SHORT TECHNICAL REPORTS

Manuscripts published in the Short Technical Reports section are shorter and less comprehensive in scope than full Research Reports.

Modulation of Non-Templated Nucleotide Addition by *Taq* DNA Polymerase: Primer Modifications that Facilitate Genotyping

BioTechniques 20:1004-1010 (June 1996)

ABSTRACT

Taq *DNA polymerase can catalyze non-templated addition of a nucleotide (principally adenosine) to the 3¢ end of PCR-amplified products. Recently, we showed that this activity, which is primer-specific, presents a potential source of error in genotyping studies based on the use of short tandem repeat (STR) markers. Furthermore, in reviewing our data, we found that non-templated nucleotide addition adjacent to a 3¢ terminal C is favored and that addition adjacent to a 3¢ terminal A is not. It was clear, however, that features of the template in addition to the 3¢ terminal base also affect the fraction of product adenylated. To define consensus sequences that promote or inhibit product adenylation, we transplanted sequences between the 5¢ ends of the reverse primers of markers that are adenylated and those of markers that are not adenylated. It proved difficult to identify a single sequence capable of protecting the products of all markers from non-templated addition of nucleotide. On the other hand, placing the sequence GTTTCTT on the 5¢ end of reverse primers resulted in nearly 100% adenylation of the 3¢ end of the forward strand. This modification or related ones (called "PIG-tailing") should facilitate accurate genotyping and efficient T/A cloning.*

INTRODUCTION

Short tandem repeat (STR) sequences are found throughout the human genome. Since the length of such a repeat sequence is inherited in a Mendelian manner, STR markers are routinely used for genotyping. Semi-automated, high-throughput genotyping has been achieved by running fluorescently labeled polymerase chain reaction (PCR) products of multiple markers in each lane of a gel and by detecting and sizing the products with an automated DNA sequencer (5,8). In the case of the Applied Biosystems (ABI) DNA sequencer, data collection and allele sizing are based on the use of GENESCAN™ and GENOTYPER™ software. At present, manual editing is required to correct allele misidentification due to variability in non-templated addition of a single nucleotide, mainly adenosine, to the 3′ end of the fluorescently labeled strand by *Taq* DNA polymerase (1).

Recently, we showed that the degree to which a marker is subject to "plus A" modification is primer-specific (7), but we did not define the structural motif(s) underlying this specificity. Instead, we demonstrated that genotyping errors due to variability in adenosine addition could be minimized by amplifying markers in one of two thermal cycling protocols: one that favors production of the "true" allele and another that favors the production of the "plus A" product. In the course of these studies, we discovered a small number of markers highly resistant to adenylation and some additional markers that gave adenylated products even when the PCR conditions used favored production of the true allele. We decided to study these markers in hopes of creating new marker panels with the following features: *(i)* uniform (all or none) plus A modification and *(ii)* a single, rapid PCR cycle for all primers. These features would aid in high-fidelity, high-throughput genotyping.

A solution to the above problem could also be useful in T/A cloning (i.e., cloning of PCR products generated by *Taq* DNA polymerase into vectors having ends with 3′ overhangs comprised of single T's). This cloning strategy has been widely adopted but has proven problematic. Some products fail to be ligated into the vector, and some vector preparations yield unacceptable background levels, which may result, in part, from incomplete T-tailing. We can now predict, to a first approximation, whether a PCR product is clonable, and we suggest a modification in PCR primer design that may increase cloning efficiency.

MATERIALS AND METHODS

PCR Conditions

All PCRs were carried out using 60 ng of template DNA (CEPH individu-

Table 1. Thermal Cycling Protocols

Protocol 1

2-step. 95°C for 5 min followed by 10 cycles of 94°C for 15 s and 55°C for 15 s, followed by an additional 23 cycles of 89°C for 15 s and 55°C for 15 s. Total time: 1 h and 8 min. This protocol favors production of "true" (i.e., non-adenylated) products (7).

Protocol 2

2-step/30-min final extension. As above, followed by a 30-min final extension at 72°C.

