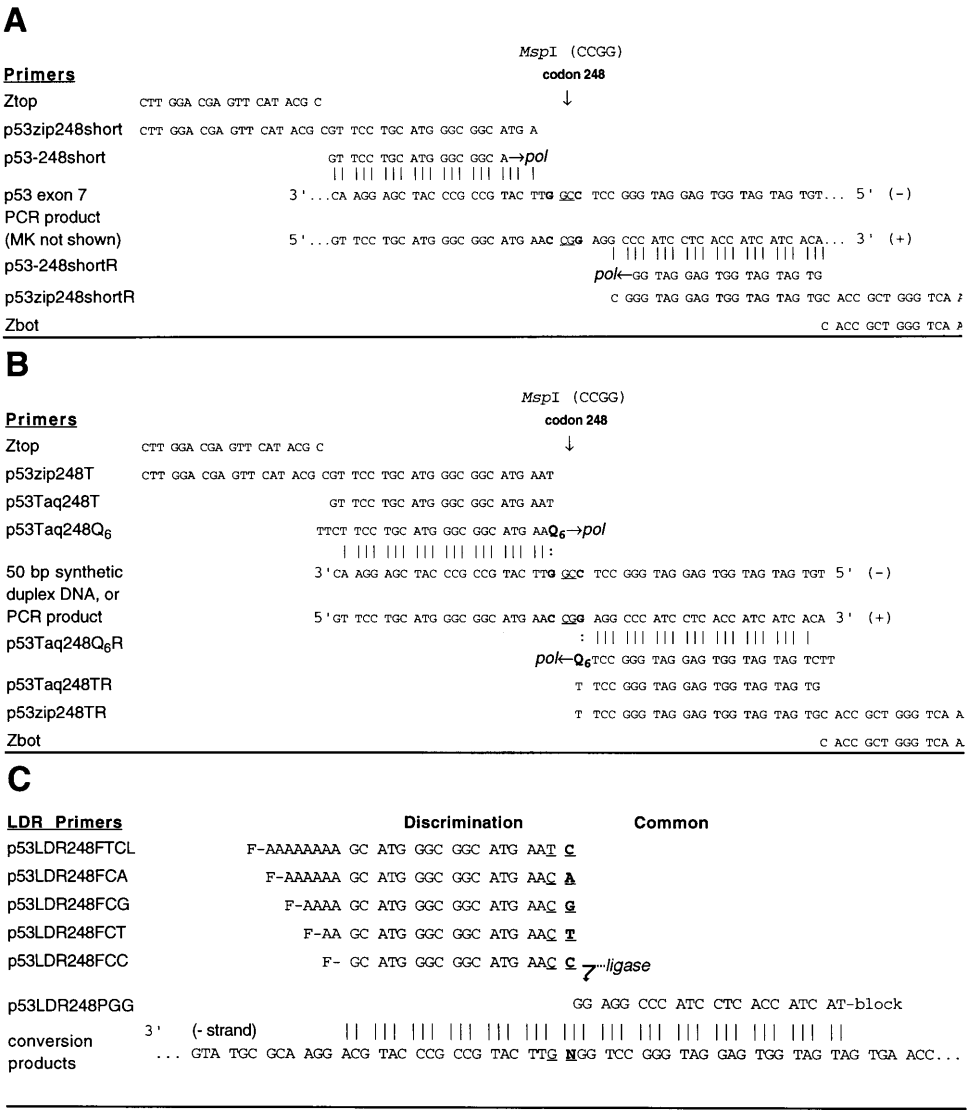

Figure 2. Primers used in PCR/RE/LDR. Complimentary (– strand) sequences are shown in reverse orientation (3′→5′), in particular reverse strand primers (having names ending in R). (**A**) PCR fidelity assay. A synthetic 50 bp duplex marker template (MK) and wild-type p53 exon 7 PCR product are mixed at known ratios in parallel reactions. Perfect match primers p53-248short and p53-248shortR amplify the wild-type CCGG and marker CGGG. Then, longer zipcode-containing primers p53zip248short and p53zip248shortR were added. Finally, wild-type was repeatedly digested and reamplified with zipcode primers (Ztop and Zbot). (**B**) Preconversion was performed using primers containing 3′ convertide, e.g. p53-248Q_6. Conversion of the *Msp*I site to a *Taq*I site with or without preconversion was performed using 3′ natural base primers p53zip248T and p53zip248TR. Long primers were added as above and conversion products further amplified. Wild-type products were digested with the RE appropriate for the new site. Mutant products were preferentially amplified with zipcode primers. (**C**) LDR primer sets were designed to query the template sequence around the point of ligation. Perfectly hybridized upstream and downstream LDR primers with no overlap or gap are preferentially ligated with high specificity. Discrimination primers have different length 5′ tails to allow specific product separation on an acrylamide gel. Primers shown were used for identification of mutations occurring in the second base of the *Msp*I site (no conversion). An extra primer (p53LDR248FTCL) was used to compare C→T transitions at the first base and second base of the *Msp*I site. A comparable set of discrimination and common primers, used to identify mutations at the second base of the *Taq*I site in conversion products, had a T at the 3′ penultimate base in the discrimination primers and A at the 5′ penultimate base in the common primer.

conversion products without contribution from the original template (except in non-conversion control reactions).

