

and cardiac aging: major shareholders in cardiovascular disease enterprises. Part II. The aging heart in health: links to heart disease. *Circulation* 107:346-354.

18. **Mitchell, R.H., A.H. Bailey, J.M. Anderson, and W.S. Gilmore.** 1992. Electrical stimulation of cultured myocardial cells. *J. Biomed. Eng.* 14:52-56.
19. **Johnson, T.B., R.L. Kent, B.A. Bubolz, and P.J. McDermott.** 1994. Electrical stimulation of contractile activity accelerates growth of cultured neonatal cardiocytes. *Circ. Res.* 74:448-459.
20. **Lauer, M.S.** 2001. Exercise electrocardiogram testing and prognosis: novel markers and predictive instruments. *Cardiol. Clin.* 19:401-414.
21. **Ramamurthy, G., J.E. Kerr, D. Harsha, and M.E. Tavel.** 1999. The treadmill test—where to stop and what does it mean? *Chest* 115:1166-1169.

Received 25 February 2004; accepted 6 April 2004.

Address correspondence to Robert J. Wessells, The Burnham Institute, La Jolla, CA 92037, USA. e-mail: wessells@burnham.org

Locking of 3' ends of single-stranded DNA templates for improved Pyrosequencing™ performance

Michael Utting¹, Jochen Hampe², Matthias Platzer¹, and Klaus Huse¹

¹Institute of Molecular Biotechnology, Jena and ²Christian Albrecht University, Kiel, Germany

BioTechniques 37:66-73 (July 2004)

In Pyrosequencing™, a DNA strand complementary to a single-stranded DNA (ssDNA) template is synthesized, whereby each incorporated nucleotide yields detectable light, and the light intensity is proportional to the incorporated nucleotides. Correct data interpretation (i.e., signal-to-noise ratio of light intensities) is hampered by artifacts due to the formation of secondary structures of single-stranded templates. Critical among these is the looping back of the template's nonbiotinylated 3' end to itself. In the resulting structure, the 3' end functions as a primer, the extension of which results in background signals. We present two ways of preventing the self-priming of a template's 3' end: (i) the use of a modified oligonucleotide, called blOligo, which is complementary to the template's 3' end and (ii) the extension of the template's 3' end with a ddNMP. In contrast to unprotected 3' ends of ssDNA templates, causing inconsistent results, we show that protecting the 3' end of an ssDNA template using either blOligos or ddNMP enables the correct interpretation of signals and results in reliable quantification.

INTRODUCTION

Pyrosequencing™ is a high-throughput method for sequencing and genotyping short DNA fragments. In the sequencing reaction, pyrophosphates that are split off from the deoxynucleoside triphosphates during PCR are quantified by a cascade of enzymatic reactions. At the cascade's end, the indicator reaction catalyzed by luciferase results in detectable light that correlates to the nucleotides initially incorporated. Compared to other genotyping methods, Pyrosequencing is attractive because it allows both single nucleotide polymorphism (SNP) detection within its sequence context and calculation of the amounts of variant alleles in a single experiment. Pyrosequencing is therefore especially useful for estimating allele frequencies with pooled genomic DNA and for detecting allelic imbalances in cDNA samples.

Pyrosequencing is performed isothermally at 28°C with single-stranded DNA (ssDNA) as a template. These conditions favor the formation of secondary structures within the template and may even result in self-priming of the ssDNA 3' ends, which generates ar-

tificial signals and ghost peaks, even in the absence of the sequencing primer (Figure 1A) (1–3).

To circumvent the problem of template folding and 3'-end self-priming, care has to be taken in defining the PCR conditions, mainly with respect to the length of the PCR product and the PCR primer positions on the target DNA. Pyrosequencing AB (Uppsala, Sweden), the manufacturer of the Pyrosequencing hardware, provides software for the design of sequencing primers, which includes the verification of the PCR product's sequence for "template looping onto itself." Furthermore, the single-stranded fragment should not exceed 300 bases because the possibility of self-priming increases with template size. This is a limitation because, in many cases, repetitive sequences around the SNP to be analyzed require larger fragment sizes. The same may be necessary when analyzing cDNA where primers for PCR must span several exons, such as to amplify different splice forms with a single primer pair. Also, for quantitative methylation analysis of CpG sites by Pyrosequencing, the probability of forming 3' loops after ssDNA preparation is high because

the sequence is of low complexity after bisulfite treatment, consisting of essentially 3 bases only.

