1810–1818 *Nucleic Acids Research, 1999, Vol. 27, No. 8*

Nucleotide analogs facilitate base conversion with 3′ mismatch primers

Joseph P. Day, Donald Bergstrom[1], Robert P. Hammer[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 Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 46208, USA and [2]Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803, USA

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

ABSTRACT

We compared the efficiency of PCR amplification using primers containing either a nucleotide analog or a mismatch at the 3′ base. To determine the distribution of bases inserted opposite eight different analogs, 3′ analog primers were used to amplify four different templates. The products from the reactions with the highest amplification efficiency were sequenced. Analogs allowing efficient amplification followed by insertion of a new base at that position are herein termed 'convertides'. The three convertides with the highest amplification efficiency were used to convert sequences containing C, T, G and A bases into products containing the respective three remaining bases. Nine templates were used to generate conversion products, as well as non-conversion control products with no base change. We compared the ability of natural bases to convert specific sites with and without a preconversion step using nucleotide analog primers. Conversion products were identified by a ligation detection reaction using primers specific for the converted sequence. We found that conversions resulting in transitions were easier to accomplish than transversions and that sequence context influences conversion. Specifically, primer slippage appears to be an important mechanism for producing artifacts via polymerase extension of a 3′ base or analog transiently base paired to neighboring bases of the template. Nucleotide analogs could often reduce conversion artifacts and increase the yield of the expected product. While new analogs are needed to reliably achieve transversions, the current set have proven effective for creating transition conversions.

INTRODUCTION

Highly sensitive assays that detect low abundance mutations rely on PCR to amplify the target sequence. However, a non-selective PCR strategy will amplify both mutant and wild-type alleles with approximately equal efficiency, resulting in low abundance mutant alleles comprising only a small fraction of the final product. If the mutant sequence comprises <25% of the amplified product, it is unlikely that DNA sequencing will be able to detect the presence of such an allele. Although it is possible to accurately quantify low abundance mutations by first separating the PCR products by cloning and subsequently probing the clones with allele-specific oligonucleotides (ASOs) (1–3), this approach is time consuming. In contrast to the above, allele-specific PCR methods can rapidly and preferentially amplify mutant alleles. For example, multiple mismatch primers have been used to detect H-*ras* mutations at a sensitivity of one mutant in 10^5 wild-type alleles (4) and claims as high as one mutant in 10^6 wild-type alleles have been reported (5,6). However, careful evaluation suggests these successes are limited to allele-specific primers discriminating through 3′ purine·purine mismatches. For the more common transition mutations, the discriminating mismatch on the 3′ primer end (i.e. G:T or C:A mismatch) will be removed in a small fraction of products by polymerase error during extension from the opposite primer on wild-type DNA. Thereafter, these error products are efficiently amplified and generate false positive signal. One strategy to eliminate this polymerase error problem is to deplete wild-type DNA early in PCR.

Several investigators have explored selective removal of wild-type DNA by restriction endonuclease digestion in order to enrich for low abundance mutant sequences. These restriction fragment length polymorphism (RFLP) methods detect approximately one mutant in 10^6 wild-type or better by combining the sensitivity of polymerase with the specificity of restriction endonucleases. One approach has used digestion of genomic DNA followed by PCR amplification of the uncut fragments (RFLP–PCR) to detect very low level mutations within restriction sites in the H-*ras* and p53 genes (7,8). Similar results have been obtained by digestion following PCR and subsequent amplification of the uncleaved DNA now enriched for mutant alleles (PCR–RFLP) (9–11). Although sensitive and rapid, RFLP detection methods are limited by the requirement that the location of the mutations must coincide with restriction endonuclease recognition sequences. To circumvent this limitation, primers that introduce a new restriction site have been employed in 'primer-mediated RFLP' (12–17). However, subsequent investigators have demonstrated

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

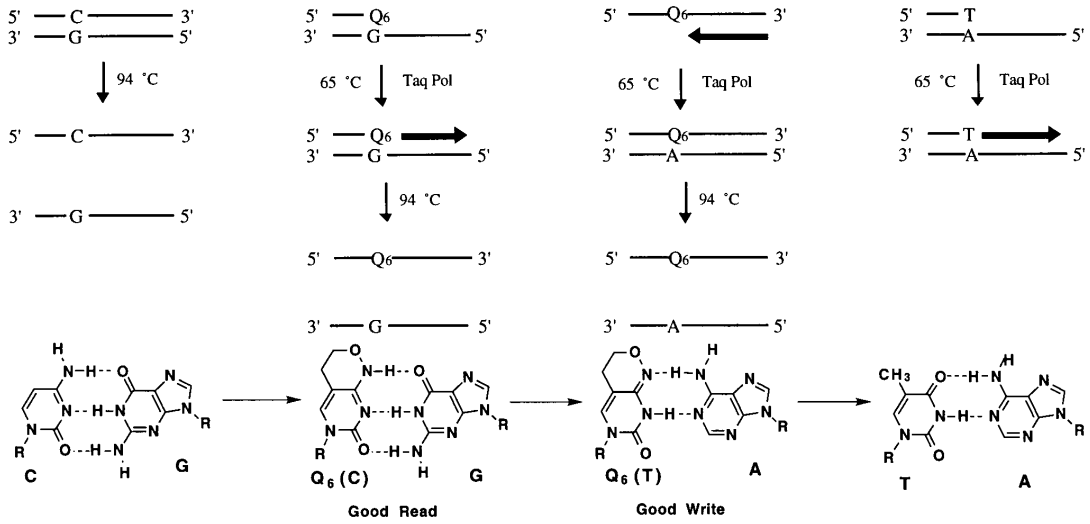

Figure 1. Conversion facilitated by nucleotide analog preconversion. A C:G base pair in a sequence is targeted for conversion to a T:A base pair. Rather than using a 3′ natural base mismatch primer to attempt direct conversion, a nucleotide analog (Q_6) primer is used for preconversion. The Q_6 analog reads the G base well and allows polymerase to efficiently extend from the 3′ Q_6 primer. During PCR, the reverse primer anneals to the Q_6 PCR product and is extended by polymerase to synthesize the opposite strand. When polymerase reaches the Q_6 analog in the template, polymerase writes A (or G; not shown) opposite the analog and continues synthesis of the strand. After a few cycles, a pool of products is made with degenerate sequence opposite the analogs. A natural base primer is then added to selectively amplify the products having the desired base change.

that errors are produced at the very next base by polymerase extension from primers having 3′ natural base mismatches (18–20). Such templates fail to cleave during restriction digestion and amplify as false positives that are indistinguishable from true positive products extended from mutant templates.