Protocol 3

3-step/10-min final extension (ABI PRISM™ protocol). 95°C for 5 min followed by 10 cycles of 94°C for 15 s, 55°C for 15 s and 72°C for 30 s, followed by an additional 20 cycles of 89°C for 15 s, 55°C for 15 s and 72°C for 30 s, followed by a final extension at 72°C for 10 min. Total time: 1 h and 35 min. This protocol favors the generation of "plus A" products (7).

Table 2. Dinucleotide Repeat Markers

Marker	Reverse Primer 5′ Sequence	% Adenylation		
		2-Step PCR	3-Step+10′ PCR	Average
D1S199	CAAAGAC	4	25	14
D16S405	TGAAGGC	12	33	22
D4S398	TCAATTT	15	15	15
D15S131	TTAAAAA	24	29	26
D15S127	AACAGTT	25	65	45
D1S207	GCAAGTC	42	64	53
D10S197	GTGATAC	50	66	58
D5S436	GTCTCCA	61	66	63
D16S511	CAGCCCA	45	80	62
D8S279	GTGTCAG	66	78	72
D1S255	GTGATGG	57	92	75
D5S406	GGGATGC	64	90	76
D6S276	GGGTGCA	71	86	78
D13S173	GTCTCTG	75	82	78

The reverse primers of the markers listed in this table were modified as described in the text. Percent adenylation of the "native" markers was calculated as indicated in the Figure 1 legend. The figures given are means of several amplifications; the figures in the last column are averages of the values in the preceding two columns. Percent adenylation of products that are not modified or products that are heavily modified are fairly consistent from amplification to amplification (7).

als 884-01, -02, -03 or -04; BIOS Laboratories, New Haven, CT, USA). Fifteen-microliter reaction volumes contained 50 mM potassium chloride, 1.5 mM magnesium chloride, 10 mM Tris-HCl, pH 8.3, 333 nM each forward and reverse primer, 0.6 U of AmpliTaq® DNA Polymerase (Perkin-Elmer, Norwalk, CT, USA) and 250 µM each dNTP (dATP, dCTP, dGTP and dTTP).

Dinucleotide Repeat Markers

Fluorescently labeled (6-Fam, Hex or Tet) forward primers of human dinucleotide repeat markers were obtained from Perkin-Elmer/Applied Biosystems Division (PE/ABI, Foster City, CA, USA) or Research Genetics (Huntsville, AL, USA). These were used with reverse primers from the same sources or with modified reverse primers synthesized on a Cyclone™ Plus DNA Synthesizer (MilliGen/Biosearch, Burlington, MA, USA). The primers were purified according to the method of Sawadogo and Van Dyke (6); two consecutive butanol precipitations were performed.

The markers studied are listed in Table 2.

PCR Thermal Cycling Conditions

Three PCR protocols were used with Model 9600 Thermal Cyclers (Perkin-Elmer) (see Table 1).

Analysis of PCR Products

The native or modified markers were used independently to amplify products from single DNA templates, and the 15-µL reaction products were diluted to 100 µL with water. Then, 1.5 µL of each product were mixed with 2.5 µL formamide, 0.5 µL blue dextran loading dye and 0.5 µL internal size standard GS-500 (PE/ABI). The size standard contains DNA fragments fluorescently labeled with the dye phosphoramidite TAMRA (red) ranging in size from 50–500 bp. After heat denaturation at 95°C for 5 min, the tubes were chilled on ice.

Two different ABI DNA sequencers were used: Model 373A and Model 377 (PE/ABI). For the Model 373A, 3.5 µL of the PCR product/size standard mixture were electrophoresed in one lane of a 7% denaturing polyacrylamide gel (Bio-Rad, Hercules, CA, USA) at 15 W constant power (12-cm, well-to-read, filter set B). For the Model 377, 2.5 µL of the mixture were electrophoresed in one lane of a 5% denaturing polyacrylamide gel at 3000 V constant voltage, 2400 scans per hour (36-cm, well-to-read). Fluorescently labeled DNA fragments were analyzed using ABI GENESCAN 672 (Version 1.2.2-1) software on the Model 373A and using ABI GENESCAN (Version 2.0.1 fc2) and ABI PRISM 377 (Version 1.1) collection software on the Model 377 (all from PE/ABI). Genotype data were

generated using ABI GENOTYPER (Version 1.1r8) DNA fragment analysis software.