In such cases, the formation of secondary structures and occurrence of template self-priming have to be overcome by the experimental design. Pyrosequencing using double-stranded DNA (dsDNA) as a template has been previously described, but the incomplete removal of PCR reagents may induce background signals (2,4). Nordstrom et al. (2) therefore developed a protocol for dsDNA Pyrosequencing that included oligonucleotides blocked at their 3' ends by a phosphate or an amino group and proposed the use of such modified oligonucleotides when sequencing with ssDNA templates forming 3'-end loops.

Another possible solution is the addition of ssDNA binding protein (SSB) in the sequencing reaction (1,5–10). However, the binding of SSB to ssDNA varies depending on the DNA sequence and its size (1,11).

Here we present two possible ways to avoid the single-stranded template's self-annealing and priming by blocking

its 3' end. First, an oligonucleotide that is identical to the nonbiotinylated PCR primer but with a dideoxynucleotide at its 3' end can be annealed to the template. This oligonucleotide competes with the ssDNA 3' end for internal annealing and prevents the template's 3' end self-annealing. We call this type of oligonucleotide a *blOligo* (blocking oligonucleotide), according to the designation coined by Nordstrom et al. (2), which cannot be extended during the sequencing reaction due to the presence of the dideoxy residue at its 3' end. Second, the 3' end OH of the template is enzymatically modified with terminal deoxynucleotidyl transferase (TdT; Amersham Biosciences) by incorporating a single dideoxynucleotide. Thus, the 3' end of the modified template will not be elongated.

Figure 1 illustrates the self-priming problem and the two possible solutions. Ideally, no secondary structures interfere with the sequencing reaction (Figure 1B). Deoxynucleotides are attached only to the sequencing primer, and the correct sequence readout is obtained. If the template's 3' end interacts with an

internal complementary sequence (Figure 1A), nucleotides are incorporated at its 3' OH, which falsify the sequence readout because the resulting signals overlay with the sequencing primer elongation. With the use of *blOligo*, the formation of secondary structures is prevented, giving rise to a correct sequence readout (Figure 1C), as is the case with TdT treatment, where the 3' end of the template is locked (Figure 1D).

MATERIALS AND METHODS

Allele-Specific Expression of *CARD15*

Total RNA from leucocytes was isolated using the RNeasy® kit (Qiagen, Hilden, Germany). Reverse transcription was performed using the rtPCR oligo dt™ kit (Qiagen), according to the manufacturer's recommendations. PCR was performed with primers whose annealing sites were localized within exons 2 and 4 of caspase recruitment domain family, member 15 (*CARD15*), respectively. As a control experiment,