Use of nucleotide analogs may reduce errors resulting from polymerase extension and improve base conversion fidelity. Nucleotide analogs that are designed to base pair with more than one of the four natural bases herein are termed ‘convertides’. Base incorporation opposite different convertides has been tested (21). For each analog, PCR products were generated using *Taq* polymerase and primers containing an internal nucleotide analog. The products generated showed a characteristic distribution of the four bases incorporated opposite the analogs. Of significance, these products retained the original sequence at all natural base positions. Convertides readily form degenerate amplification products by virtue of their ability to assume different hydrogen bonding patterns through either tautomeric shift (22), bond rotation (23) or base stacking (24,25). Thus, PCR primers containing convertides may be used to facilitate base conversion. In principle, using the 6*H*,8*H*-3,4-dihydropyrimido[4,5-*c*][1,2]-oxazine-7-one analog (Q_6), which is known to exhibit both the C-like and T-like tautomeric forms at the 3′-end of the primer (22), a C-G base pair may be converted to a T-A base pair (Fig. 1). Due to the better geometry, DNA polymerases may ‘read’ or extend better from a Q_6·G pair than a T·G mismatch (wobble base pair). Similarly, DNA polymerases may ‘write’ or incorporate both G and A bases opposite Q_6 (26), whereas A is always inserted opposite a T base. Thus, the Q_6 analog primer serves as an intermediary, providing a ‘preconversion’ step before a natural base primer is added to selectively amplify the desired product from the degenerate pool. While nucleotide analogs have great potential, they have not been tested in high sensitivity assays.

We synthesized several PCR primers containing one of eight different nucleotide analogs at the 3′-end (Fig. 2). PCR extension

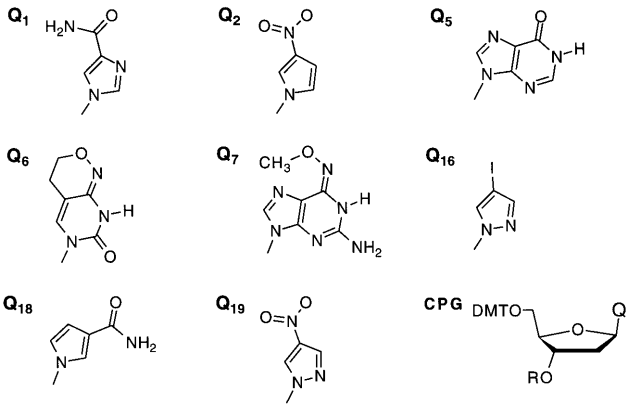

Figure 2. Nucleotide analogs used in PCR primers. In the final deprotected oligonucleotide, the name of the nucleoside containing the base analog shown is: Q_1, 1-(2′-deoxy-β-D-ribofuranosyl)imidazole-4-carboxamide; Q_2, 1-(2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole; Q_5, 2′-deoxyinosine; Q_6, 6-(2′-deoxy-β-D-ribofuranosyl)-6*H*,8*H*-3,4-dihydropyrimido[4,5-*c*][1,2]oxazine-7-one; Q_7, 2-amino-7-(2′-deoxy-β-D-ribofuranosyl)-6-methoxyaminopurine; Q_{16}, 1-(2′-deoxy-β-D-ribofuranosyl)-4-iodopyrazole; Q_{18}, 1-(2′-deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide; Q_{19}, 1-(2′-deoxy-β-D-ribofuranosyl)-4-nitropyrazole. Base analogs (Q) are attached to the 1′ position of deoxyribofuranose. The nucleoside analogs are attached to the controlled pore glass (CPG) column via a succinoyl linker (R, linker to CPG). The oligonucleotide is synthesized from the 5′-hydroxyl after removal of the dimethoxytrityl (DMT) protecting group, placing the analog at the 3′-end. After cleavage from the CPG column and deprotection, the oligonucleotide is extended by polymerase from the 3′ base analog hydroxyl group (R = H).

efficiency and fidelity were measured and the mutations in PCR products identified by sequencing and ligation detection reaction (LDR) (27–29). We found that primer-mediated RFLP–PCR using natural base 3′ mismatch primers is prone to high levels of misextension errors. Specific misextension errors in each reaction

were quantified in the range 0.1–100% using LDR (30). However, conversion fidelity could be significantly improved if preconversion with 3′ convertide primers was performed.

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 analogs 2′-deoxyinosine (Q_5), 6-(2′-deoxy-β-D-ribofuranosyl)-6*H*,8*H*-3,4-dihydropyrimido[4,5-*c*][1,2]oxazine-7-one (Q_6) and 2-amino-7-(2′-deoxy-β-D-ribofuranosyl)-6-methoxyaminopurine (Q_7) were synthesized using 2′-deoxyinosine-CPG, dP-CPG and dK-CPG, respectively (Glen Research) (Fig. 1). The oligonucleotide primers containing Q_1, Q_2 and Q_{18} at the 3′-position were synthesized from Q_1-, Q_2- and Q_{18}-derived CPG synthesized from Q_1 (31), Q_2 (24) and Q_{18} (25) by the method of Pon *et al.* (32). Details of the synthesis of the iodopyrazole (Q_{16}) and nitropyrazole (Q_{19}) nucleosides will be reported separately.

PCR polymerases and buffers

The DNA polymerases used were Ampli*Taq*, Ampli*Taq* Stoeffel Fragment, Ampli*Taq* Fluorescent Sequencing (Applied Biosystems), Vent and Vent(exo–) (New England Biolabs) and Expand polymerase mix (*Taq* and *Pfu* polymerase mixture, in Expand High Fidelity kit; Boehringer Mannheim). The commercially available PCR buffers used were supplied in the Ampli*Taq* and Expand High Fidelity kits. An alternative buffer, CiNF, is described elsewhere (33). Briefly, CiNF reactions contain 20 mM citrate, pH 7.6, 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, 10% formamide, primers and template DNA. All PCR and LDR reactions described below were performed under paraffin oil.

Mismatch extension efficiency

Primers containing natural bases and nucleotide analogs were used in PCR to measure the efficiency of product formation from synthetic duplex p53 exon 7 templates having *Msp*I (CCGG), *Taq*I (TCGA), *Hha*I (GCGC) or *Tai*I (ACGT) sites at the *Msp*I position containing codon 248. The primers hybridized to wild-type sequence on either side of the *Msp*I site with the 3′-ends of the primers extending one base into the site on each side (Fig. 3A). Eight different analogs and the four natural bases were tested in parallel reactions on each of the four synthetic templates. PCR was performed using *Taq* Stoeffel Fragment or *Taq* Fluorescent Sequencing polymerases with the buffer supplied for each polymerase. We used 10 pmol of each primer and 20 fmol of duplex template, 0.2 mM each dNTP and 4 mM $MgCl_2$. Parallel reactions underwent 10, 20, 30, 40 and 50 PCR cycles of 94°C for 15 s, 65°C for 1 min. Efficiency and yield were determined from samples run on 3% agarose gels and stained with ethidium bromide.