'Zipcode' PCR. Wild-type sequences or wild-type conversion products were removed by restriction digestion. The appropriate RE was added to the reaction tube and supplemented with additional $MgCl_2$ as required to allow efficient digestion. *Msp*I digestion was performed at 37°C for 15 min using no additional $MgCl_2$, except when using citrate buffer. *Taq*I digestion was performed at 65°C for 30 min at 6 mM Mg^{2+} by adding 1 µl of enzyme diluted in 140 mM $MgCl_2$. The undigested conversion products were reamplified from 1 µl of a 10× dilution added to a 20 µl PCR reaction containing 10 pmol of the 'zipcode' primers Ztop and Zbot (Fig. 2B). These zipcode primers each contain a DNA sequence that is not similar in sequence to any genomic sequences present in the sample, thus only the products of previous PCR using primers containing the zipcode sequences will be efficiently amplified. Conversion products were amplified using Expand polymerase mix and buffer (see PCR polymerases and buffers). After an initial RE digest, zipcode PCR reamplification followed by redigestion was performed as follows: reactions were

preincubated at 94°C for at least 1.5 min then initiated with a hot start by adding 0.1 µl of RE-digested sample (1 µl of a 10× dilution) to a 20 µl reaction; 10 cycles of 94°C for 15 s, 65°C for 2 min. Zipcode PCR amplification products were redigested as described above.

Ligase detection reaction. Ligase detection reactions were performed in standard LDR buffer (25 mM Tris pH 7.6, 12 mM $MgCl_2$, 65 µg/ml bovine serum albumin, 100 mM KCl and 10 mM DTT). Each 20 µl reaction contained ~500 fmol of dsDNA (1 µl of PCR sample), 500 fmol of each discrimination primer and 750 fmol of common primer (Fig. 2C). Sets of discrimination and common primers were synthesized to detect the expected conversion products, i.e. converted to C**N**GG or T**N**GA at the *Msp*I position. The common primer was synthesized using 3′-Spacer C3 CPG columns and the 5′-end was phosphorylated on the column using phosphorylation reagent (see Oligonucleotide synthesis). Discrimination primers of each set varied at the 3′-terminal base to query the base in that location, i.e. the second base of the *Msp*I position. Discrimination primers had 5′ tails of different length and a FAM label for fluorescence detection. The tail size identified the primer and allowed physical separation of different LDR products on an acrylamide gel.

The LDR reaction was preincubated for 1.5 min at 94°C prior to the addition of 5 nmol *Tth* ligase enzyme under a layer of mineral oil. We used 10 LDR cycles of 94°C for 15 s, 65°C for 2 min. The reactions were then held at 94°C until cold quenched on ice and stored at –70°C. The LDR products were separated on 10% acrylamide gels containing 7 M urea with 0.6× TBE (1× TBE contains 90 mM Tris base, 90 mM borate, 2 mM EDTA) used in the gel and for the running buffer. Data were collected using an ABI 373 automated DNA sequencer and Applied Biosystems Genescan 672 software (GS Collection and GS Analysis).

Image processing

Raw gel pictures were produced by the ABI GS Analysis software. Dye-specific pictures were opened in Adobe Photoshop 3.0, cropped, resized and converted to grayscale. The grayscale images were opened in NIH Image 1.59, inverted and 1D vertical background was subtracted. Optionally, NIH Image could render a three-dimensional plot from a corrected two-dimensional picture. Background corrected pictures were reinverted and rendered in pseudocolor by Photoshop by replacing the color table to make subtle intensity differences easier to compare. Except for color replacement, only linear image processing was performed in order to preserve relative intensities.

RESULTS AND DISCUSSION

We developed PCR/RE/LDR to detect and identify low abundance mutations occurring within the *Msp*I site (CCGG) at codon 248 in the p53 gene (Fig. 1). An initial PCR amplifies exon 7 from genomic DNA. This product serves as the template for a second PCR that amplifies the central CpG dinucleotide in the *Msp*I site. To generate a restriction site in sequence lacking a pre-existing site, mismatch primers are used to alter one or more bases flanking the CpG dinucleotide. This results in a conversion PCR that creates a restriction site (NCGN→TCGA *Taq*I site, for example). In a generalized method for introducing contiguous Type II restriction sites, conversion PCR primers by necessity have 3′-terminal mismatches. To avoid unfavorable natural base mismatches that may result in insertion of an erroneous base at the next site (18,30), preconversion with 3′ nucleotide analog primers is performed. However, extension with 3′ analog primers produces a pool of degenerate products (19). Thus, after this preconversion step, natural base primers are used to selectively amplify the desired products.

We assessed mismatch conversion error relative to PCR error by performing parallel non-conversion control reactions and true conversion reactions with and without preconversion. Non-conversion reaction products retained the *Msp*I site (CCGG), while conversion introduces a *Taq*I site (TCGA). All PCR/RE/LDR steps were performed under similar conditions, varying only the primers and RE (*Msp*I or *Taq*I). In both cases, non-cleavable DNA is preferentially amplified. When wild-type DNA is selectively removed by digestion, it is necessary to determine the proportion of DNA with incorrect sequence produced relative to the initial quantity of DNA in the sample, which is nearly 100% wild-type. Parallel reactions were performed in which known fractions of MK DNA were present. The MK DNA contained a single base change in the *Msp*I site (C<u>G</u>GG), rendering it uncleavable by *Msp*I. C→G transversions are unlikely to occur through polymerase error. The MK standard curve allows quantification of mutations detected by LDR. PCR conditions were tested to minimize PCR error (observed in the non-conversion reactions) and mismatch extension errors (additional errors observed in the conversion reactions).