RESULTS AND DISCUSSION

After listing the markers we studied earlier (7) in order of non-templated nucleotide addition, it was clear that the 5′ base of the reverse primer of resistant (i.e., non-adenylated) markers was frequently T and that the 5′ base of heavily adenylated markers was frequently G. In other words, *Taq* DNA polymerase prefers to add a non-templated 3′ terminal nucleotide adjacent to a C and prefers not to add one adjacent to an A. These data are consistent with earlier primer extension data reported by Hu (3), but exceptions to the rule did occur, prompting us to study the phenomenon further.

We decided initially to look for a consensus sequence to add to the 5′ end of reverse primers that would protect forward-strand products from adenylation. We hoped to be able to routinely use the rapid 2-step Protocol 1 (Table 1) for all amplifications involving such primers. Not only is this protocol faster than the 3-step Protocol 3, (Table 1), but it also typically gives better product yields (7).

We chose to work with two primers having opposite properties: D4S398 and D16S511. About 15% of the PCR product of marker D4S398 is plus A-modified when either Protocol 1 or 3 is used. Marker D 16S511, on the other hand, yields products that are roughly 40% and 75% plus A-modified, respectively, when these protocols are used. When the three (TCA), four (TCAA), five (TCAAT) or six (TCAATT) bases comprising the 5′ end of the reverse primer of marker of D4S398 were added to the 5′ end of the D16S511 reverse primer, there was a progressive decrease in product adenylation using Protocol 3 from 75% for the native marker to 43%, 41%, 39% and 33% for the three, four, five and six base-tailed primers, respectively (data not shown). We sought an optimal sequence by progressively substituting bases at each of the six positions of the tail; the results of these studies are illustrated in Figure 1. A weak consensus emerged for the sequence that resists adenylation: 5′ T (or C, but not A or G); A or C; A, C or G; A, C or G; N; N 3′ (N = A, C, G or T). While the base chosen for positions 5 and 6 seemed to have little impact, the overall length of the added tail was important. Six (or 7) bases were required for best activity. We chose TCACAC, TCAGGG (see Figure 1), TAACTG and TAACTGG as sequences representative of the consensus and added these tails to each of the markers listed in Table 2. All the tails markedly inhibited adenylation of the PCR products, but none was universally effective. In general, markers that were strongly modified to plus A when the native reverse primer was used, were most resistant to the action of the tail. This