Mismatch conversion product sequencing

Products most efficiently amplified by each analog were diluted 1000-fold in water. The diluted DNA products were reamplified for 20 cycles of 94°C for 15 s, 65°C for 2 min using the same polymerase and buffer as in the previous PCR, but with the addition of 10 pmol of ‘zipcode’-containing primers p53zip248 and p53zip248R (Fig. 3A). Zipcode sequences are oligonucleotides with no known sequence similarity to DNA sequences in any organism. Amplification with zipcode primers is intended to specifically amplify the zipcode-containing products of the previous PCR, i.e. only converted DNA (containing zipcodes) and not the nearly identical unconverted DNA (lacking zipcodes) will be amplified. Conversion products were run on 3% agarose gels and bands of the expected size excised. DNA was extracted from the gel slices by centrifugation in a 235C microcentrifuge (Fisher) for 30 min through a 0.45 µm HVLP filter (Millipore). The conversion product was dried and resuspended in ABI Dye Terminator Cycle Sequencing reaction mix with one of the zipcode primers according to kit instructions (Applied Biosystems). An equal volume (3 µl each) of sequencing reaction was combined with dye mix consisting of 83% formamide (Eastman), 4 mM EDTA and 8 mg/ml Blue Dextran (Sigma). Samples were electrophoresed on a 7 M urea–10% acrylamide gel (19:1 bis, 0.6× TBE in gel and running buffer) in an ABI 373 DNA Sequencer. Data were analyzed using ABI 373A DNA Sequencer Data Analysis software v.1.2.0.

Conversion product identification

Conversion fidelity was tested using nine different synthetic templates, with and without preconversion using three primers containing Q_5, Q_6 and Q_7 (see Oligonucleotide synthesis). Preconversion PCR was performed with 3′ analog primers prior to adding the desired natural base primers, in an effort to avoid mismatch primer extension. The 50 bp duplex DNA templates contained the wild-type p53 sequence surrounding codon 248 (Fig. 3B), except for the bases corresponding to the *Msp*I site (CCGG). The following sequences were substituted at the *Msp*I position: 1) CCGG (wild-type); 2) CTGG; 3) CGGG; 4) CAGG; 5) TCGA; 6) GCGC; 7) ACGT; 8) ACGT; 9) GCGC. Preconversion was performed with hot start using 50 fmol/µl p53-248Q_N and p53-248Q_NR primers and Vent(exo–) in CiNF buffer and 10 fmol/µl of duplex template. Preconversion used two PCR cycles of 94°C for 15 s, 55°C for 1 min, 60°C for 1 min. Product was stored at 4°C. Conversion reactions were started with 1 µl of preconversion reaction containing the same polymerase and buffer, but no additional template. Each reaction required 10 pmol of each primer, using one of the four pairs p53zip248N and p53zip248NR (N = C, T, G or A). Parallel conversion reactions with no preconversion were initiated with a hot start by adding 10 fmol of synthetic duplex template instead of preconversion reaction aliquot. PCR cycles were as follows: five cycles of 94°C for 15 s, 55 + 1°C/cycle for 1 min, 60°C for 1 min; 20 cycles of 94°C for 15 s, 60°C for 2 min. A final extension was performed at 60°C for 5 min. Polymerase was inactivated by freezing and