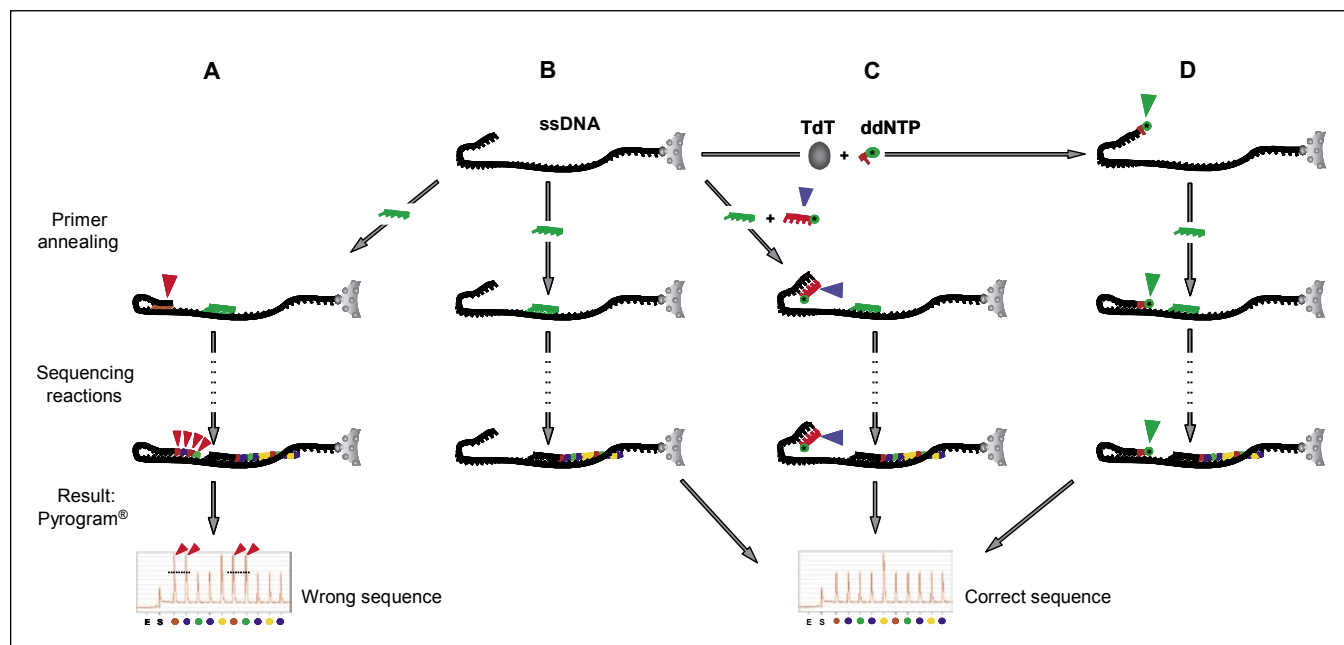


Figure 1. The self-priming problem in Pyrosequencing and two methods of circumventing the problem. (A) Elongation of single-stranded DNA (ssDNA) template 3' end due to self-priming. The formation of a secondary structure of the template causes the elongation of the sequencing primer (green) as well as of the template's 3' end (red arrowheads). An incorrect sequence readout will result. (B) Ideal performance of Pyrosequencing without 3' end self-priming, where only the sequencing primer is elongated. (C) Sequencing reaction in the presence of a *blOligo* (blocking oligonucleotide) designed to avoid self-priming of the template's 3' end. The *blOligo* anneals to the 3' end (blue arrowhead) of the template and, because it has a ddNMP at its own 3' end (green dot), cannot be elongated. (D) Sequencing reaction of a template with ddNMP-modified 3' end. Prior to sequencing, the template's 3' end was modified by a deoxynucleotidyl transferase (TdT)-catalyzed addition of a dideoxynucleotide that will not be elongated even if self-annealing takes place. (B–D) A correct sequence readout is obtained.