We tested various proofreading and non-proofreading polymerases, as different polymerase properties are required during target amplification from genomic DNA, conversion and reamplification steps in PCR/RE/LDR. Since it is essential throughout PCR/RE/LDR to minimize any alteration of the bases assayed by LDR, proofreading polymerases might seem the logical choice for maintaining the highest fidelity (31), however, they may interfere with conversion by mismatch primer extension. Hence, PCR conditions must be found which maximize the fidelity of non-proofreading polymerases (32).

Initially, we used PCR/RE/LDR as a high sensitivity assay to determine PCR conditions that maintain the highest fidelity throughout the procedure. Two main sources of error were expected: (i) polymerase misincorporation; (ii) DNA template degradation. Raising the PCR buffer pH improves long PCR, probably by decreasing depurination which leads to strand cleavage (22–24). While higher pH may decrease template damage, higher pH is also known to adversely affect polymerase fidelity (25–27). Therefore, we tested tricine, EPPS and citrate buffers which have pK_a values in the range 7–8 and $|\Delta pK_a|$ lower than Tris (see Materials and Methods). Tris cannot meet the dual constraints of neutral pH at high temperature to maintain template integrity and lower pH at the extension temperature to maintain polymerase fidelity, although most PCR fidelity and long PCR studies use Tris. Some investigators have explored the use of alternative buffers with lower $|\Delta pK_a|$ (25,26,28). Buffer-specific effects on PCR were tested by making PCR buffers containing identical components except for the buffering compound. We tested salt effects by making one set of test PCR buffers with ammonium sulfate and another with potassium acetate. The ΔpK_a of each buffer was determined in pure solution and in 1× PCR buffer mixtures (data not shown). Our results agreed with published ΔpK_a values of pure buffers (33,34) corrected by a small constant (0.005 pH units/°C), possibly due to a temperature

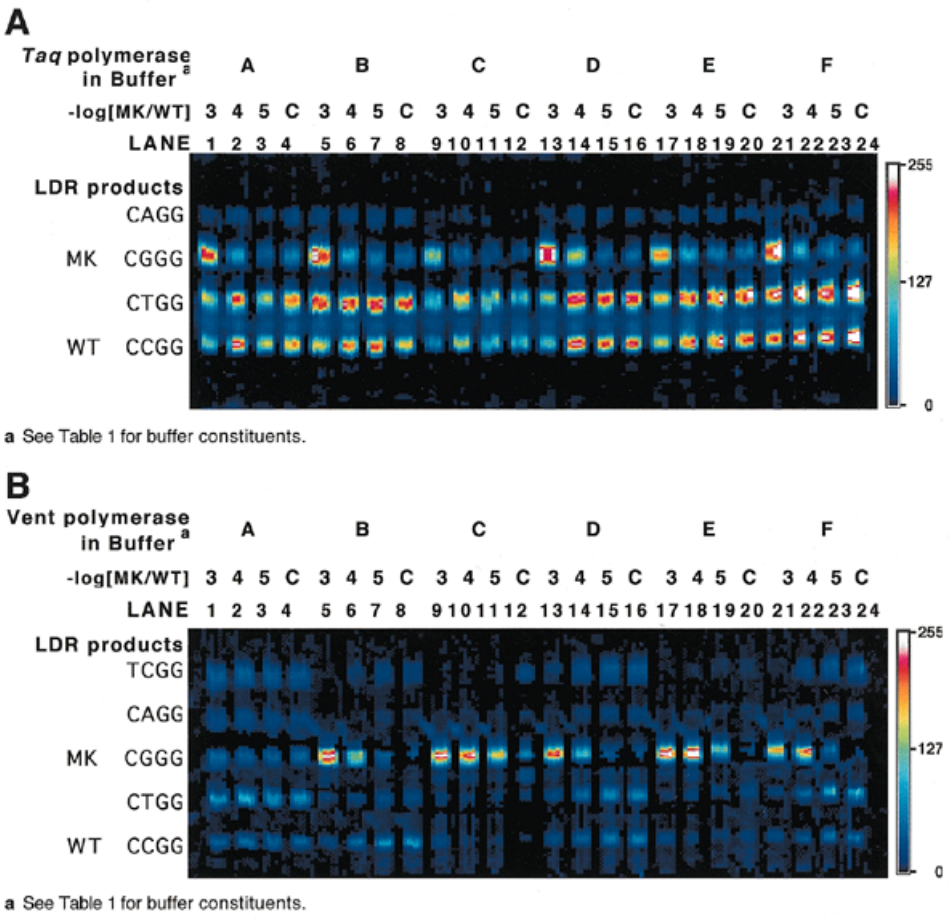

Figure 3. Buffer- and enzyme-dependent PCR errors detected by PCR/RE/LDR. The indicated polymerase/buffer combinations were used to amplify p53 exon 7 from genomic DNA. The same buffers were used in reactions with perfect match primers to reamplify the *Msp*I site. (**A**) *Taq* polymerase used in various test PCR buffers. (**B**) Vent polymerase used in various test buffers. Vent polymerase did not amplify p53 exon 7 from genomic DNA in TsK buffer. In this case only, two different enzyme/buffer sets were used for preamplification and 'conversion' (not actual conversion, since perfect match primers were used). The Ampli*Taq*/TsK exon 7 genomic DNA PCR product was substituted in the Vent/TsK reamplification. C indicates no MK was added (control reaction).