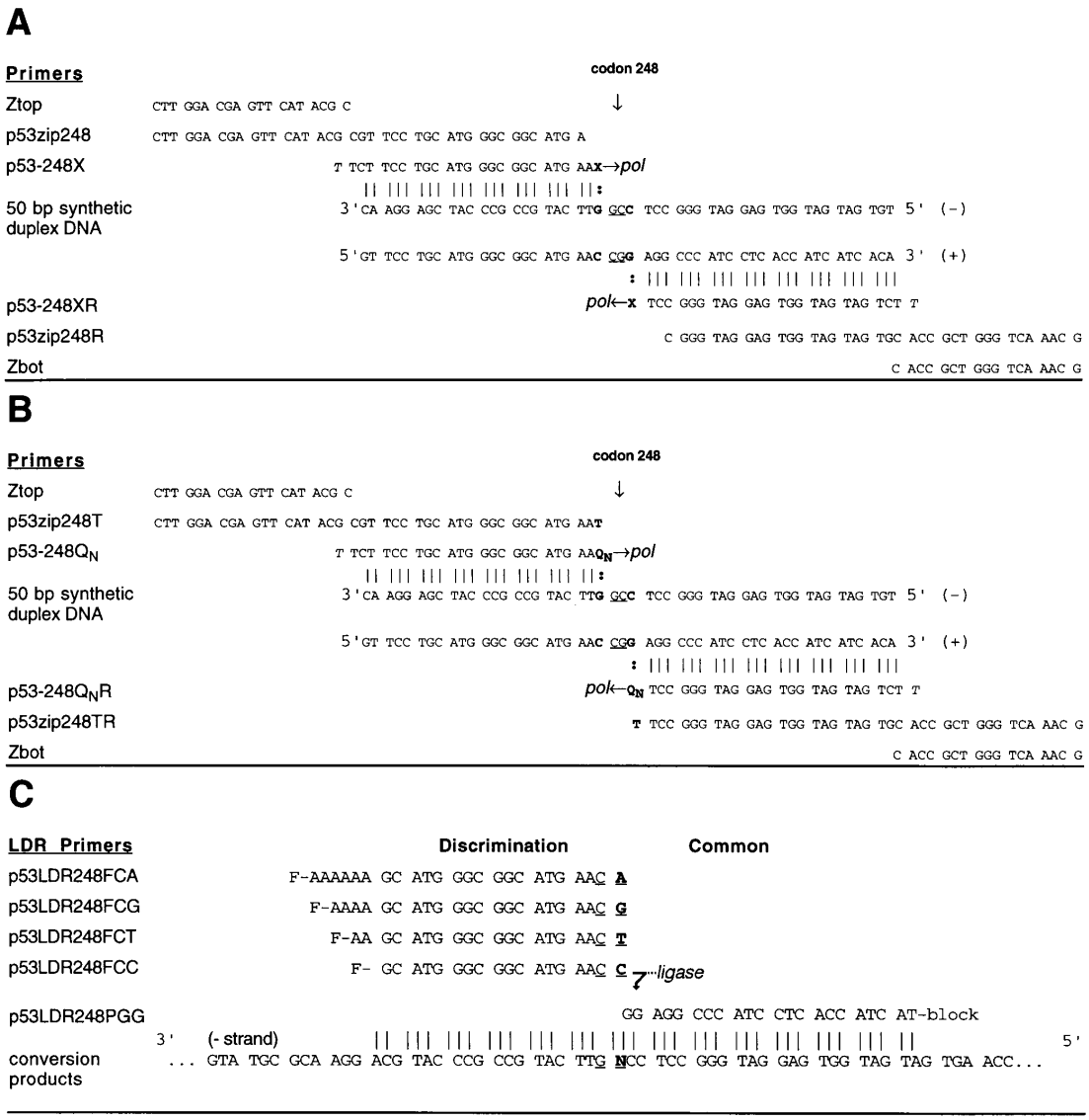

Figure 3. Primers used in mismatch extension and PCR/LDR. Complimentary (– strand) sequences are shown in reverse orientation (3′→5′), e.g. reverse strand primers (names ending in R). (**A**) One of nine different synthetic 50 bp duplex templates is shown melted with primers aligned to complementary sequence. Primer extension was performed using 3′ natural base and nucleotide analog primers (p53-248X and p53-248XR). Some extension products were reamplified using truncated zipcode primers p53zip248 and p53zip248R and sequenced using one of the zipcode primers (Ztop or Zbot). (**B**) Preconversion was performed on nine different 50 bp synthetic duplex templates using 3′ nucleotide analog primers, e.g. p53-248Q_6 and p53-248Q_6R. Conversion, with or without preconversion, was performed using primers containing the 3′ natural base, e.g. primers p53zip248T and p53zip248TR. These conversion products were reamplified using zipcode primers and identified by LDR. (**C**) LDR primer sets were designed to identify specific base changes in conversion products. LDR primers anneal in competition with each other to conversion products. Perfectly complementary upstream and downstream LDR primers with no overlap or gap ligate with high specificity. Discrimination primers had different length 5′ tails to allow specific product separation on an acrylamide gel. Shown are a set of primers used to identify PCR error products in non-conversion of wild-type template.

thawing twice. Products were diluted 10× in water and reamplified by adding 1–20 µl of Expand polymerase and buffer mix. PCR was performed for 20 cycles (30 cycles for low yield reactions) of 94°C for 15 s, 65°C for 2 min using 12 pmol of zipcode primers Ztop and Zbot (Fig. 3). LDR was performed as described below to identify the conversion products generated.

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. 3C). Sets of discrimination and common primers were synthesized to perform LDR on the expected conversion products and varied at the bases (B_i) corresponding to the *Msp*I position sense strand ($B_1B_2B_3B_4$ = CCGG for wild-type). The discrimination primers had wild-type sequence and terminate in -B_1B_2(-OH-3′). The discrimination primers were synthesized as a set of four primers each with C, T, G and A in turn at B_2. The common LDR primers had (5′-PO_4-)B_3B_4- followed by wild-type sequence and hybridized to the template with its 5′ base adjacent to the 3′ base of a discrimination primer. Discrimination

primers varied the 3′-terminal base to identify error products at B_2 of the *Msp*I position. For simplicity, only B_2 was monitored. LDR primers matched the expected conversion products; for example, conversion of -CCGG- template to -ACGT- required discrimination primers ending in -AC, -AT, -AG and -AA and a common primer with 5′-pGT-. Discrimination primers had 5′ tails of different length and a FAM label for fluorescence detection. The tail length allowed physical separation of different LDR products on an acrylamide gel and thus identification of the LDR products.

LDR reactions were preincubated for 1.5 min at 94°C prior to the addition of 5 nmol *Tth* ligase, followed by 10 LDR cycles of 94°C for 15 s, 65°C for 2 min and a final hold briefly at 94°C. Reactions were cold quenched and stored at –70°C. The LDR products were separated on 10% acrylamide gels containing 7 M urea, with 0.6× TBE (1× TBE is 90 mM Tris base, 90 mM borate, 2 mM EDTA) in the gel and running buffer. Data were collected using an ABI 373 DNA sequencer with Genescan 672 software.

Image processing

Gel pictures were produced by the ABI 672 Analysis software. Dye-specific images 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. The background subtracted images were reinverted and rendered in pseudocolor by Photoshop to make intensity differences easier to compare. Except for color replacement, only linear image processing was performed to preserve relative intensities.

RESULTS AND DISCUSSION

Initial experiments were designed to determine the efficiency of generating PCR products when using primers containing 3′-terminal nucleotide analogs (Materials and Methods). Eight different analogs were designed to pair with more than one of the four natural bases in order to convert one base to another base at a specific position in a sequence. Primer pairs containing either a nucleotide analog or one of the four natural bases at their 3′-ends were used to amplify four different templates (Fig. 3A). Each nucleotide analog and natural base was mispaired (or paired) in turn with all four natural bases on the opposite strand and amplification was attempted with either *Taq* Stoeffel Fragment or *Taq* Fluorescent Sequencing polymerases. The relative amplification efficiency was determined by the number of cycles required to generate visible product on an ethidium bromide stained agarose gel (Table 1). We found that both *Taq* Stoeffel Fragment and *Taq* Fluorescent Sequencing polymerases produced comparable amounts of product (data not shown). Perfectly matched natural base primers generated visible product after 10 cycles, however, some analog primers generated no product after 50 cycles. The analogs that did amplify with high efficiency were those that were best able to ‘read’ the opposite strand sequence (Fig. 1).

One product for each analog (as well as the natural base controls) was reamplifed and sequenced to determine polymerase preference in inserting nucleotide bases opposite the analog (Table 1). We found that the Q_1, Q_5, Q_6, Q_{16} and Q_{18} primers generated detectable true conversion product, however, only Q_5 primers generated almost exclusively true conversion product. No single analog functioned as a ‘universal base’ (26) capable of generalized conversion. Unexpectedly, some products contained sequences that were difficult to read across the middle four bases, suggesting single base insertions or deletions occurred during PCR extension. This was especially prevalent in products generated from mismatched natural bases (see below).

Table 1. Extension efficiency and conversion with 3′ natural base and nucleotide analog primers

Primer 3′ base	TCGA template reads A writes (efficiency)	CCGG template reads G writes (efficiency)	GCGC template reads C writes (efficiency)	ACGT template reads T writes (efficiency)
T	A (+++)	A (++)	(++)	(++)
C	(++)	G (+++)	(++)	(++)
G	(++)	(++)	C (+++)	(+++)
A	(+)	(+)	T (+++)	T (+++)
Q_1	A,T (±)	(±)[a]	(–)	(+++)[a]
Q_2	(±)[a]	(±)[a]	(–)	T (++)[a]
Q_5	(++)	(++)	(+++)	C (+++)
Q_6	A,G (+++)	(+++)	(++)	(++)
Q_7	(+)	(+)	(+++)	T (+++)
Q_{16}	A,T (+)[a]	(–)	(–)	(–)
Q_{18}	(+)[a]	(±)[a]	(±)[a]	T,A (+++)
Q_{19}	A (++)[a]	(–)	(±)[a]	(+)[a]

[a]Low product yield.

Four different templates were used to test primer extension from a 3′ base or analog paired in turn with A, G, C and T. Relative efficiency was determined by the number of cycles required to generate visible product with *Taq* Stoeffel Fragment polymerase: (+++), 10 cycles; (++), 20 cycles; (+), 30 cycles; (±), 40–50 cycles; (–), no product. Two of the natural base mismatch primer products were sequenced. Generally, the most efficiently amplified template for each analog was reamplified with truncated primers and sequenced to determine which bases are written opposite each analog. In one case (Q_1) a lower efficiency extension product with higher yield was selected for sequencing. Mixed base writing preference for some analogs is indicated, with most frequent product listed first.

