Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA

Manjit Kaur and G. Mike Makrigiorgos*

Department of Radiation Oncology, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

Received November 24, 2002; Revised December 23, 2002; Accepted January 16, 2003

ABSTRACT

Errors introduced during PCR amplification set a selectivity limit for microsatellite analysis and molecular mutation detection methods since polymerase misincorporations invariably get confused with genuine mutations. Here we present hairpin-PCR, a new form of PCR that completely separates genuine mutations from polymerase misincorporations. Hairpin-PCR operates by converting a DNA sequence to a hairpin following ligation of oligonucleotide caps to DNA ends. We developed conditions that allow a DNA hairpin to be efficiently PCR-amplified so that, during DNA synthesis, the polymerase copies both DNA strands in a single pass. Consequently, when a misincorporation occurs it forms a mismatch following DNA amplification, and is distinguished from genuine mutations that remain fully matched. Error-free DNA can subsequently be isolated using one of many approaches, such as dHPLC or enzymatic depletion. We present feasibility for the main technical steps involved in this new strategy, conversion of a sequence to a hairpin that can be PCR-amplified from human genomic DNA, exponential amplification from picogram amounts, conversion of misincorporations to mismatches and separation of homoduplex from heteroduplex hairpins using dHPLC. The present hairpin-PCR opens up the possibility for a radical elimination of PCR errors from amplified DNA and a major improvement in mutation detection.

INTRODUCTION

PCR-based amplification is used in almost every aspect of genetic diagnosis, mutation detection and basic research. A major problem with PCR, however, is that all polymerases invariably generate errors during amplification (1–3). Polymerase misincorporations produce slippage errors that complicate microsatellite analysis (4) and set a selectivity limit for molecular mutation detection methods of one mutant sequence in a background of 10^5–10^6 wild-type sequences (5). Thus, high selectivity mutation detection, which relies on PCR, often falls short by one to two orders of magnitude of the selectivity required to investigate mechanisms of spontaneous mutagenesis (6), to identify cancer cells at an early stage (7), to detect mutations in single cells (8) or to reliably identify minimal residual disease (9). In all these applications, polymerase misincorporations invariably become disguised as mutations and result to false positives (10).

Here we present hairpin-PCR a new form of PCR that completely separates genuine mutations from polymerase misincorporations. This opens up the possibility of eliminating misincorporations and generating ‘error-free’ amplified DNA for mutation detection or other applications (Fig. 1A). Hairpin-PCR is based on the observation that, unlike regular DNA hairpins (Fig. 1B), which tend to suppress PCR (11), a DNA hairpin with non-complementary ends (Fig. 1C) can be efficiently PCR-amplified. A DNA sequence that needs to be PCR-amplified is first converted to a hairpin following ligation of an oligonucleotide ‘cap’ on one end and a pair of non-complementary linkers on the other end (Fig. 1A). Next, primers corresponding to the two non-complementary linkers are used to allow DNA polymerase to displace the opposite strand and amplify the entire complement of the hairpin. These primers can optionally overlap the sequence of interest to confer sequence specificity. Following hairpin amplification, the double-stranded PCR product is heat-denatured and placed rapidly on ice to allow the original hairpins to re-form. With subsequent removal of the hairpin caps, the original DNA sequence is recovered. The error-free amplified DNA can then be processed for mutation detection without the limitation of polymerase errors. Below we present feasibility for the main technical steps involved in this strategy, which can lead to a radical elimination of PCR errors from amplified DNA.

*To whom correspondence should be addressed at Brigham and Women’s Hospital, Level L2, Radiation Therapy, 75 Francis Street, Boston, MA 02115, USA. Tel: +1 617 632 6905; Fax: +1 617 632 6900; Email: mmakrigiorgos@partners.org
MATERIALS AND METHODS