dependence of the pH probe itself. We adjusted the pH of our test PCR buffers to produce approximately neutral pH at 65°C. However, the 1× PCR buffers had somewhat different ΔpK_a values compared to the pure buffers; for example, 1× TsN had $\Delta pK_a = -0.033/°C$ versus $-0.030/°C$ for 100 mM Tris and 1× TcK had $\Delta pK_a = -0.022/°C$ versus $-0.025/°C$ for 100 mM tricine.

Test PCR buffers containing Tris, tricine or EPPS were used to test PCR fidelity with no conversion of the *Msp*I site (CCGG) at codon 248 of p53 (Fig. 3). Our objective in this experiment was to test the error rate of PCR using various buffers and polymerase enzymes. Since introduced errors create template that cannot be cleaved by the selected restriction enzyme, false positives accumulate as this error template continues to amplify alongside true mutant DNA. This experiment establishes the conditions necessary to achieve amplification while minimizing error. The same polymerase and buffer set was used in both preamplification of p53 exon 7 from genomic DNA and in the 'conversion' step. As mentioned, the 'conversion' step maintains the *Msp*I site by using perfectly matched primers whose 3′-ends terminate on the C and G bases flanking the central CpG. After an initial *Msp*I digest, template and amplification products were periodically redigested every 10 cycles during reamplification to remove WT sequence. Synthetic marker mutant MK with the sequence CGGG was present in these reactions at 10^{-3}, 10^{-4}, 10^{-5} or 0 ratio to wild-type (WT). MK will not be cleaved by *Msp*I restriction digestion, but will amplify with each PCR cycle to provide an internal control to measure product quantitities (see below). The MK product will also maintain its sequence, as the perfect match primers in the conversion step will again terminate on the C and G bases flanking the central GG. Error products resulting from MK PCR will in general lack *Msp*I sites and will be indistinguishable from regular MK template. If an *Msp*I site is accidentally created, the product will be destroyed by digestion.

For each buffer, LDR detected MK products in each of the four parallel reactions, with the 0 MK control indicating the background level of CGGG error produced. The intensities of other error products detected by LDR were compared to MK to estimate the fraction of each error product generated. Ampli*Taq* generated few transversions (C→G or C→A), but a large amount of C→T transition was observed (Fig. 3A). Vent generated much less of the C→T transition compared to Ampli*Taq* (Fig. 3B). Ampli*Taq* showed little dependence on the presence of potassium acetate in buffers A, C and E (Fig. 3A, lanes 1–4, 9–12 and 17–20) versus ammonium sulfate in buffers B, D and F (lanes 5–8, 13–16 and

21–24). Vent polymerase amplified template more efficiently in Tris/ammonium sulfate buffer B than Tris/potassium acetate buffer A (Fig. 3B, lanes 1–4 versus lanes 5–8), as described previously (22,31,35,36). However, Vent exhibited improved fidelity in tricine/potassium acetate buffer C (lanes 9–12) and EPPS buffer E (lanes 17–20) compared to tricine/ammonium sulfate buffer D (lanes 13–16) and EPPS/ammonium sulfate buffer F (lanes 21–24).

The relative fidelities of the different polymerase–buffer combinations may be described by their 'sensitivity' expressed as the $-\log_{10}$ of the ratio of MK to WT initially present. The C→T error for Ampli*Taq* amplification in Tris/potassium acetate buffer A can be taken as an example. If the signal for the CTGG error product (Fig. 3A, lane 2) is compared to the MK CGGG signal (Fig. 3A, lanes 1–3), the intensity of the signal most resembles the 10^{-3} MK:WT dilution (Fig. 3A, lane 1). Thus, the C→T error rate is 10^{-3}; the sensitivity is 3, since $-\log[MK/WT] = -\log[10^{-3}] = 3$. From this it can be seen that the higher the sensitivity, the lower the error rate. Reactions with higher sensitivities for each mutation had the highest overall fidelity (results summarized in Table 1). Many of the Vent reactions had sensitivities of 1 in 10^5 for every mutation (Fig. 3B), while the Ampli*Taq* reactions had sensitivities of 1 in 10^3 (Fig. 3A). Sensitivity indicates the usefulness of the assay rather than the error rate of the polymerase. Error (ER) per base per cycle may be estimated from the fraction (F) of all mutations occurring at one base which accumulated over 65 cycles (D) before digestion. For our purposes, the number of cycles is an estimate of the number of duplications, since multiple non-saturating PCRs were performed. From ER = F/D, Vent polymerase had an error rate of $<1 \times 10^{-7}$/base/cycle in tricine/potassium acetate buffer C, $\sim 2 \times 10^{-7}$/base/cycle in tricine/ammonium sulfate buffer D and 2×10^{-6}/base/cycle in TcN buffer. We observed an error rate of 2×10^{-5}/base/cycle mainly due to the C→T transition for Ampli*Taq* in Tris/potassium acetate buffer A. Elimination of this artifact could improve Ampli*Taq* fidelity by more than 10-fold. Others have used cloning and screening methods to estimate polymerase error (25,27,36,37) and denaturing gradient gel electrophoresis (DGGE) has also been used (31,32,35,38). However, these methods do not directly measure mutated DNA and do not detect all mutations. By cloning and DGGE methods, Vent polymerase has an error rate estimated as from 0.3 to 4×10^{-5}/base/cycle (27,35,38). The error rate of *Taq* polymerase has been estimated as from 0.8 to 9×10^{-5}/base/cycle (25,27,28), comparable to the error rate we observed for Ampli*Taq* in TsK buffer. Of the thermostable polymerases, *Pfu* has the lowest reported error rate estimated as from 0.7 to 1.6×10^{-6}/base/cycle (27,39,40). *Pfu* polymerase may also exhibit improved fidelity in tricine or other low $|\Delta pK_a|$ buffers.