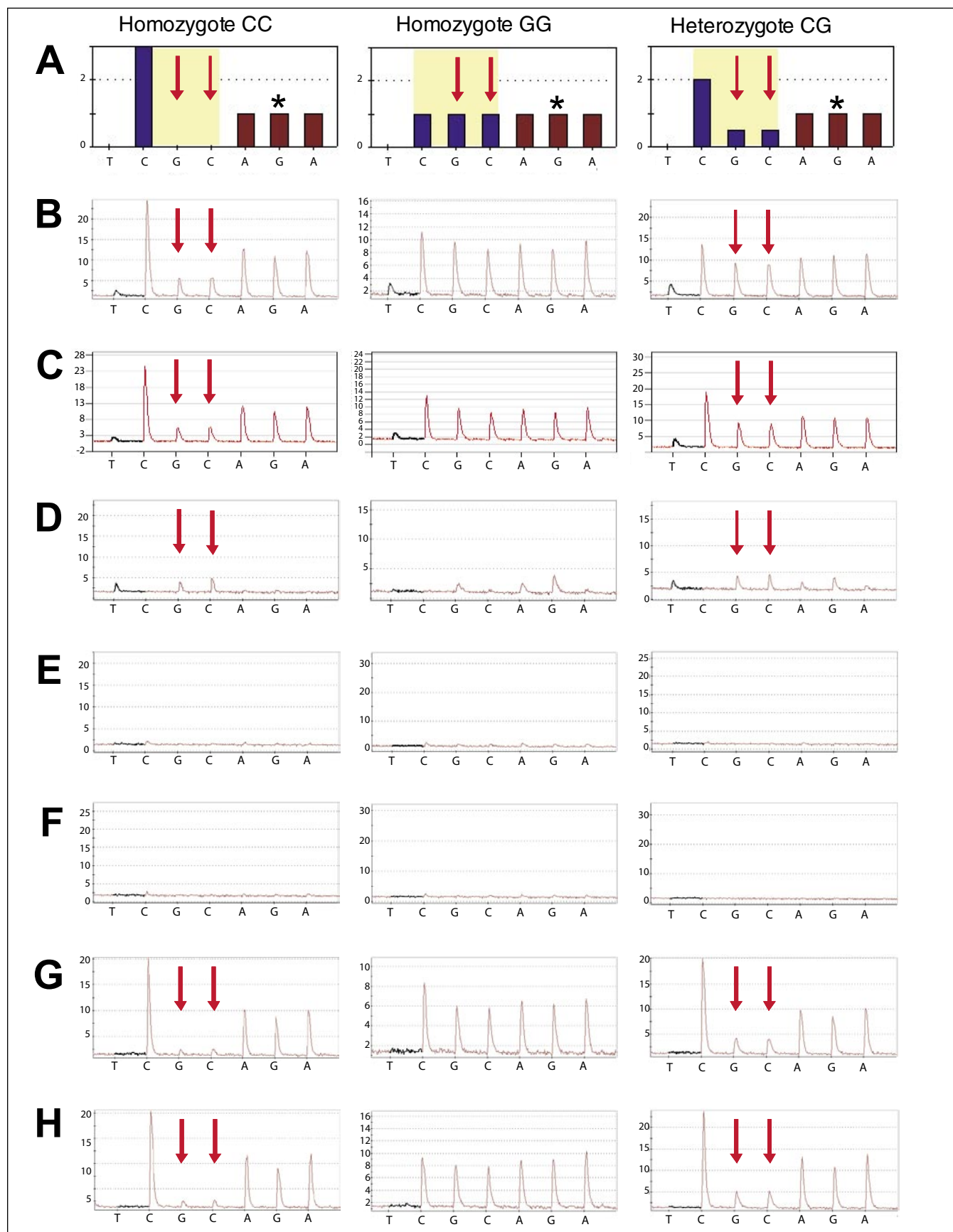


Table 1. Oligonucleotide Sequences

Purpose	Target	Sequence
PCR	<i>CARD15</i> ^a	5'-biotin-CGTCTCTGCTCCATCATAGG-3'
	<i>CARD15</i>	5'-TGGGAGCGGGGTTTCTCAG-3'
	<i>A2M</i> ^a	5'-biotin-GGACTCTAGGTTTCATGCTTCAC-3'
	<i>A2M</i>	5'-TTAGCCCGCCAGGAATTAAC-3'
Pyrosequencing	SNP rs2067085	5'-GTTGCCGATCTTCACACCGT-3'
	<i>CARD15</i>	
	SNP rs226379	5'-TTCCACCCACACAAG-3'
	<i>A2M</i>	
Blocking Self-Priming	<i>CARD15</i>	5'-TGGGAGCGGGGTTTCTCAGddC-3'
	<i>A2M</i>	5'-TTAGCCCGCCAGGAATTAACddC-3'

CARD15, caspase recruitment domain family, member 15; *A2M*, α -2-macroglobulin.

^aForward PCR primers.

the PCR products of *CARD15* exon 2-exon 4 fragments were cloned into the PCR2.1TOPO™ vector (Invitrogen, Karlsruhe, Germany), and the recombinant vector was used as a PCR template instead of cDNA. The calculation of allele amounts occurs in reference to the common G (Figure 2A, labeled with an asterisk).

Allele Frequency of *A2M* SNP rs226379

For the determination of allele frequency of α -2-macroglobulin (*A2M*) SNP rs226379, PCR with pooled genomic DNA (from 80–100 Caucasian individuals; Roche, Mannheim, Germany) was performed according to the manufacturer's instructions. For the calculation of allele frequencies, the PSQ™ 96MA 2.1 software (Pyrosequencing AB) was used.

PCR Conditions

Amplification was done in 25 μ L containing MasterAmp™ 2 $\times$ PCR-Premix-Buffer D (Biozym Scientific GmbH, Oldendorf, Germany), 1 U of *Taq* DNA polymerase (Amersham Biosciences, Freiburg, Germany), 4 pmol

of each primer, and with the following conditions: 1 cycle of 96°C for 5 min, followed by 45 cycles of 96°C for 30 s, 59°C for 35 s, and 72°C for 35 s, with a final cycle of 72°C for 5 min. PCR primer pairs used are listed in Table 1.