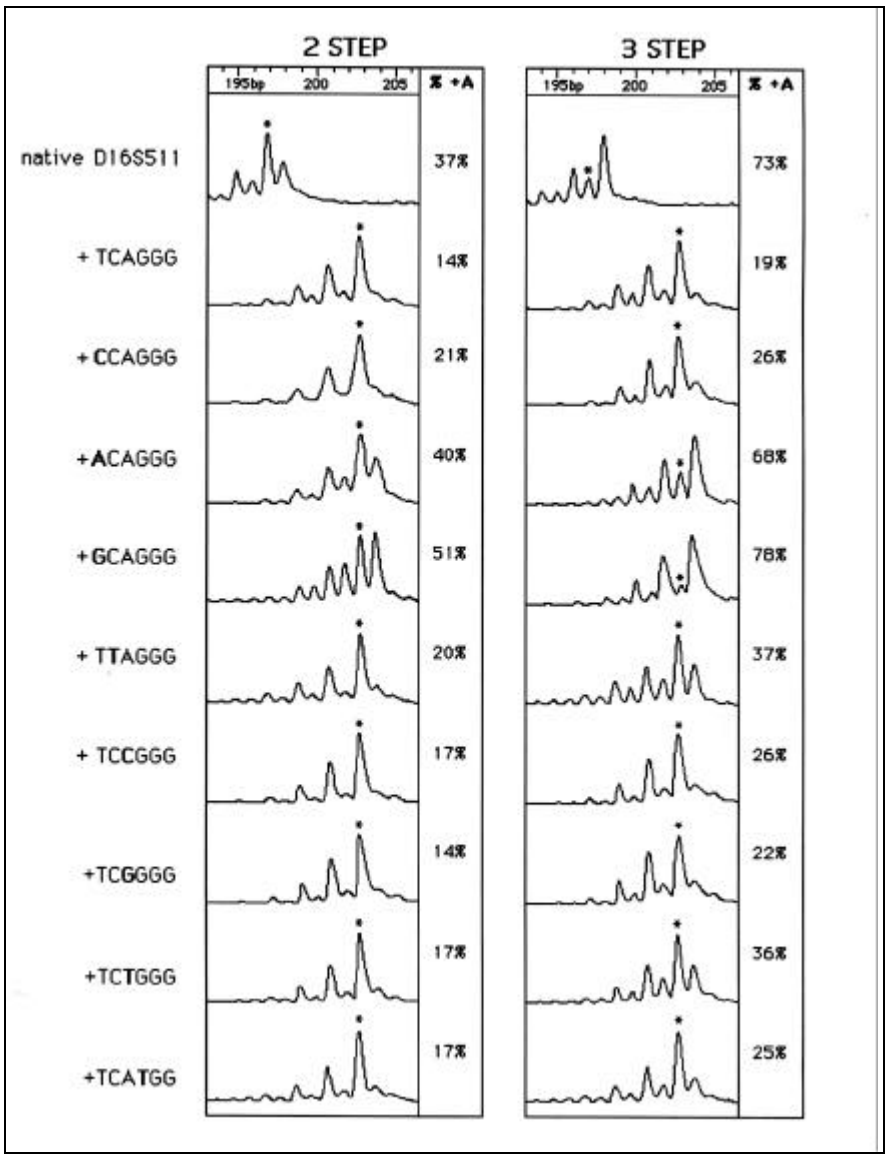

Figure 1. Primer modifications that favor formation of non-adenylated products. The top electrophoretogram shows that marker D16S511 generates non-adenylated 197-bp (63%) and adenylated 198-bp (37%) products when it is amplified using a 2-step PCR protocol (Protocol 1), which does not favor adenylation. Using a 3-step protocol followed by a 10-min final extension (Protocol 3), 73% of the product is adenylated. Addition of TCAGGG to the 5′ end of the reverse primer of marker D16S511 dramatically changes its performance. The 203-bp product generated resists adenylation, even when Protocol 3 is used. Replacing the T in position 1 of the tail by G, A or C (in order of potency) causes a reduction in its ability to protect from adenylation. Similarly, substitutions of T for C in position 2, T for A in position 3 or T for G in position 4 cause some deterioration in the performance of the tail. Asterisks (*) indicate the true (non-adenylated) products. Percent +A is calculated by dividing the height of the +A peak by the sum of the heights of the true and +A peaks.

suggests that *Taq* DNA polymerase interacts with a stretch of base pairs longer than the six or seven encoded by the tails. The mechanism for this is unknown, but it is tempting to speculate that certain nucleotide sequences stabilize conformations of the polymerase that are more or less catalytically active.

Since it was clear that tails would have to be custom-tailored for certain troublesome markers in each primer set to protect them from non-templated addition of nucleotide, we decided to look for a tail that would consistently have the opposite effect. Note that the products of the last five markers in Table 2 could be predicted to be heavily adenylated based on the findings summarized above. In designing a plus A-modifying tail, we placed G on the 5′ end because of our observation that this favors adenylation (Figure 1). We put T in positions 2, 3 and 4 because we found in the studies summarized above that this also favors adenylation (Figure 1). The 5′ sequences of the last six primers in Table 2 fit our design quite well. On the basis of these sequences, we decided on GTTTCT as a reasonable tail to test. Addition of this tail to the reverse primers of the markers listed in Table 2 resulted in nearly complete adenylation of each one when they were amplified using thermal cycling Protocols 1, 2 or 3 (Table 1). An example of this effect is shown in Figure 2. As mentioned earlier, adding a seventh, arbitrarily chosen, base to the tail improves its performance somewhat. This spacer may somehow buffer the tail from the effects of internal sequences.

