An Unexpected Product From Polymerase Chain Reaction-Mediated Site-Directed Mutagenesis Due to Misalignment of the Mismatched Primer

DARREN W. BAKER, PAUL G. ROTHBERG

Kansas City, Missouri

Background: The I1307K (T3920→A) variant of the *APC* gene has been identified as a potential risk factor for colorectal cancer and is present in 6% of Ashkenazi Jews. Screening for this mutation may allow identification of people at elevated risk who would benefit from increased surveillance.

Methods and Results: We designed an assay to detect the T3920→A allele using a primer mismatched at the 3′ terminal nucleotide in the polymerase chain reaction (PCR) to generate a recognition site for the restriction enzyme *Mse* I. After optimization of the PCR for magnesium ion concentration and annealing temperature, the amplicon did not cut completely with the restriction enzyme in each of four tested DNAs. Sequence analysis of the PCR product that was resistant to digestion revealed that the T3920→A variant was not present. The artifact was caused by a single nucleotide loop-out in the genomic DNA template under the 3′ region of the primer, which allowed the 3′ terminal base of the primer to hybridize properly. As a result, the mismatched primer created a modified product different from that originally planned. At a magnesium ion concentration below the optimum for product yield, most of the product was digested by *Mse* I. Sequence analysis showed that, under these conditions, the intended product was produced.

Conclusions: Mismatched primers can produce unintended products in a PCR due to looping out of a nucleotide in the template or the primer. The magnesium ion concentration can influence the sequence and amount of the product.

Key words: *APC*, DNA diagnosis, artifact, mismatched primer.

One of the most useful methods for rapid detection of known mutations involves the digestion of DNA amplified using the polymerase chain reaction (PCR) with a restriction enzyme that cuts differentially depending on which allele is present. When a useful restriction enzyme is not available, it is usually possible to generate a recognition site by PCR-mediated site-directed mutagenesis using a mismatched primer in the reaction [1]. For example, a mutation in the prothrombin gene that causes a predisposition for thrombophilia is commonly detected using a mismatched primer to generate a *Hin*d III recognition site in DNA with the mutation [2].

We attempted to design such a system for detection of the T3920→A mutation in the *APC* gene. This mutation was found in 6% of Ashkenazi Jews and was associated with an increased risk of colorectal cancer due to creation of a hypermutable region [3]. We designed a primer with an intentional mismatch to create a recognition site for the re-

From the Molecular Genetics Laboratory, Children's Mercy Hospital, Kansas City, Missouri.

Reprint requests: Paul G. Rothberg, PhD, Molecular Genetics Laboratory, Children's Mercy Hospital, 2401 Gillham Rd, Kansas City, MO 64108.

1084-8592/98/0303-0006$8.00/0

striction enzyme *Mse* I in the normal sequence. In the presence of the T3920→A mutation, the *Mse* I recognition site would not be created. This assay failed because of an interesting artifact involving misalignment of the mismatched primer during the early cycles of the PCR.

Materials and Methods

Sequences

We used the *APC* nucleotide sequence from the Entrez nucleotide sequence database (www.ncbi.nlm.nih.gov/Entrez/nucleotide.html) accession M74088 [4] for design of primers and description of the position of the mutations. The sequences used to design primers were checked against accession M73548 [5].

PCR

A 267-bp fragment from exon 15 of the *APC* gene was generated using the PCR. The reaction was done in a volume of 25 μL using 25 pmol of each primer (upstream primer APC18: TTCTGCTAA TACCCTGCAAATAGCAGAAT, and downstream primer APC19: CTAAACATGAGTGGGGT CTCCTGAA), 0.1 μg of genomic DNA, 250 μM each dNTP, 1.25 units of a mixture of Stratagene (La Jolla, CA) Taq2000 DNA Polymerase and Clontech (Palo Alto, CA) TaqStart antibody, and Perkin-Elmer (Foster City, CA) Taq polymerase buffer II (without Mg). $MgCl_2$ was added as indicated from a 25 mM solution. The reaction was cycled 35 times between 96°C for 30 seconds, 55°C for 1 minute and 72°C for 30 seconds, preceded by 2 minutes at 95°C and followed by 5 minutes at 72°C, in an MJ Research thermal cycler (Watertown, MA). Oligodeoxyribonucleotides were obtained from Oligos Etc. (Wilsonville, OR).