Hairpin-forming, long oligonucleotides

Five long oligonucleotides expected to form hairpins were synthesized by Oligos Etc and HPLC-purified (OR, USA). The sequence of hairpins A and D are depicted in Figure 1B–C. Sequences of hairpins B, C and E, which were designed to contain non-complementary ends like hairpin D, were: hairpin B, 5′-ACC GAC GTC GAC TAT CCG GGA ACA CAT GAT TTA AAT GTT TAA ACA CGC GGT GGA CTT AAT TAA CTA GTG CCT TAG TTA GCG TGA TAA ATT AAG TCA CCG CAT GTT TAA ACA TTT AAA TGT ACA GCA CTC TCC AGC CTC TCA CGG CA-3′; hairpin C, 5′-ACC GAC GTC GAC TAT CCG GGA ACA CAA GAT TTA AAT ACA CGC GGT GAC TTA ACA GCC GGC CCT TAA CTA GTG CCT TAG GTA GCG TGA AAG TTA ATT AAG CGC GCC TGT TAA GTA GCC ACC GCC TGT TTA AAC ATT TAA ATC TGG AGC ACT CTC CAG CCT CTC ACC GCA-3′; hairpin E, 5′-ACC GAC GTC GAC TAT CCG GGA ACA GAT CCA TGC ACT GCC CAA CAA CAC CAG CCT TCC CCA GCC AAA GAA GAA ACC ACT GTA TGG AGA AGA TTT CGA CCC TTC AGA AAA CTG AAG GGT CGA AAT ATT CTC CAT CCA GTG TCT TCT TGG GCT GGG GAG AGG AGC TGG TGT TGG GCA GTG CAT GGA TCA GCA CTC AGC TCC AGC TCA CGG CA-3′.

Hairpin-PCR

Designated amounts of hairpins B–D were used in a 25 µl PCR reaction using Titanium Advantage® polymerase (Clontech, Palo Alto, CA) and the primers 5′-GTG AGA GGC TGG AGA GTG CTG CT-3′, forward, and 5′-AGC TCG ACT ATC CGG GAA CAT CCA TGC ACT GCC CAA CAA CAC CAG CCT CTC TCC CCA GCC AAA GAA GAA ACC ACT GTA TGG AGA AGA TTT CGA CCC TTC AGA AAA CTG AAG GGT CGA AAT ATT CTC CAT CCA GTG TCT TCT TGG GCT GGG GAG AGG AGC TGG TGT TGG GCA GTG CAT GGA TCA GCA CTC AGC TCC AGC TCA CGG CA-3′.

dHPLC analysis of oligonucleotide hairpins

To perform separation of mixtures of heteroduplex and homoduplex hairpins, 1 ng hairpins were injected into a WAVE™ dHPLC system (Transgenomics Inc., Cambridge, MA) and run under denaturing conditions at different temperatures, following the company-supplied protocol [Transgenomics Inc. (2002) Single Nucleotide Polymorphism (SNP), Insertion & Deletion on the WAVE® Nucleic Acid Fragment Analysis System; http://www.transgenomic.com/pdf/AN112.pdf]. The dHPLC system was equipped with a fraction collector that allows selection of the elution product according to the DNA retention time on the dHPLC column.

Conversion of native DNA sequences to a hairpin, and PCR amplification

The primers 5′-AGG CCT TCA TGA CTG ATA CCA-3′ (forward) and 5′-TGA GAT CGA CTG AGA CCC CAA-3′ (reverse) were used to amplify from genomic DNA a 137 bp p53 sequence (nucleotides 2215–2352 of Genbank sequence X54156) flanked by TaqI and AluI restriction sites near each end. Following double digestion of this sequence with TaqI (65°C, 1 h) and AluI (37°C, 1 h) the restricted p53 DNA fragment was purified via QiAquick™ centrifugation columns (Qiagen Inc., Valencia, CA) and then ligated to the hairpin-shaped sequences Cap1, 5′-phosphate)-CGACGCGGCGCC-GGCTTAAGTGAGCGCCGCGCCGCGCTG-3′, which ligates TaqI sites and Cap2, 5′-phosphate)-CGACGAGTTCTCG-GCTTTCGATGCTCTGTTGAGUACGCTGACATCTCGT-GAACACAAACTCGGACAG-3′, which ligates AluI (blunt) sites, following the protocol described by Horie and Shimada (12). Briefly, ligation of the two caps to DNA was performed by adding a 100-fold molar excess of each Cap into 10 µM DNA template in the presence of T4 DNA polymerase and incubating the 50 µl reaction volume overnight at 15°C. A 2 µl ligation mixture was then treated with uracil glycosylase (Roche Diagnostics), at 37°C, 30 min, in the company-supplied buffer, 20 µl final volume, in a PCR tube. Upon addition of PCR components and buffer, a reaction was carried out using Titanium polymerase for 35 cycles and the following thermocycling conditions: 94°C, 30 s; (94°C, 30 s/68°C, 60 s) × 25 cycles; 68°C, 60 s; 4°C; Hold. The PCR products were then examined via ethidium-stained agarose gel electrophoresis. Alternatively, PCR amplification was conducted using high fidelity Advantage HF-2R polymerase (Clontech) or Pfu Turbo® (Strategene Inc.). In addition, using the same thermocycling conditions, quantitative real-time PCR in the presence of SYBR Green I dye was performed in a Cepheid i SmartCycler™ machine. Primers used for PCR of hairpin A were 5′-TAA ATG TTT AAA CAC CGC GT-3′, reverse; and 5′-TAA ATG TTT AAA CAC CGC GT-3′ reverse. To amplify picogram amounts of hairpin D spiked into 100 ng human genomic DNA from HL-60 cells, touchdown PCR was applied: 94°C, 30 s; (94°C, 20 s/65°C, 20 s/68°C, 20 s) × 30 cycles, with annealing temperature decreasing 1°C/cycle; (94°C, 10 s/55°C, 20 s/68°C, 20 s) × 15 cycles; 68°C, 6 min; 4°C; Hold.