It is not necessary to add the entire 6- or 7-base tail to all reverse primers. Certain primers (e.g., the last five in Table 2) can be used as is. Other primers should be modified to fit the following consensus: G; T; T; T; (A, C or G). (The assignment of A, C or G in position 5 was determined experimentally, data not shown.) Add GTTT to reverse primers with A, C or G on their 5′ ends; add GTT to primers starting with T followed by A, C or G; add GT to primers ending with TT followed by A, C or G. When the primers listed in Table 2 were modified according to this scheme, they generated products that were greater than 85% adenylated using PCR Protocols 2 or 3. The products were more than 70% adenylated when Protocol 1, the protocol of choice for rapid throughput genotyping, was used.

We have modified more than 50 markers with the plus A consensus tail. All yielded predominantly adenylated products readily identified by the GENOTYPER software. As stated above, there is no need to routinely add a 6- or 7- base tail, but there is no obvious disadvantage in doing so; we have not seen an increase in noise due to nonspecific priming. In fact, tailing the primers has allowed us to alter magnesium concentrations to reduce noise and/or increase signal. The increase in signal results in part from the fact that the entire product is associated with a single, adenylated DNA species.

The results we have presented may also shed light on the problems encountered in doing T/A cloning (inserting adenylated PCR products into vectors with complementary 3′-T overhangs). We have shown that some PCR products resist non-templated addition of nucleotide. Presumably such products would ligate into T-tailed vectors poorly. We feel that modifying reverse and/or forward primers as described above should result in efficient cloning. Similarly, a vector that is completely T-tailed might be produced by linearizing a plasmid to yield blunt ends with 5′ sequences such as GTTTCT. There is no convenient restriction enzyme that can be used to make such a product, so we recommend instead synthesizing an oligonucleotide containing the palindromic sequence AAGACAC/GT-**G**TCTT and inserting it into the plasmid of one's choice. Substituting the G in bold font for T causes little or no deterioration in the performance of the resulting tail (Figure 2). Cutting the above plasmid with *Bbr*PI (Boehringer Mannheim, Indianapolis, IN, USA) or *Pml*I (New England Biolabs, Beverly MA, USA) at the site indicated yields