To test the ability of convertides to reduce mismatch extension errors, we assessed the effects of preconversion PCR cycles on fidelity. PCR products generated from template amplified with only natural base conversion primers were compared to products resulting from two initial PCR cycles using convertides followed by selective amplification using specific natural base primers. We performed preconversion PCR with primer pairs containing Q_5, Q_6 and Q_7 analogs, since these convertides had been shown to be the most efficiently extended. To improve overall PCR fidelity and 3′ mismatch primer extension, CiNF buffer (Materials and Methods) was used (33). Nine different synthetic duplex templates containing mutated *Msp*I sites were amplified with or without preconversion using 3′ analog preconversion primers. Both natural base conversion primers and 3′ analog preconversion primers were designed to manipulate the outside bases CCGG of the *Msp*I position (Fig. 3). Some conversions were intended to serve as controls. In these cases, the original bases in the template were either restored after analog preconversion or never changed with full-length perfect match primers. All steps were performed identically between preconversion and non-preconversion reactions, except that preconversion reactions used as template the product of two cycles of convertide PCR for succeeding rounds of amplification,

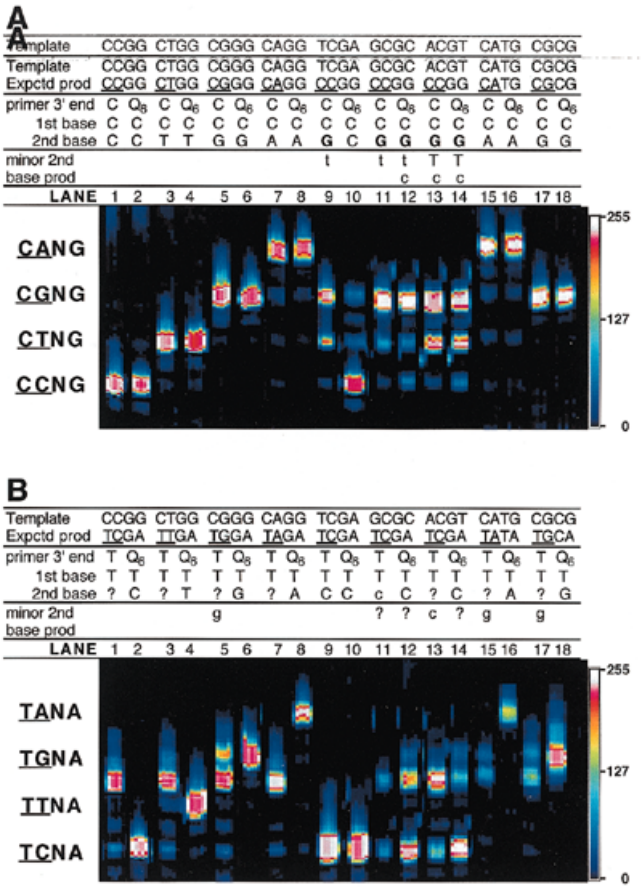

Figure 4. Conversion by natural base and Q_6 convertide. Conversion products from nine templates were detected by PCR/LDR (Materials and Methods). Each template was a 50 bp synthetic duplex DNA of identical sequence except for the central four bases which have the sequence indicated. Conversion occurred within these four bases. The expected conversion products produced by starting with the conversion primers having the indicated 3′ natural base or convertide are shown. (**A**) Conversion of the first base to C with and without Q_6 preconversion. (**B**) Conversion of the first base to T with and without Q_6 preconversion.

while synthetic duplex served as the starting material for PCRs with no preconversion. In both cases, 3′ natural base primers were used to selectively amplify the desired end product. These primers contained non-hybridizing zipcode sequences on their 5′-ends (Materials and Methods), which ultimately served as primer binding sites for the final 20–30 cycles of PCR (Fig. 3B). Conversion products were quantified by LDR (Fig. 3C).

We found that overall, natural base mismatch conversion generated >80% incorrect conversion products (Fig. 4A, lane 9, and B, lanes 1, 3, 5, 7, 15 and 17), but preconversion could improve the fidelity and/or the yield of some conversions. In general, transversions were difficult to achieve even with preconversion. G→C and A→C conversion generated very little of the expected product with either the natural base or Q_6 primers (Fig. 4A, lanes 11–14). Use of Q_6 preconversion improved the yield of G→T and A→T conversion products (compare natural base conversion in Fig. 4B, lanes 11 and 13, with Q_6 preconversion in lanes 12 and 14). In the case of transitions, C→T conversion produced unexpected one base shortened artifacts with natural base mismatch primers on the CXGG templates (Fig. 4B, lanes 1, 3, 5, 7, 15 and 17), but the correct products were generated when using Q_6 preconversion (Fig. 4B, lanes 2, 4, 6, 8, 16 and 18). In addition, Q_6 primers did improve the yield of the expected T→C conversion product (Fig. 4A, lanes 9 and 10). The controls performed as expected: all C→C and T→T non-conversion reactions worked correctly without convertides (Fig. 4A, lanes 1, 3, 5, 7, 15 and 17, and B, lane 9) and the corresponding Q_6 preconversion products were restored to the original sequence (Fig. 4A, lanes 2, 4, 6, 8, 16 and 18, and B, lane 10). In summary, Q_6 preconversion reduced or eliminated artifacts produced by natural base C→T and T→C conversion and facilitated transitions in general. Transversions were only partially successful: G→T and A→T conversions could be improved with preconversion, but G→C and A→C conversion could not be achieved.

Apparently correct conversions were observed with attempted C→G and C→A transversions, however, carefully designed control templates revealed that these ‘conversions’ were artifactual. C→G and C→A conversion appeared to be successful for templates containing a central CpG dinucleotide (Fig. 5A and B, lanes 1–3 and 13–21). However, the same final conversion products were observed with other templates lacking the central CpG dinucleotide, now clearly incorrect. For example, a GCGC product resulted during G conversion in reactions where the expected product should have contained T, G or A in the second position (Fig. 5A, lanes 4–12). Also, an ACGT product resulted during A conversion where the expected product should have inserted a non-C base in the second position (Fig. 5B, lanes 4–12 and 22–27). The mismatch primers used to alter the outer bases of the recognition site did not reach the central dinucleotide, yet these bases were altered. It is doubtful the ‘successful’ conversions occurred through the intended mechanism and thus represent fortuitous artifacts. The yield of LDR product was low for two palindromic templates despite efficient PCR (Fig. 5A and B, lanes 22–27). These conversion reaction products presumably contain a large fraction of insertions or deletions, which cannot be detected by the current set of LDR primers. In summary, C→G conversion was partially accomplished by both Q_5 (Fig. 5A, lanes 5, 8, 11 and 23) and the natural base G (Fig. 5A, lanes 4, 7, 10 and 22), however, preconversion does not appear to improve conversion. C→G conversion exhibits sequence dependence.