RESULTS AND DISCUSSION

Amplification of DNA hairpins with non-complementary ends

The observation that, if DNA is amplified in a hairpin structure, mismatches should be almost always separated from mutations, urged the development of hairpin-PCR. Indeed, if the polymerase introduces an A>G mutation on the upper DNA strand it is unlikely that, during synthesis of the bottom
strand of a single hairpin, it will perform the exact opposite error (T>C mutation) at the same position of the complementary strand. Even for a polymerase with a large error rate of $10^{-4}$/base the odds for a double-error event are $10^{-4} \times 10^{-4} = 2.5 \times 10^{-9}$, i.e. less than the expected spontaneous mutation rate in somatic tissues (13,14). On the other hand, practically all genuine mutations should remain fully matched following hairpin-PCR, as these reside in both strands from the beginning (Fig. 1A). This complete discrimination of polymerase errors from the mutations should allow subsequent isolation of error-free amplified hairpins by one of many strategies, such as dHPLC (15), CDCE (16), DGGE (17) or enzymatic depletion of mismatches using mismatch recognition proteins, MutS (18), MutY (19) and TDG (20).

To develop the basic technical aspects of this approach, we designed long oligonucleotides (B, C, D and E, 149, 168, 200 and 218 nucleotides, respectively) expected to form hairpins with non-complementary ends that do not inhibit primer binding at their ends (Fig. 1C), as well as a regular hairpin A, 131 bp, which lacks the non-complementary ends (Fig. 1B), for comparison. Hairpins D and E encompass the complete sequence of p53 exon 9. One nanogram each hairpin was then used in a 25 μl PCR reaction using Titanium Advantage polymerase and primers designed to operate on the non-complementary ends of hairpins B–D, or alternatively on the complementary ends of hairpin A. Hairpins B–D produce a PCR product, while hairpin A does not (Fig. 2A, lanes 1–5). The data indicate that hairpins are readily amplified as long as primers are allowed to bind, and the polymerase is able to synthesize the hairpin complement, presumably by displacing the opposite strand. Omission of either forward or reverse primers abolishes the product (Fig. 2A, lanes 6–7) which indicates that amplification requires both primers and that the full-length hairpin is replicated by the polymerase. Hairpin-PCR was repeated using two proof-reading polymerases, Pfu Turbo, or Advantage-HF2 and amplification was obtained (Fig. 2B). Figure 2C depicts quantitative real-time hairpin-PCR profiles of hairpin D serial dilutions, using SYBR Green I dye. The exponential nature of amplification is evident. Because of the way hairpin-PCR operates (Fig. 1A), the PCR products are expected to result in double-stranded DNA molecules, each strand of which is a full hairpin. To separate the two strands, and to recover the original hairpins, following purification of the PCR product the samples are denatured at 95°C, 1 min, and rapidly cooled by placing them directly on ice. This procedure does not allow time for substantial cross-hybridization of different DNA strands, while each strand is expected to rapidly form a hairpin due to its self-complementary sequence. Figure 2D demonstrates that rapid cooling converts the hairpin amplification product (lanes 1 and 2) to a band approximately half the size (lanes 3 and 4), which corresponds to the expected monomer hairpin. Next, the forward and reverse primers used for the amplifications in Figure 2A were re-designed to encompass an additional nine nucleotide extension (20 + 9 = 29mers) inside the p53 exon 9 hairpin D sequence. Figure 2E demonstrates that, although the
3’ end of the primers falls within the hairpin portion of the sequence, amplification remains almost unhindered. The data are consistent with the occurrence of primer binding by means of the 20 base overlap with the non-complementary end of the hairpin, and that the 3’ end of the primers temporarily displaces the hairpin sequence. This ‘invasion’ by hybridized oligonucleotides at the DNA ends, also reported by Guilfoyle et al. (21), presumably happens frequently enough to allow polymerase binding and primer extension to occur. Therefore restricting the primers on the non-complementary ends amplifies every hairpin sequence that contains those ends, while using primers with 3’ ends extending into the hairpin sequence renders hairpin-PCR sequence specific.