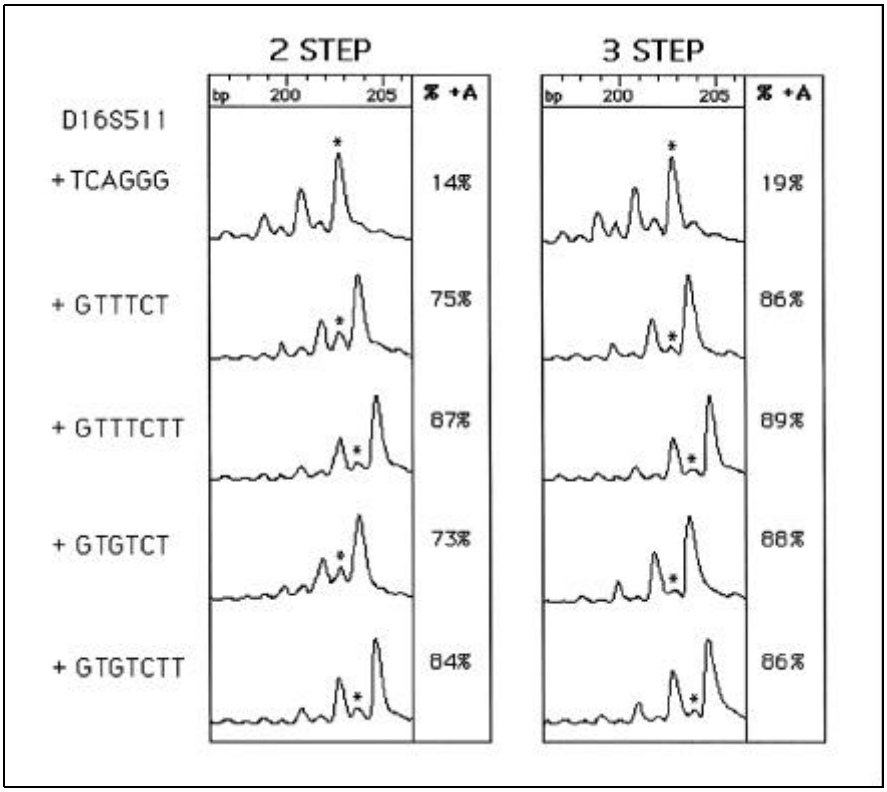

Figure 2. Primer modifications that strongly promote adenylation. Addition of GTTTCT or GTTTCTT to the reverse primer of marker D16S511 has just the opposite effect than addition of TCAGGG. The TCAGGG-tailed primer generates a 203-bp non-adenylated product; GTTTCT- and GTTTCTT-tailed primers generate 204-bp and and 205-bp adenylated products, respectively, even when the PCR conditions used do not favor non-templated addition of nucleotide. Substitution of G for T in the third position of the tail has little or no effect on its performance. Asterisks (*) indicate the non-adenylated products. Percent +A was calculated as described in the Figure 1 legend.

a blunt-ended product that should readily be T-tailed by conventional protocols (2,4).

It may be possible to create pairs of PCR primers, the products of which can be directionally cloned. One would place a tail that resists adenylation on one primer and a tail that promotes it on the other. A vector for cloning such products could be made by inserting the appropriate *Xcm*I site (CCACACG**T**▼GTCTTGG) into the polylinker of any one of several vectors and cutting the vector with *Xcm*I and an enzyme that generates blunt ends (e.g., *Eco*RV or *Sma*I). The resulting vector should have a single T overhang (indicated in bold font) complementary to the single A found on one end of the PCR product.

Finally, note that we refer to the primer modification described in this manuscript as "PIGtailing" because we comprise the Prostate Investigation Group of the National Center for Human Genome Research.

REFERENCES

1. **Clark, J.M.** 1988. Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. *16*:9677-9686.
2. **Holton, T.A. and M.W. Graham.** 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. *19*:1156.
3. **Hu, G.** 1993. DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3′ end of a DNA fragment. DNA Cell Biol. *12*:763-770.
4. **Marchuk, D., M. Drumm, A. Saulino and F.S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. *19*:1154.
5. **Reed, P.W., J.L. Davies, J.B. Copeman, S.T. Bennett, S.M. Palmer, L.E. Pritchard, S.C.L. Gough, Y. Kawaguchi, et al.** 1994. Chromosome-specific microsatellite sets for fluorescence-based semi-automated genome mapping. Nature Genet. *7*:390-395.
6. **Sawadogo, M. and M.W. Van Dyke.** 1991. A rapid method for the purification of deprotected oligodeoxynucleotides. Nucleic Acids Res. *19*:674.
7. **Smith, J.R., J.D. Carpten, M. Brownstein, S. Ghosh, V. Magnuson, D.A. Gilbert, J.M. Trent and F.S. Collins.** 1995. An approach to genotyping errors caused by non-templated nucleotide addition by *Taq* DNA polymerase. Genome Res. *5*:312-317.
8. **Ziegle, J.S., Y. Su, K.P. Corcoran, L. Nie, P.E. Mayrand, L.B. Hoff, L.J. McBride, M.N. Kronick and S.R. Diehl.** 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics *14*:1026-1031.

The authors would like to thank Drs. J. Trent and F. Collins for support, encouragement and helpful discussions. Address correspondence to Michael J. Brownstein, National Institute of Mental Health, National Institutes of Health, Building 36, Room 3A17, Bethesda, MD 20892, USA.

Received 19 September 1995; accepted 13 December 1995.

Michael J. Brownstein, John D. Carpten[1] and Jeffrey R. Smith[2]

National Institute of Mental Health
[1]National Center for Human Genome Research
National Institutes of Health
Bethesda, MD
[2]University of Michigan
Ann Arbor, MI, USA