While high fidelity proofreading enzymes appeared to improve amplification from genomic DNA, proofreading still must be avoided in the conversion step. We tested different high fidelity genomic amplification conditions in combination with fixed conversion conditions. Genomic amplification was performed with either Vent(exo–) in citrate/ammonium sulfate buffer G or Vent(exo–) in citrate/ammonium sulfate buffer G with 10% formamide (Table 1). Non-conversion primers were used with Vent(exo–) to optimize PCR fidelity in anticipation of conversion by mismatch primer extension. Our highest fidelity conditions were as follows. Genomic amplifications with Vent/buffer G were initiated by spiking genomic amplification product from Expand/buffer C with 10% formamide reactions after three cycles. These Vent/buffer G reactions required 4 mM Mg^{2+} and PCR primers, but no additional genomic DNA was provided (see Table 1 for observed error rate with other conditions tested).

Table 1. Comparison of fidelity using proofreading and non-proofreading polymerases in different buffers for PCR to amplify the target sequence from genomic DNA and for conversion PCR

Polymerase enzymes (1) Genomic DNA PCR→(2) conversion	Buffer (1)→(2)	Limiting error	Error rate Total	Per cycle[a]
Taq→*Taq*	A→A	C→T	10^{-3}	2×10^{-5}
Taq→*Taq*	B→B	C→T	10^{-3}	2×10^{-5}
Taq→*Taq*	C→C	C→T	10^{-3}	2×10^{-5}
Taq→*Taq*	D→D	C→T	10^{-3}	2×10^{-5}
Taq→*Taq*	E→E	C→T	10^{-3}	2×10^{-5}
Taq→*Taq*	F→F	C→T	10^{-3}	2×10^{-5}
Taq[b]→Vent	A→A	C→T	$>10^{-3}$	$>2 \times 10^{-5}$
Vent→Vent	B→B	C→T	10^{-5}	2×10^{-7}
Vent→Vent	C→C	C→T	$<10^{-5}$	$<2 \times 10^{-7}$
Vent→Vent	D→D	C→T	10^{-4}	2×10^{-6}
Vent→Vent	E→E	C→T	$<10^{-5}$	$<2 \times 10^{-7}$
Vent→Vent	F→F	C→T	10^{-5}	2×10^{-7}
Vent(exo–)→Vent	C→G	C→T	10^{-4}	2×10^{-6}
Vent(exo–)→Vent(exo–)	C→G	C→T	10^{-3}	2×10^{-5}
Vent→Vent(exo–)	C→G	C→T	10^{-4}	2×10^{-6}
Vent→Vent(exo–)	C→G(f)	C→T	10^{-5}	2×10^{-7}
Vent[c]→Vent(exo–)	G (4)→G	C→T	10^{-4}	2×10^{-6}
Vent→Vent(exo–)	G (8)→G(f)	C→T	10^{-5}	2×10^{-7}
Vent[c]→Vent(exo–)	G (4)→G(f)	C→T	$<10^{-5}$	$<2 \times 10^{-7}$

[a]Based on a minimum of 50 total cycles, i.e. observed error/50.
[b]No Vent PCR product from genomic DNA. *Taq* amplified product used for Vent conversion PCR.
[c]Template added by taking 1 µl after the third PCR cycle from a parallel genomic DNA amplification using Expand polymerase mix in buffer C.
Taq and Vent polymerases were initially tested using one buffer for genomic amplification and conversion. During the conversion step, only non-conversion of the *Msp*I site near p53 codon 248 was performed using short perfect match primers (Fig. 2A) to determine the background level of polymerase error. LDR quantified *Msp*I site mutations at the second position (CCGG→CNGG). Fidelity was compared in parallel reactions using proofreading and non-proofreading polymerases in genomic amplification and conversion. Expand polymerase mix, *Taq* with proofreading *Pfu* polymerase added, was used to initiate target amplification from genomic DNA for subsequent high fidelity Vent polymerase PCR. Vent polymerase was substituted with non-proofreading Vent(exo–) to determine whether proofreading was required and also in the conversion step where proofreading is not permitted. The effect of 10% formamide in the conversion PCR buffer was also tested. All buffers contained 200 µg/ml bovine serum albumin, 2.5 mM $MgCl_2$ and 200 µM dNTP (each). Specific components were: A (TsK), 20 mM Tris pH 9.1, 50 mM potassium acetate (standard *Taq* polymerase buffer); B (TsN), 20 mM Tris pH 9.1, 16 mM ammonium sulfate (standard Vent polymerase buffer); C (TcK), 20 mM tricine pH 8.7, 50 mM potassium acetate; D (TcN), 20 mM tricine pH 8.7, 16 mM ammonium sulfate; E (EpK), 20 mM EPPS pH 8.4, 50 mM potassium acetate; F (EpN), 20 mM EPPS pH 8.4, 16 mM ammonium sulfate; G (CiN), 20 mM citrate pH 7.6, 16 mM ammonium sulfate. (f), presence of 10% formamide; (4), increase to 4 mM $MgCl_2$; (8); increase to 8 mM $MgCl_2$.