The results of the preconversion study indicate that errors in natural base conversion were prevalent, but the use of Q_5, Q_6 and Q_7 convertides in preconversion reduced polymerase error in certain cases. In terms of conversion reactions, transitions were easier to accomplish than transversions. This is in agreement with previous findings. Newton *et al.* observed more errors in extension of primers with 3′-terminal C·T, A·A and T·T mismatches (transversions) than with purine·pyrimidine mismatches (transitions) (34). In our hands, pyrimidine–pyrimidine conversion usually generated the expected product, especially when using convertides. In cases of purine–pyrimidine and pyrimidine–purine conversion, incorrect products were often generated (summary of results in Table 2). Formation of incorrect conversion products can be explained in part by a transient base pair slippage of the primer 3′ nucleotide (or analog) to a misaligned position on the template (Fig. 6). As a result, the sequence following the mismatch is not complementary to the original template. Consistent with this hypothesis is the observation of unreadable sequence immediately following the analog in our initial sequencing experiments. Palindromic products, especially

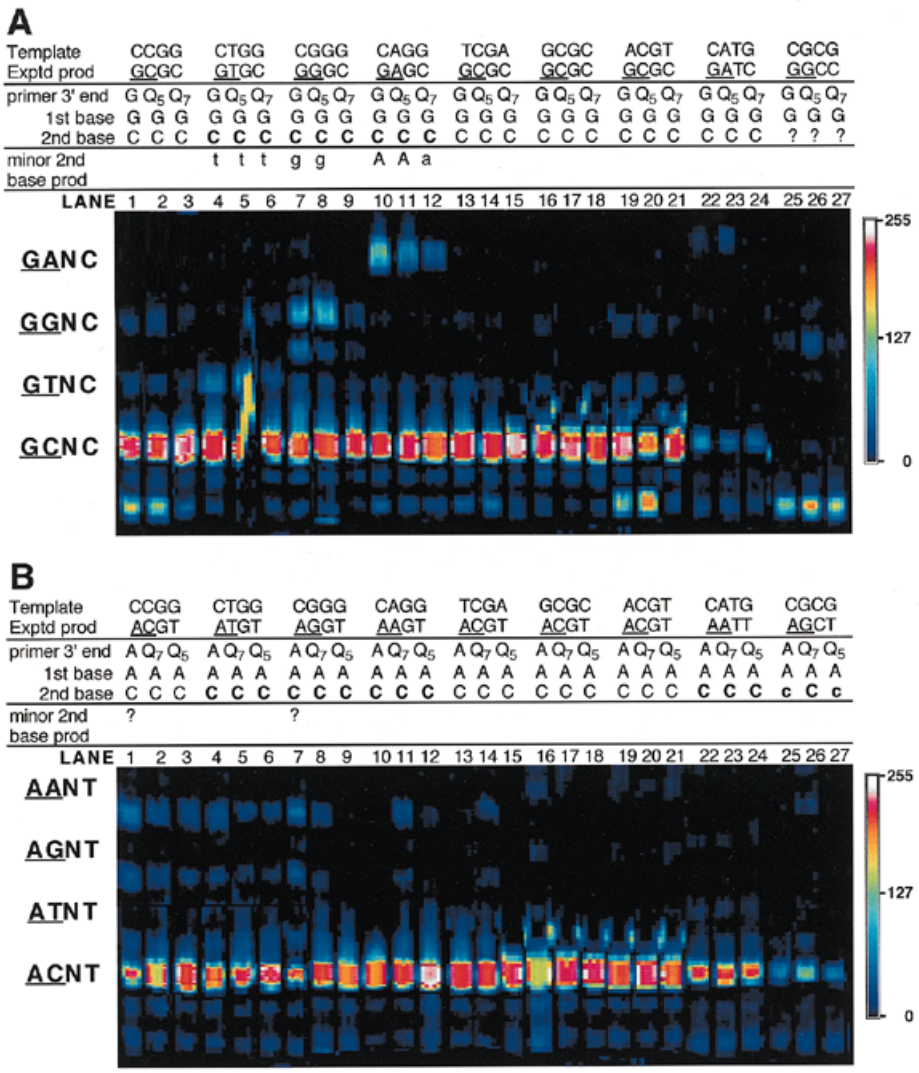

Figure 5. Conversion by natural base and Q_5 and Q_7 convertides. Conversion products from nine templates were detected by PCR/LDR (Materials and Methods). Each template was a 50 bp synthetic duplex DNA of identical sequence except for the central four bases which have the sequence indicated. Conversion occurred within these four bases. The expected conversion products produced by starting with the conversion primers having the indicated 3′ natural base or convertide are shown. (**A**) Conversion of the first base to G with and without Q_5 or Q_7 preconversion. (**B**) Conversion of the first base to A with and without Q_5 or Q_7 preconversion.

CpG dinucleotides, are themselves prone to slippage and extension. We observed palindromic products were frequently produced from non-palindromic templates. These artifacts were reduced by the presence of 10% formamide in the PCR buffer, presumably through destabilization of misaligned structures. Finally, nucleotide analogs produced fewer artifacts than natural bases. Different analogs produced different kinds and quantities of artifacts, perhaps according to their relative ability to base pair and stabilize a slippage misalignment. Thus, if polymerase extension is slow from an analog poorly base paired with the template, extension from a strong transient base pair generated by slippage could exceed the rate of extension from a weakly base paired 3′-terminal base.

As discussed earlier, PCR–RFLP has been widely used to detect rare mutations. A limitation of this technique is reliance on serendipity to yield mutations that can be modified to create restriction sites in either the wild-type or the mutant template. A second limitation imposed on this approach is the need to avoid using 3′-terminal mismatch primers, since extension from these primers is known to be error prone. To date, the majority of successful attempts have used interrupted palindromic restriction sites to avoid using 3′-terminal mismatch primers. Mutations in the cancer-causing genes K-*ras* and H-*ras* were detected at a sensitivity of 1 in 10^5 using PCR–RFLP with interrupted palindromic enzymes *Xmn*I (9), *Alw*NI (35) and *Bst*NI or *Mva*I (36,37). These PCR–RFLP experiments and others (18,38–42) avoid 3′-terminal mismatches, however, most cancer mutations are in sequences that cannot be converted to interrupted palindromes, for example at CpG dinucleotides.

Table 2. Most effective conversion (Figs 4 and 5)

Starting template[a]	First base converted to			
	C	T	G	A
1 CCGG	C	Q_6	Q_7 (FP)	Q_5 (FP)
2 CTGG	C	Q_6	X (err C)	X (err C)
3 CGGG	C	Q_6	Q_5 (err C)	X (err C)
4 CAGG	C	Q_6	G (err C)	X (err C)
5 TCGA	Q_6	T or Q_6	Q_7 (FP)	Q_5 (FP)
6 GCGC	X (err G)	Q_6	G	Q_5 or Q_7
7 ACGT	X (err G)	Q_6 weak	Q_7	A or Q_7
8 CATG	C	X	Q_5 (err C)	X (err C)
9 CGCG	C	Q_6	X	Q_7 (err C)

[a]The 50 bp synthetic duplex DNA templates containing p53 sequence spanning codon 248 are distinguished by the four bases replacing the *Msp*I site, which are shown.