Pyrosequencing

Biotin-labeled PCR products were immobilized on 10 μ L streptavidin-coated Dynabeads® M280 (Dyna, Oslo, Norway) by mixing with 20 μ L of PCR product and 30 μ L 2 $\times$ BW buffer II (Pyrosequencing AB). The samples were incubated by shaking at 43°C for 30 min, and afterwards they were transferred into 50 μ L 0.3 M NaOH using the Multi Magnet PSQ 96 Sample Prep Tool (Pyrosequencing AB). The samples were washed in 100 μ L washing buffer (Pyrosequencing AB) for 1 min and transferred into 40 μ L annealing buffer, containing 4 pmol of sequencing primer, and kept at 80°C for 5 min. After equilibration to room temperature, the sequencing reaction was performed with the PSQ 96 SNP Reagent Kit, according to the manufacturer's directions, on a PSQ 96MA machine (both from Pyrosequencing AB).

A blank control was done with

template without sequencing primer. *blOligo* reactions contained 4 pmol of the blocking oligonucleotide with a di-deoxy residue at its 3' end. Sequencing primers and *blOligos* used are listed in Table 1.

TdT Treatment

For TdT (E.C. 2.7.7.31) treatment, Dynabeads with single-stranded templates were incubated in 5 μ L of One-Phor-All-Buffer PLUS (Amersham Biosciences) containing 0.5 mM ddCTP and 2.5 U TdT in a total volume of 50 μ L at 37°C for 30 min. Afterwards, they were washed in 100 μ L washing buffer and treated by Pyrosequencing as described above.

RESULTS AND DISCUSSION

As outlined in Figure 1, self-priming of ssDNA template may falsify the outcome of Pyrosequencing. Figures 2 and 3 give two experimental examples for the self-priming problem. In Figure 2, Pyrosequencing is performed on single-stranded templates obtained from the reverse transcription PCR (RT-PCR) of *CARD15* gene transcripts to analyze allele-specific expression. SNP rs2067085 was used as a marker for the allelic imbalance in heterozygotes reported previously using another SNP (12). Theoretical histograms for homozygotes and heterozygotes are clearly informative and distinguishable (Figure 2A). However, using cDNA of a CC homozygotic individual (all individuals were genotyped using genomic DNA and Sanger sequencing), ghost peaks appear at positions informative for the G allele, although there is no G allele present (Figure 2B). For the heterozygotic individual, the SNP determining peaks appear too high, indicating an imbalance in allelic expression (Figure

Figure 2. Characterization of a coding single nucleotide polymorphism (SNP) on *CARD15* transcripts by Pyrosequencing. SNP rs2067085 was analyzed on cDNA from leucocytes of three genotypically different individuals. (A) Theoretical histograms for the two homozygotes and the heterozygote with the chosen pipetting scheme. The histogram for the heterozygote was calculated for an equal expression of both alleles. Bases informative for the SNP alleles are shown in blue while bases in common are shown in brown. (B) Experimental Pyrograms obtained by using the standard protocol and cDNA from leucocytes. (C) Experimental Pyrograms obtained by using the standard protocol and cloned PCR products. (D) Results from blank control experiments in which Pyrosequencing was performed with single-stranded template without the addition of sequencing primer. (E) Blank control as in row D with the addition of *blOligo* (blocking oligonucleotide). (F) Blank control as in row D with the template ddCMP modified by deoxynucleotidyl transferase (TdT) treatment. (G) Pyrograms obtained by using *blOligo* and sequencing primer. (H) Pyrograms obtained by using ddCMP-modified template and sequencing primer. Red arrows indicate allele-specific bases used, and the asterisks (*) indicate the common G that was used as a reference for the calculation of allele amounts. *CARD15*, caspase recruitment domain family, member 15.

Table 2. Analysis of Allele-Specific Expression of *CARD15* in Leucocytes of Four Individuals

	Template	Allele Frequency G	Allele Frequency C	G:C	Σ G + C
Individual 1	standard	0.62	0.48	1.29	1.10
	<i>blOligo</i>	0.38	0.72	0.52	1.10
	ddNMP	0.39	0.65	0.60	1.04
Individual 2	standard	0.72	0.49	1.50	1.21
	<i>blOligo</i>	0.57	0.58	0.98	1.15
	ddNMP	0.46	0.57	0.81	1.03
Individual 3	standard	0.81	0.54	1.50	1.35
	<i>blOligo</i>	0.47	0.58	0.81	1.05
	ddNMP	0.48	0.52	0.92	1.00
Individual 4	standard	0.88	0.58	1.52	1.46
	<i>blOligo</i>	0.53	0.68	0.78	1.21
	ddNMP	0.52	0.54	0.96	1.06

Values were calculated through the analysis of single nucleotide polymorphism (SNP) rs2067085 on caspase recruitment domain family, member 15 (*CARD15*) cDNA by Pyrosequencing. For each individual, Pyrosequencing was performed using the standard procedure and after locking the 3' ends of PCR products with the *CARD15 blOligo* (blocking oligonucleotide). Bold numbers indicate optimal Pyrosequencing performance after deoxynucleotidyl transferase (TdT) treatment.