To investigate the amplification efficiency of hairpin-PCR, 100 ng purified human genomic DNA from a cell line that lacks the p53 gene (HL-60 cells), was mixed with decreasing amounts of the p53 exon 9-containing hairpin D. One human genome (~3 × 10⁹ bp) is ~1.5 × 10⁷ times the size of hairpin D, therefore spiking 10⁻² pg hairpin D into 100 ng genomic DNA is approximately equivalent to adding a single p53 exon 9 allele in the genome. Figure 2F and G demonstrate hairpin-PCR amplification of p53 exon 9 using two different polymerases. Amplification from 0.01–0.1 pg hairpin D in the presence of genomic DNA is obtained. The amplification efficiency of hairpin-PCR appears comparable with that of regular PCR.
DHPLC separation of homoduplex from heteroduplex hairpins

To demonstrate that hairpins containing a single base mismatch, such as those expected to result from polymerase misincorporations, can be distinguished from fully matched hairpins via DHPLC, we injected homoduplex hairpin into a WAVE™ DHPLC system equipped with a fraction collector. Two more hairpins were synthesized. These were identical to the homoduplex hairpin except that they were synthesized to contain sequence changes, 56G>A or 46insACA, respectively. Upon folding, these hairpins form mismatches which simulate a potential misincorporation by Taq polymerase (18) and a Taq slippage error (4), respectively. One nanogram of each heteroduplex and homoduplex hairpin was injected separately into DHPLC, or, alternatively, mixed (1:1) and injected as a mixture. At a partially denaturing temperature of 61°C, the peaks from the heteroduplex hairpins could be distinguished from the fully matched, homoduplex hairpin, Figure 2H. Setting the threshold of the fraction collector on the trailing part of the homoduplex peak allows the collection of mainly (70–80%) homoduplex hairpin out of this mixture. This example simulated a worse case scenario, where the heteroduplex DNA was 50% of the overall sample. Normally, however, the heteroduplex peak resulting from PCR errors will be a smaller fraction (~1–10%) of the homoduplex peak (2). From the data in Figure 2H it can be estimated that if PCR errors are confined to 10 or 1% of the sequences, one would collect >95 and >99% homoduplex DNA, respectively, resulting in a radical elimination of heteroduplex hairpins from the mixtures. In DHPLC chromatography almost all possible base changes and PCR errors are detectable [Transgenomics Inc. (2002) Transgenomic Optimase™ Polymerase Delivers Highest Fidelity in PCR for WAVE™ System Analysis (US); http://www.transgenomic.com/pdf/AN119u.pdf], however individual base changes can result to varying degrees of separation of heteroduplexes from the homoduplex peak (15). Nevertheless, homoduplex DNA tends to have the longest retention time on the column (15). By recycling the collected homoduplex through the DHPLC for a second time and by collecting the trailing portion of the homoduplex each time should practically filter-out the misincorporations.