Nine duplex DNA templates were used in conversion reactions. Each contained sequence identical to p53 surrounding codon 248, except the *Msp*I site was replaced by a different four base sequence ($B_1B_2B_3B_4$). B_1 and B_{4-} (opposite strand) were simultaneously converted in turn to C, T, G and A either directly by PCR using natural base primers or by preconversion PCR with nucleotide analog primers followed by PCR with natural base primers. In non-conversion control reactions the 'conversion' product is identical to the original template. A natural base is used to indicate control reactions and cases in which preconversion did not improve conversion. Preconversion was performed using Q_6 to facilitate conversion to C and T and using Q_5 and Q_7 to facilitate conversions to G and A. Conversion primers determine B_1 and B_4; LDR was performed to detect unintended base changes in B_2 (which ideally is unchanged after conversion). Conversion improved by preconversion is indicated by the nucleotide analog used. Preconversion equally as effective in control reactions as natural base primers is also indicated by the analog used. Low conversion fidelity results in large B_2 error. Major B_2 error products are identified (e.g. err C indicates C at B_2) and the absence of correct product indicated no conversion method was successful (X, no correct product). Apparently correct product probably formed through a fortuitous mechanism is indicated (FP, false positive).

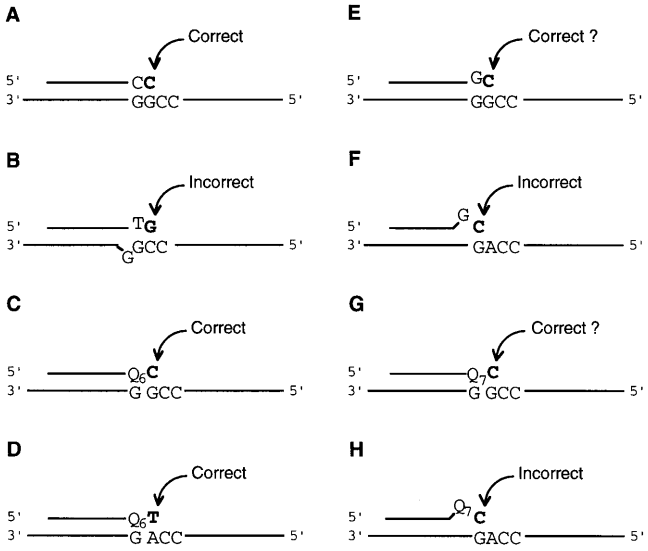

Figure 6. Fidelity of polymerase extension. Primer slippage accounts for many of the observed products of extension (Figs 4 and 5). (**A**) Perfectly complimentary primer gives correct product. (**B**) T:G mismatch at the second base explains TGGA (or TGCA) product. (**C**) Extension from a Q_6:G pairing with no slippage on the minus strand of the CCGG template (followed by 3′ T conversion primers) resulted in the expected TCGA product. (**D**) Extension from a Q_6:G pairing with no slippage on the minus strand of the CTGG template and several other templates (followed by 3′ T conversion primers) resulted in the expected products. (**E**) G·G mismatch extension apparently gave the expected GC product on one template, but perhaps only fortuitously (see F). (**F**) All extensions from G·G mismatches gave GC extension products, consistent with a G·T mismatch formed by slippage at the preceding base (Fig. 3). (**G**) Q_5:G and Q_7:G extension products apparently gave the expected GC product on one template, but perhaps only fortuitously (see H). (**H**) All extensions from Q_5:G and Q_7:G mismatches (followed by 3′ G conversion primers) gave GC extension products consistent with a Q_5:T or Q_7:T mismatch at the preceding base (Fig. 3).

A larger fraction of mutations would be made into targets for detection if contiguous recognition sequences could be introduced with as few errors as interrupted palindromic recognition sequences. Currently, contiguous restriction sites are introduced by terminal 3′ mismatch primer extension, which is prone to errors. O'Dell *et al*. tested a general method for introducing different restriction sites at CpG dinucleotides using mismatch PCR (19). The outer bases of four different CpG dinucleotides in the human LDL receptor gene were altered to create *Taq*I (TCGA), *Msp*I (CCGG) or *Hha*I (GCGC) sites. In these targets, *Taq*I sites were successfully generated by 3′ T mismatch primers. The method was able to detect homozygous and heterozygous individuals, however, the ratio of products representing each allele was not equal, as is expected in germline mutations. We have shown several cases where T mismatch conversion failed to create a *Taq*I site, thus the method is sequence dependent. O'Dell *et al*. found that C and G mismatch conversion failed. We agree with their conclusion that stronger base pairing leads to mispriming, possibly through stabilization of primer slippage on the template. Gotoda *et al*. claim to have successfully used PCR–RFLP to introduce an *Mae*II site (ACGT) by extension of a 3′ C·A mismatch to produce a T→C transition (43). Athma *et al*. used PCR extension of a 3′-terminal mismatch primer to create a restriction site for discriminating between two alleles (44). A G·T mismatch produced a *Mva*I site (CC A/T GG) through an A→G transition. We successfully performed A→G conversion using a natural base mismatch, but encountered difficulties with T→C conversion by natural base primers. In our hands, transitions can be accomplished more readily than transversions, but the yield of correct product can be sequence dependent. Others have also found that PCR–RFLP can produce false positive results (20). Our use of the ligase detection reaction allowed us to determine the precise amounts of misextension products generated.

We have measured the fidelity of polymerase extension from primers containing 3′ natural bases and nucleotide analogs. Our results indicate that natural base mismatch primer extension cannot be used as a general technique to create restriction sites in any given sequence for RFLP analysis. Primer slippage appears to be an important mechanism for producing error in mismatch primer extension. This source of error may have a dramatic impact on some allele-specific PCR and other methods of high sensitivity mutation detection. With further development and testing of nucleotide analogs to facilitate conversion, mismatch primer extension may become a technique that can efficiently introduce desired mutations with few artifacts. We have found some nucleotide analogs improve mismatch primer extension (Table 3). Further improvement of 3′ mismatch extension will be required to minimize the high degree of context-dependent error

seen in transversions and lead to improved levels of sensitivity and broader scope of PCR–RFLP-based mutation detection.

Table 3. Summary of conversion strategy

Starting base	Conversion to			
	C	T	G	A
C	C	Q_6		
T	Q_6	T		
G			G	A or Q_7
A			Q_5 or Q_7	A

A Q_n convertide indicates preconversion is required using the indicated convertide prior to final conversion using natural base primers. In some cases, an additional convertide or using only the natural base will result in the desired conversion.

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. We also thank Peiming Zhang and Travis Johnson from the Bergstrom laboratory who synthesized the Q_1, Q_2 and Q_{18} primers and Melissa Cameron from the Hammer laboratory who synthesized the Q_{16} and Q_{19} primers.