Restriction Enzyme Digestion

The PCR product was digested with 4 units of *Mse* I (New England Biolabs, Beverly, MA) by mixing 8 μL of the PCR product with 1 μL of the restriction enzyme and incubating at 37°C for 2 or more hours.

Gel Electrophoresis

PCR products were analyzed by electrophoresis on 8 × 10 cm TBE 10% polyacrylamide minigels, obtained from Novex (San Diego, CA). Gels were stained with ethidium bromide and photographed using ultraviolet transillumination. Agarose gel electrophoresis was used to purify DNA fragments for further analysis. After electrophoresis, the gels were stained with minimal amounts of ethidium bromide and illuminated with a long-wave ultraviolet hand lamp to excise the desired fragments. The DNA was purified from the gel by the GeneClean (Bio 101; La Jolla, CA) procedure. pGEM DNA markers from Promega (Madison, WI) were used to estimate sizes of DNA fragments.

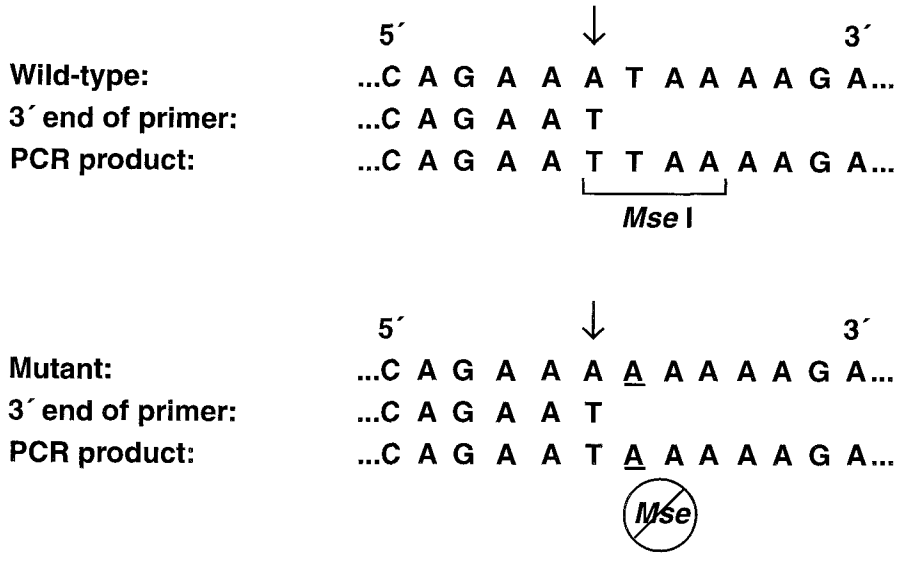

Fig. 1. Strategy for detection of T3920→A in the *APC* gene. The upper panel shows how a mismatch in the primer (arrow) produces an *Mse* I site (TTAA) in the PCR product. The lower panel shows that in the presence of T3920→A, the *Mse* I site is not produced. The variant A in the lower panel is underlined.

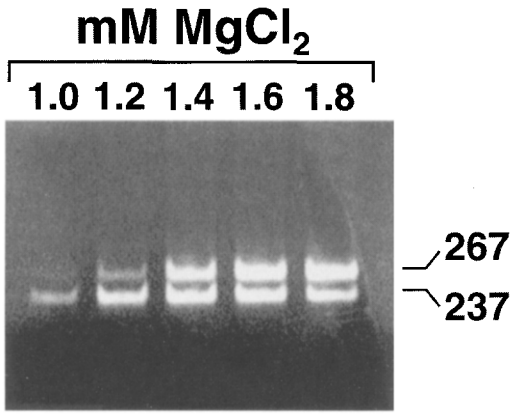

Fig. 2. Incomplete digestion of PCR products with *Mse* I. The PCR was done with the indicated concentration of $MgCl_2$ and a genomic DNA presumed to be wild-type. Next the products were digested with *Mse* I as described in Materials and Methods. The figure shows a gel stained with ethidium bromide and transilluminated with ultraviolet light. The size of the DNA fragments in base pairs is indicated. The pattern expected from *Mse* I digestion of an amplicon generated from wild-type genomic DNA was only seen at a suboptimal $MgCl_2$ concentration (1.0 mM).

DNA Sequencing

DNA generated by the PCR was sequenced using fluorescently labeled dye terminators, AmpliTaq DNA polymerase and the ABI Prism 377 automated DNA sequencer (PE Applied Biosystems; San Jose, CA), by the University of Missouri Kansas City School of Biological Sciences Sequencing Facility.