Conversion of native DNA sequences to hairpins and PCR amplification

To enable the scheme in Figure 1A, conversion of a native DNA fragment to a hairpin that can be amplified directly from human genomic DNA is required. To convert a sequence to a hairpin with non-complementary ends we performed ligation of two different oligonucleotide ‘caps’, Cap1 and Cap2, at the positions of two restriction sites encompassing the sequence (Fig. 3A). Cap1 and Cap2 are small oligonucleotides designed to form a hairpin that ligates both top and bottom strands at the respective DNA restriction site (12). In addition, Cap2 contains two centrally located uracils. Following the simultaneous ligation of both caps at the two DNA ends, a treatment with uracil glycosylase removes the uracils and generates abasic sites at the center of Cap2. During the heating step of the subsequent PCR reaction the glycosylase is inactivated and a strand break is expected to form via beta elimination at the abasic sites (22), which allows the hairpin to obtain a structure that can be PCR-amplified. To demonstrate the application, a 91 bp p53 sequence flanked by AluI and NlaIII restriction sites was generated following a double digestion of a larger DNA fragment, which had been first amplified from genomic DNA using regular PCR. Following ligation of caps 1 and 2, the resulting 145 bp fragment was amplified using primers overlapping the non-complementary linkers and the p53 sequence itself. A ~290 bp double-stranded product was generated following a double enzymatic digestion was subjected to the same procedure. A ~290 bp was generated when the full scheme of Figure 3A was applied.
(lane 2) but not when DNA ligase was omitted (lane 3) or when a single primer was used in the hairpin-PCR reaction (lanes 4 and 5). The DNA fragment was then excised from the gel and sequenced. Sequencing verified that the correct sequence had been amplified and that the expected hairpin structure of the amplified sequence had formed.

Both unknown and known mutation detection methods are affected by PCR errors and the most selective methods are those that are affected most. The principal limitation for mutation scanning via constant denaturant capillary electrophoresis (CDCE) is the fidelity of the polymerase used (3,23). High selectivity mutation scanning via DGGE and dHPLC is ultimately limited by polymerase error rate (3,24) [Transgenomics Inc. (2002) Transgenomic Optimase™ Polymerase Delivers Highest Fidelity in PCR for WAVE® System Analysis (US); http://www.transgenomic.com/pdf/AN119u.pdf]. Some of the high selectivity assays for RFLP-based known mutation detection [PCR/RE/LCR (25); Radioactivity-based PCR–RFLP (26); RSM (27,28); APRIL-ATM (29) and others reviewed in (5)] utilize PCR in at least one stage prior to RFLP-selection. Therefore these are also limited by PCR errors (30). The ability to amplify DNA without being limited by polymerase-introduced errors would significantly impact mutation detection and cancer diagnosis. A mismatch-binding protein, MutS, was used previously to deplete mismatches caused by PCR errors, in order to improve DNA synthesis fidelity (18). However, low frequency genuine mutations are also converted to mismatches and eliminated in this process, thus there is no benefit to mutation detection. In contrast, hairpin-PCR converts polymerase errors to mismatches while it also retains mutations in the homoduplex DNA. Forcing the polymerase to keep a double record of the sequence effectively boosts the DNA replication fidelity, as it is unlikely that a misincorporation will happen at the same position in both DNA strands simultaneously. We demonstrated amplification of small (75–145 bp) sequences in hairpin formation. However, polymerases can displace much longer (>1 kb) DNA stretches during synthesis (31) thus it is possible that DNA amplification in a hairpin structure will also occur for larger sequences. On the other hand, if dHPLC-based separation of mismatch-containing hairpins from homoduplexes is applied, an upper DNA size limit of 450–500 bp is imposed, as dHPLC separation is inefficient beyond this (15). Most genotyping/mutation detection methods operate on DNA sizes <500 bp (32), therefore the DNA size limitation should not hinder application of hairpin-PCR in the area of diagnostics. Non-dHPLC-based methods based on enzymatic degradation of heteroduplex hairpins [e.g. via MutS (18); MutY (19); TDG (20,33) or other mismatch-binding proteins] may have to be developed in order to apply this approach to larger DNA fragments.

In summary, we demonstrated that DNA hairpins designed to have non-complementary ends are efficiently PCR-amplified and that dHPLC can discriminate among homoduplex and heteroduplex hairpins. Native DNA sequences can be converted to a hairpin structure and amplified from human genomic DNA. The present hairpin-PCR leads to a strategy for a radical elimination of PCR errors that would allow a major boost to the detection of mutations and STR repeats in human tissue.

ACKNOWLEDGMENTS

The assistance of Hema Liyane and John McCloskey, Transgenomics Inc., Cambridge, MA with dHPLC analysis is gratefully acknowledged.

REFERENCES