REFERENCES

1 Saiki,R.K., Bugawan,T.L., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1986) *Nature*, **324**, 163–166.
2 Sidransky,D., Tokino,T., Hamilton,S.R., Kinzler,K.W., Levin,B., Frost,P. and Vogelstein,B. (1992) *Science*, **256**, 102–105.
3 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.
4 Cha,R.S., Zarbl,H., Keohavong,P. and Thilly,W.G. (1992) *PCR Methods Applicat.*, **2**, 14–20.
5 Haliassos,A., Chomel,J.C., Grandjouan,S., Kruh,J., Kaplan,J.C. and Kitzis,A. (1989) *Nucleic Acids Res.*, **17**, 8093–8099.
6 Chen,Z.-Y. and Zarbl,H. (1997) *Anal. Biochem.*, **244**, 191–194.
7 Sandy,M.S., Chiocca,S.M. and Cerutti,P.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 890–894.
8 Pourzand,C. and Cerutti,P. (1993) *Mutat. Res.*, **288**, 113–121.
9 Kumar,R. and Barbacid,M. (1988) *Oncogene*, **3**, 647–651.
10 Kumar,R. and Dunn,L.L. (1989) *Oncogene Res.*, **4**, 235–241.
11 Jacobson,D.R. and Mills,N.E. (1994) *Oncogene*, **9**, 553–563.
12 Jacobson,D.R. and Moskovits,T. (1992) *PCR Methods Applicat.*, **1**, 299.
13 Chen,J. and Viola,M.V. (1991) *Anal. Biochem.*, **195**, 51–56.
14 Di Giuseppe,J.A., Hruban,R.H., Offerhaus,G.J., Clement,M.J., van den Berg,F.M., Cameron,J.L. and van Mansfeld,A.D. (1994) *Am. J. Pathol.*, **144**, 889–895.
15 Kahn,S.M., Jiang,W., Culbertson,T.A., Weinstein,B., Williams,G.M., Tomita,N. and Ronai,Z. (1991) *Oncogene*, **6**, 1079–1083.
16 Levi,S., Urbano-Ispizua,A., Gill,R., Thomas,D.M., Gilbertson,J., Foster,C. and Marshall,C.J. (1991) *Cancer Res.*, **51**, 3497–3502.
17 Mitsudomi,T., Viallet,J., Mulshine,J.L., Linnoila,R.I., Minna,J.D. and Gazdar,A.F. (1991) *Oncogene*, **6**, 1353–1362.
18 Hattori,M., Nanko,S., Dai,X.Y., Fukuda,R. and Kazamatsuri,H. (1994) *Biochem. Biophys. Res. Commun.*, **202**, 757–763.
19 O'Dell,S.D., Humphries,S.E. and Day,I.N. (1996) *Genome Res.*, **6**, 558–568.
20 Hodanova,K., Hrebicek,M., Cervenkova,M., Aerts,J.M. and Zeman,J. (1997) *J. Inherit. Metab. Dis.*, **20**, 611–612.
21 Hoops,G.C., Zhang,P., Johnson,W.T., Paul,N., Bergstrom,D.E. and Davisson,V.J. (1997) *Nucleic Acids Res.*, **25**, 4866–4871.
22 Brown,D.M. and Lin,P.K.T. (1991) *Carbohydrate Res.*, **216**, 129–139.
23 Bergstrom,D.E., Zhang,P. and Johnson,W.T. (1996) *Nucleosides Nucleotides*, **15**, 59–68.
24 Bergstrom,D.E., Zhang,P., Toma,P.H., Andrews,P.C. and Nichols,R. (1995) *J. Am. Chem. Soc.*, **117**, 1201–1209.
25 Zhang,P., Johnson,W.T., Klewer,D., Paul,N., Hoops,G., Davisson,V.J. and Bergstrom,D.E. (1998) *Nucleic Acids Res.*, **26**, 2208–2215.
26 Hill,F., Loakes,D. and Brown,D.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 4258–4263.
27 Barany,F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
28 Barany,F. (1991) *PCR Methods Applicat.*, **1**, 5–16.
29 Day,D.J., Speiser,P.W., White,P.C. and Barany,F. (1995) *Genomics*, **29**, 152–162.
30 Khanna,M., Park,P., Zirvi,M., Cao,W., Picon,A., Day,J., Paty,P. and Barany,F. (1999) *Oncogene*, **18**, 27–38.
31 Johnson,W.T., Zhang,P. and Bergstrom,D.E. (1997) *Nucleic Acids Res.*, **25**, 559–567.
32 Pon,R.T., Usman,N. and Ogilvie,K.K. (1988) *BioTechniques*, **6**, 768–775.
33 Day,J.P., Berstrom,D., Hammer,R.P. and Barany,F. (1999) *Nucleic Acids Res.*, **8**, 1819–1827.
34 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.
35 Anderson,J.A., Irish,J.C. and Ngan,B.Y. (1992) *J. Otolaryngol.*, **21**, 321–326.
36 Urban,T., Ricci,S., Grange,J.D., Lacave,R., Boudghene,F., Breittmayer,F., Languille,O., Roland,J. and Bernaudin,J.F. (1993) *J. Natl Cancer Inst.*, **85**, 2008–2012.
37 Ronai,Z. and Minamoto,T. (1997) *Hum. Mutat.*, **10**, 322–325.
38 Beutler,E., Gelbart,T. and West,C. (1990) *Clin. Chim. Acta*, **194**, 161–166.
39 Hingorani,A.D. and Brown,M.J. (1995) *Biochem. Biophys. Res. Commun.*, **213**, 725–729.
40 Kuwata,S., Yanagisawa,M., Saeki,H., Nakagawa,H., Etoh,T., Tokunaga,K., Juji,T. and Shibata,Y. (1995) *J. Allergy Clin. Immunol.*, **96**, 1051–1060.
41 Nishiwaki,Y., Kamino,K., Yoshiiwa,A., Nagano,K., Takeda,M., Tanabe,H., Nishimura,T., Kobayashi,T., Yamamoto,H., Nonomura,Y., Yoneda,H., Sakai,T., Imagawa,M., Miki,T. and Ogihara,T. (1996) *Clin. Genet.*, **49**, 119–123.
42 Ishihara,M., Ohno,S., Ishida,T., Naruse,T., Kagiya,M., Mizuki,N., Maruya,E., Saji,H. and Inoko,H. (1997) *Tissue Antigens*, **49**, 107–110.
43 Gotoda,T., Yamada,N., Murase,T., Shimano,H., Shimada,M., Harada,K., Kawamura,M., Kozaki,K. and Yazaki,Y. (1992) *J. Lipid Res.*, **33**, 1067–1072.
44 Athma,P., Fidahusein,N. and Swift,M. (1995) *Biochem. Biophys. Res. Commun.*, **210**, 982–986.