We found PCR conditions for each step in PCR/RE/LDR that maintain high fidelity when no mismatch conversion was performed. With known high fidelity PCR conditions, we next tested conversion by changing the p53 codon 248 *Msp*I site

(CCGG) into a *Taq*I site (TCGA). MK (CGGG) was added as before in parallel reactions to measure fidelity relative to the initial wild-type DNA present. High fidelity PCR was performed as described above and some (but not all) reactions were subjected to preconversion. Preconversion was performed using primers containing the degenerate pyrimidine nucleotide analog Q_6 at their 3′-ends (Fig. 2B). The final conversion was accomplished using natural base 3′ T mismatch primers. Products were detected using LDR to interrogate the second base position in the *Msp*I, *Taq*I and MK sequence: CNGG or TNGA. Fidelity for conversion with and without preconversion was compared to a non-conversion control. Successful conversion will change the *Msp*I site (CCGG) into a *Taq*I site (TCGA); MK will also be converted from CGGG to TGGA. However, the main issue of conversion success is the maintenance of the central bases in all cases: CpG for *Taq*I conversions and GpG for MK. Figure 4 shows the results of conversion. Lanes 1–4 (C:G) are non-converted reactions that were digested with *Msp*I; lanes 5–8 (Q_6:G) are preconverted/converted reactions that were digested with *Taq*I; lanes 9–12 (T:G) are converted reactions lacking preconversion that were digested with *Taq*I. PCR/RE/LDR with no conversion was sensitive to better than 1 in 10^4 using the previously determined best conditions for preamplification and conversion (Fig. 4, C:G, lanes 1–4). PCR/RE/LDR with conversion of the *Msp*I site to a *Taq*I site by T mismatch primers was apparently very successful at first glance (T:G, lanes 9–12). As would be expected for successful conversion, no *Msp*I product can be detected in the _CG_ region of the figure, hence, it appears that the site was converted to *Taq*I and then digested. However, although a very large fraction of MK (CGGG) is observed in the reactions with added MK (T:G, lanes 9–11), the same large fraction is also observed in the 0 MK control lane (T:G, lane 12). Thus, the entire quantity of MK is an artifact produced by mismatch extension of the natural base T primers. This event would convert the second position C in the *Msp*I site to a G during extension, mimicking the internal sequence of MK (CCGG→TGGA). Preconversion with Q_6 primers eliminates the MK artifact (Q_6:G, lanes 5–8).

The greater amount of WT present in non-converted samples (Fig. 4, lanes 1–4) compared to Q_6 converted samples (lanes 5–8) may be due to inhibition of *Msp*I digestion by formamide. Formamide apparently inhibits *Msp*I digestion as evidenced by the presence of strong wild-type LDR bands (WT) in the non-conversion control (C:G lanes), which are not present after *Taq*I digestion of the converted sequence (Q_6:G and T:G lanes).

The low amount of MK product seen in the Q_6 10^{-4} and 10^{-5} MK lanes (Fig. 4, Q_6:G, lanes 6 and 7) compared to the respective non-conversion control reactions (C:G, lanes 2 and 3) may be due to low efficiency of MK conversion. The production of a *Taq*I site actually requires two conversions, one on each side of the central CpG dinucleotide. Lowering the concentration of MK 10-fold may reduce MK conversion far more than 10-fold. Thus, with only one side of the MK sequence converted in a large amount of its product, one half of the LDR primers will be unable to properly hybridize to this sequence and ligation will not occur. LDR detection will only reveal the lesser quantity of fully converted template. Nevertheless, the amount of MK product is greater than the control in these two lanes (compare Fig. 4 lane 8 to lanes 6 and 7). While formamide may reduce conversion efficiency, conversion fidelity is greatly improved.