Results

Our strategy is described in Figure 1. We designed the upstream PCR primer to have a mismatched T at its 3′ terminus to create a modified product with a recognition site for the restriction enzyme *Mse* I (TTAA) in the normal sequence (Fig. 1, top panel). After digestion, the wild-type modified product will be cut into fragments of 237 and 30 bp. In the presence of the T3920→A mutation, the *Mse* I site will not form (Fig. 1, bottom panel), and the modified product will remain uncut at 267 bp. Although an abundant clean product of the expected size of 267 bp was produced at the optimum $MgCl_2$ concentration of 1.6 to 2.0 mM, it would not digest completely with *Mse* I using four different genomic DNAs (data not shown). As shown in Figure 2, when lower levels of $MgCl_2$ were used, the PCR product was much less abundant but was a better substrate for the restriction enzyme. Even at the lowest $MgCl_2$ concentration that gave a product (1.0 mM), we still could not get consistent 100% digestion with *Mse* I. We were unable to produce an adequate quantity of a digestible product despite attempts to optimize the $MgCl_2$ concentration, annealing temperature, and buffer pH. At this point, our goal changed from producing a useful assay to trying to understand a potentially interesting artifact.

We considered it unlikely that all four genomic DNAs tested had a mutation that prevented *Mse* I digestion. We also considered it unlikely that $MgCl_2$ was inhibiting the restriction enzyme because the manufacturer (New England Biolabs) recommends an even higher concentration (10 mM) than the 1.0 to 2.0 mM that we used. Therefore, we hypothesized that the mismatched primer may have generated an unintended mutation during the PCR because of a looping out (bulge loop) of a nucleotide in the genomic DNA template allowing the 3′ terminal T in the primer to base pair with the A (at position 3920) in the template (Fig. 3). To test this hypothesis, we amplified the DNA using a $MgCl_2$ concentration of 2.0 mM, digested the product with *Mse* I, and gel purified the uncut 267 bp

Fig. 3. Proposed mechanism for resistance of the high $MgCl_2$ product to digestion by *Mse* I. The figure shows how a loop-out of a T in the template strand during the early cycles of the PCR will generate a product which lacks the recognition sequence for *Mse* I. Both strands of the PCR product are shown for comparison with the sequence in Figure 4.

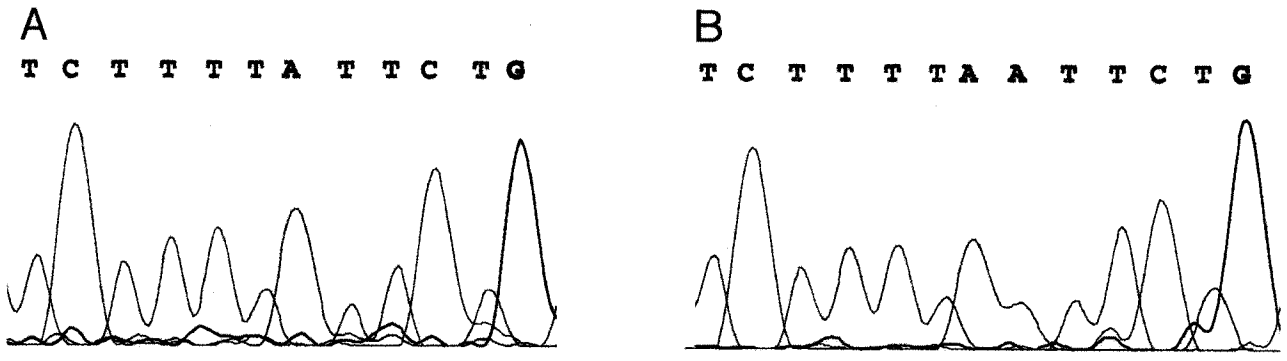

Fig. 4. Sequence analysis of the PCR product produced in high $MgCl_2$ that was resistant to *Mse* I digestion (A), and the PCR product produced in low $MgCl_2$ (B). Electropherogram traces are shown. The downstream PCR primer was used for the sequencing primer, creating a sequence of the lower strand.

product. This DNA was sequenced using fluorescently labeled dye terminators and the downstream PCR primer, APC19. As shown in Figure 4A, the sequence TCTTTTATTCTG is consistent with our hypothesis (lower strand of PCR product in Fig. 3). If the lack of cutting by *Mse* I were due to the T3920→A mutation, the sequence of the uncut product would have been TCTTTTTATTCTG, the reverse complement of the PCR product sequence shown in the bottom panel of Figure 1. As a control, we made the 1.2 mM $MgCl_2$ product, using three PCRs to generate sufficient material, and sequenced it without prior digestion with *Mse* I. The sequence, as shown in Figure 4B, is as expected (Fig. 1, top panel).

Discussion

The original assay for the I1307K mutation in the *APC* gene involved PCR followed by allele-specific hybridization [3]. We attempted to develop an alternate assay to detect this mutation using a mismatched primer to create a diagnostic restriction enzyme site. The assay failed because the PCR did not generate a sufficient quantity of the desired product. The most abundant product did not contain the expected *Mse* I restriction enzyme site. Our data are consistent with this amplicon being due to misalignment of the primer, involving the looping out of a single nucleotide in the genomic DNA template during the early cycles of the PCR (model in Fig. 3). The fact that this mismatched product, rather than the simple mismatched T without a loop-out, was the primary product under optimized conditions was unexpected. Looping out of a nucleotide in the template or the primer must be considered as a potential pitfall in the design of experiments based on mismatched primers for both site-directed mutagenesis and allele-specific PCR.

The most obvious cause of this artifact is the presence of an A residue at the next position of the template strand with which the 3′-terminal mismatched T of the primer could base pair (Fig. 3). However, this arrangement does not necessarily cause a misalignment artifact. A similar strategy to detect the N1303K mutation in the cystic fibrosis gene (CFTR) involved a 3′ terminal mismatched G to create a site for the restriction enzyme *Bst*N I [6,7]. This assay was successful even though the next template base in the CFTR sequence is a C [8].

The sequence context prior to the mismatched base might also have played a role in generating the undesired product. Figure 3 shows a loop-out of the middle T of the three contiguous thymidylate nucleotides in the template. Actually, a single base loop-out of any of these three nucleotides in the template would produce the same product. The presence of three contiguous thymidylates at this position in the sequence probably facilitated the loop-out event.

We also found that the magnesium ion concentration during the PCR had a substantial influence on the sequence of the product. This phenomenon may cause errors in the interpretation of assays that are susceptible to misalignment artifacts. Because we do not know the mechanism of the observed magnesium effect, it is not possible to generalize about its significance. However, it might be exploitable, in some situations, to improve the

specificity of a reaction that fails because of a misalignment artifact.

Acknowledgment

The authors thank Mary Ann Kuchta for preparation of the figures, the University of Missouri Kansas City School of Biological Sciences DNA sequencing facility for assistance in their specialty, and J. Steven Leeder and John F. Bradley for helpful comments on the manuscript.

Received May 1, 1998
Received in revised form June 18, 1998
Accepted July 2, 1998

References

1. Haliassos A, Chomel JC, Tesson L, Baudis M, Kruh J, Kaplan JC, Kitzis A: Modification of enzymatically amplified DNA for the detection of point mutations. Nucleic Acid Res 1989;17:3606
2. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM: A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 1996;88:3698–3703
3. Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW, Vogelstein B: Familial colorectal cancer in Ashkenazim due to a hypermutable tract in *APC*. Nature Genet 1997;17:79–83
4. Kinzler KW, Nilbert MC, Su L-K, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D, Finniear R, Markham A, Groffen J, Boguski MS, Altschul SF, Horii A, Ando H, Miyoshi Y, Miki Y, Nishisho I, Nakamura Y: Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661–664
5. Joslyn G, Carlson M, Thliveris A, Albertsen H, Gelbert L, Samowitz W, Groden J, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes JP, Warrington J, McPherson J, Wasmuth J, LePaslier D, Abderrahim H, Cohen D, Leppert M, White R: Identification of deletion mutations and three new genes at the familial polyposis locus. Cell 1991;66:601–613
6. Ng ISL, Pace R, Richard MV, Kobayashi K, Kerem B, Tsui L-C, Beaudet AL: Methods for analysis of multiple cystic fibrosis mutations. Human Genet 1991; 87:613–617
7. Friedman KJ, Highsmith WE Jr, Silverman LM: Detecting multiple cystic fibrosis mutations by polymerase chain reaction-mediated site-directed mutagenesis. Clin Chem 1991;37:753–755
8. Zielenski J, Rozmahel R, Bozon D, Kerem BS, Grzelczak Z, Riordan JR, Rommens J, Tsui L-C: Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Genomics 1991;10:214–228