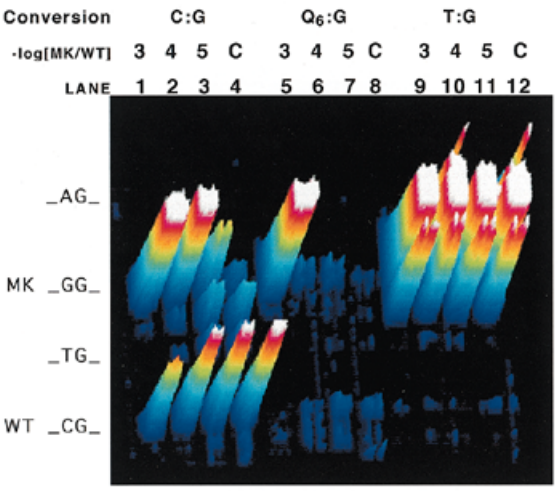

Figure 4. Comparison of conversion fidelity. The relative intensities of conversion reaction products is indicated by color and the height of each peak in a 3-dimensional plot (Materials and Methods). Marker (MK) DNA (with CGGG replacing the *Msp*I site) was added at known ratios to wild-type (WT) in parallel reactions. The –log(MK:WT) indicates relative fraction of MK present, e.g. –log(MK:WT) = 3 means the ratio of MK to WT was 1:1000. C indicates no MK was added (control reaction). Non-conversion control reactions (C:G) were performed using perfect match 3′ C primers. Conversion of the *Msp*I site (CCGG) to a *Taq*I site (TCGA) was performed using natural base 3′ T primers with and without preconversion using 3′ Q_6 nucleotide analog primers (Q6:G and T:G reactions, respectively). LDR products from *Msp*I non-conversion contain CNGG and products from *Taq*I conversion contain TNGA, but only the the central bases (second and third bases) are indicated as _NG_. The LDR products were designed to separate on acrylamide gels by two base differences in size. Some undetermined bands of intermediate size were also observed. Lanes 1–4 were digested with *Msp*I, while lanes 5–12 were digested with *Taq*I during PCR/RE/LDR.

Newton *et al*. found that C·T, A·A and T·T mismatches were all far more difficult to extend with *Taq* polymerase than purine·pyrimidine mismatches (7). These results reflect PCR efficiency of extension rather than fidelity. Others have observed low fidelity in extending natural base mismatches (18,30). Use of a nucleotide analog with structural similarities to multiple bases could potentially be used to allow polymerase extension (reading) from the analog when paired with different bases and insertion of different bases opposite the analog (writing). For the purposes of this assay, the efficiency of the process need not necessarily be high. However, successful conversion requires high PCR fidelity to ensure that only the bases targeted for conversion are altered. False positive mutation artifacts will result from alterations of bases not targeted for conversion within the sequence probed for mutations. Preconversion using 3′ Q_6 primers forming a Q_6:G mismatch avoids starting polymerase extension with a G·T mismatch. In subsequent amplification cycles, A is apparently written frequently opposite Q_6. This observation is consistent with the results of Hill *et al*. in which Q_6 base paired like C and T with nearly equal frequency (41). Facile tautomerization allows Q_6 to mimic either pyrimidine when base paired and avoids mismatch wobble. When the natural base primer is added after preconversion, a significant quantity of perfect match template already exists, otherwise the MK artifact would appear in the reaction regardless of preconversion. Other nucleotide analogs in addition to Q_6 may serve as a bridge for more efficient conversions (19).

We have measured the fidelity of polymerase extension from primers in PCR and found conditions which in some cases improve fidelity. Presumably, higher fidelity resulted from a decrease in polymerase misincorporation, primer slippage and template degradation. PCR/RE/LDR allows the measurement of very low level 'mutant' sequences by preferentially amplifying non-wild-type sequences. Our results clearly demonstrate that natural base mismatch primer extension cannot be used as a general technique to create restriction sites at will in any sequence for RFLP analysis. As demonstrated here (Fig. 4) and observed previously (18,30), natural base mismatch extension is prone to error. To perfectly engineer a restriction site from existing sequence, an error-free approach is required. Our results indicate that the use of nucleotide analogs combined with high fidelity PCR conditions may radically decrease errors. Monitoring the true specificity of primer extension was possible in these studies because LDR can measure specific PCR errors accurately and with high sensitivity. Thus, the products of different polymerases and buffers could be assayed at different steps during PCR/RE/LDR to maximize both PCR efficiency and fidelity. As a result, a PCR/RE/LDR strategy could be assembled to achieve the goal of 10^5 sensitivity. However, this highest sensitivity was achieved only in the special case of no conversion at a pre-existing *Msp*I site. At this time, primer slippage remains an important mechanism through which mismatch primer extension errors can arise (19). Although the importance of this source of error *in vivo* is uncertain, it may have a dramatic impact on allele-specific PCR and other *in vitro* methods of mutation detection. An additional source of error arises from using natural base primers to select specific sequences for amplification following preconversion with nucleotide analogs. This is because a fraction of the selective natural base primers may form a mismatched pair with bases other than the intended base. It is known that a characteristic set of different bases insert opposite nucleotide analogs (19,41). Thus, a high fidelity mismatch primer extension protocol awaits the development of new convertides that can overcome these problems. In combination with high fidelity PCR and LDR monitoring of efficiency, mismatch primer extension may become a technique for the precise introduction of desired mutations without artifacts.

ACKNOWLEDGEMENTS

We thank Michael Wigler, Thierry Soussi, Mark Sobel, Jerard Hurwitz, Saul Silverstein, Harry Ostrer, Michael Osborne, Daniel Knowles, Pat Paty, Reyna Favis and members of the Barany and Paty laboratories for technical assistance and helpful discussion. This work was supported by grants from the National Cancer Institute (CA65930-02), Strang Cancer Prevention Center and the Applied Biosystems Division of the Perkin Elmer Corporation.

REFERENCES

1 Cotton,R.G. (1993) *Mutat. Res.*, **285**, 125–144.
2 Bottema,C.D. and Sommer,S.S. (1993) *Mutat. Res.*, **288**, 93–102.
3 Fan,E., Levin,D.B., Glickman,B.W. and Logan,D.M. (1993) *Mutat. Res.*, **288**, 85–92.
4 Cariello,N.F. and Skopek,T.R. (1993) *Mutat. Res.*, **288**, 103–112.
5 Sidransky,D., Tokino,T., Hamilton,S.R., Kinzler,K.W., Levin,B., Frost,P. and Vogelstein,B. (1992) *Science*, **256**, 102–105.
6 Brennan,J.A., Mao,L., Hruban,R.H., Boyle,J.O., Eby,Y.J., Koch,W.M., Goodman,S.N. and Sidransky,D. (1995) *N. Engl. J. Med.*, **332**, 429–435.
7 Newton,C.R., Graham,A., Heptinstall,L.E., Powell,S.J., Summers,C., Kalsheker,N., Smith,J.C. and Markham,A.F. (1989) *Nucleic Acids Res.*, **17**, 2503–2516.
8 Tada,M., Omata,M., Kawai,S., Saisho,H., Ohto,M., Saiki,R.K. and Sninsky,J.J. (1993) *Cancer Res.*, **53**, 2472–2474.
9 de Fromentel,C.C. and Soussi,T. (1992) *Genes Chromosomes Cancer*, **4**, 1–15.
10 Barany,F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
11 Barany,F. (1991) *PCR Methods Applicat.*, **1**, 5–16.
12 Day,D.J., Speiser,P.W., White,P.C. and Barany,F. (1995) *Genomics*, **29**, 152–162.
13 Khanna,M., Park,P., Zirvi,M., Cao,W., Picon,A., Day,J., Paty,P. and Barany,F. (1999) *Oncogene*, **18**, 27–38.
14 Luo,J., Bergstrom,D.E. and Barany,F. (1996) *Nucleic Acids Res.*, **24**, 3071–3078.
15 Sandy,M.S., Chiocca,S.M. and Cerutti,P.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 890–894.
16 Pourzand,C. and Cerutti,P. (1993) *Mutat. Res.*, **288**, 113–121.
17 Kwok,S., Kellogg,D.E., McKinney,N., Spasic,D., Goda,L., Levenson,C. and Sninsky,J.J. (1990) *Nucleic Acids Res.*, **18**, 999–1005.
18 O'Dell,S.D., Humphries,S.E. and Day,I.N. (1996) *Genome Res.*, **6**, 558–568.
19 Day,J.P., Hammer,R.P., Bergstrom,D. and Barany,F. (1999) *Nucleic Acids Res.*, **8**, 1810–1818.
20 Kumar,R. and Barbacid,M. (1988) *Oncogene*, **3**, 647–651.
21 Anderson,J.A., Irish,J.C. and Ngan,B.Y. (1992) *J. Otolaryngol.*, **21**, 321–326.
22 Barnes,W.M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2216–2220.
23 Cheng,S., Fockler,C., Barnes,W.M. and Higuchi,R. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 5695–5699.
24 Sang,N., Condorelli,G., De Luca,A., MacLachlan,T.K. and Giordano,A. (1996) *Anal. Biochem.*, **233**, 142–144.
25 Eckert,K.A. and Kunkel,T.A. (1990) *Nucleic Acids Res.*, **18**, 3739–3744.
26 Eckert,K.A. and Kunkel,T.A. (1991) *PCR Methods Applicat.*, **1**, 17–24.
27 Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) *Nucleic Acids Res.*, **24**, 3546–3551.
28 Brail,L., Fan,E., Levin,D.B. and Logan,D.M. (1993) *Mutat. Res.*, **303**, 171–175.
29 Cha,R.S., Zarbl,H., Keohavong,P. and Thilly,W.G. (1992) *PCR Methods Applicat.*, **2**, 14–20.
30 Eiken,H.G., Odland,E., Boman,H., Skjelkvale,L., Engebretsen,L.F. and Apold,J. (1991) *Nucleic Acids Res.*, **19**, 1427–1430.
31 Keohavong,P., Ling,L., Dias,C. and Thilly,W.G. (1993) *PCR Methods Applicat.*, **2**, 288–292.
32 Keohavong,P. and Thilly,W.G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9253–9257.
33 Good,N. (1976) In Fasman,G.D. (ed.) *Handbook of Biochemistry and Molecular Biology.* CRC Press, Cleveland, OH, pp. 367–369.
34 Blanchard,J.S. (1984) *Methods Enzymol.*, **104**, 404–414.
35 Cariello,N.F., Swenberg,J.A. and Skopek,T.R. (1991) *Nucleic Acids Res.*, **19**, 4193–4198.
36 Mattila,P., Korpela,J., Tenkanen,T. and Pitkanen,K. (1991) *Nucleic Acids Res.*, **19**, 4967–4973.
37 Huang,H. and Keohavong,P. (1996) *DNA Cell Biol.*, **15**, 589–594.
38 Ling,L.L., Keohavong,P., Dias,C. and Thilly,W.G. (1991) *PCR Methods Applicat.*, **1**, 63–69.
39 Lundberg,K.S., Shoemaker,D.D., Adams,M.W., Short,J.M., Sorge,J.A. and Mathur,E.J. (1991) *Gene*, **108**, 1–6.
40 Andre,P., Kim,A., Khrapko,K. and Thilly,W.G. (1997) *Genome Res.*, **7**, 843–852.
41 Hill,F., Loakes,D. and Brown,D.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 4258–4263.