2B, right column). The same problems were also detectable for cDNA, with templates obtained from cloned PCR products. Ghost peaks also appeared at the SNP determining positions and falsified the results in a similar way (Figure 2C).

A blank control experiment in which only the template was used in the Pyrosequencing reaction (i.e., without sequencing primer) indicated that self-priming of the template's 3' end results in the occurrence of signals at positions of the informative bases (Figure 2D). Although the Pyrograms differ slightly for the three genotypes, ghost peaks are consistent at positions of the bases informative for SNP rs2067085. Beside the examples presented, in our hands, such blank control experiments reveal a 3'-end self-priming for nearly all inspected sequences and are an indispensable control when setting up Pyrosequencing analyses.

If the *CARD15 blOligo* was added during Pyrosequencing, no signals at all were obtained with either template (Figure 2E). Consequently, if the *blOligo* was used together with the sequencing primer, the Pyrogram gave the clear-cut expected result (Figure 2G). With the same targets, we tested the second possible method of locking the 3' end of the single-stranded template; in this case, by attaching a

ddCMP using TdT. As with the use of the *blOligo*, the Pyrosequencing reactions with the TdT-modified templates obtained from the cDNA and cloned PCR fragments indicated no self-priming in the blank control (Figure 2F) and gave the expected result in the analysis (Figure 2H).

Aside from SNP typing, quantitative data analysis is the main advantage of Pyrosequencing. The quantification is done by the calculation of relative peak heights. Informative bases are compared with a base common to all templates (calibration to an adenine should be avoided because the A-peaks often appear too high).

Table 2 summarizes the quantitative analysis of allele-specific expression of *CARD15* transcripts in four individuals. By standard Pyrosequencing, the results indicate an allelic imbalance in all samples with the G allele being overexpressed. However, in every case, the total amount of the allele-specific bases is conflicting with the amount of a common base. They sum to values larger than 1, which indicates the presence of background signals (Table 2, right column). Both the use of *blOligo* and TdT treatment shift the results significantly so that Σ G + C becomes closer to 1. If we accept a maximum deviation from 1.0 of 10%, which is obtained in all cases of TdT treatment, an

allelic imbalance is seen for only two of the samples and, furthermore, with a favored C allele expression.

Self-annealing of single-stranded template was also the problem with the determination of allele frequency of SNP rs226379 in the *A2M* gene (Figure 3). A ghost peak at the SNP determining position (Figure 3B) falsified the calculation of allele frequency in a pooled sample, resulting in a minor allele frequency of about 8% (Figure 3C). However, after locking 3' ends of the single-stranded template by the *A2M blOligo* or by TdT treatment, the PCR fragments indicate no self-priming, and the calculation of the minor allele frequency results in about 22%, which is in much closer agreement with the National Center for Biotechnology Information Single Nucleotide Polymorphism Database (NCBI) dbSNP; http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=226379) (average allele frequency within a Caucasian sample set of 125 individuals of the C allele: 0.34) (Figure 3, D and E).

In summary, we demonstrated that locking the 3' end of single-stranded templates considerably improves the accuracy of allele calling by Pyrosequencing. It also allows the user to design PCR primers more or less independently of the methodology of Pyrosequencing. This is highly advantageous when a PCR has been already established in mutational profiling of a genomic region because the respective primer pairs from the screening can be used immediately for Pyrosequencing (with one of the primers being biotinylated). Thus, a special PCR design for Pyrosequencing is not necessary. Both the *blOligo* and TdT treatment gave similar results, and the method chosen may depend on different parameters.

The *blOligo* has to be custom-made for each kind of template. However, it has the advantage of no additional hands-on time during the sequencing process because it anneals together with the sequencing primer. The principle of locking the 3' end by a *blOligo* and avoiding self-priming of the template can also be applied to the destruction of secondary structures of single-stranded templates (e.g., hairpins). Moreover, an additional *blOligo* could be annealed a few bases downstream of

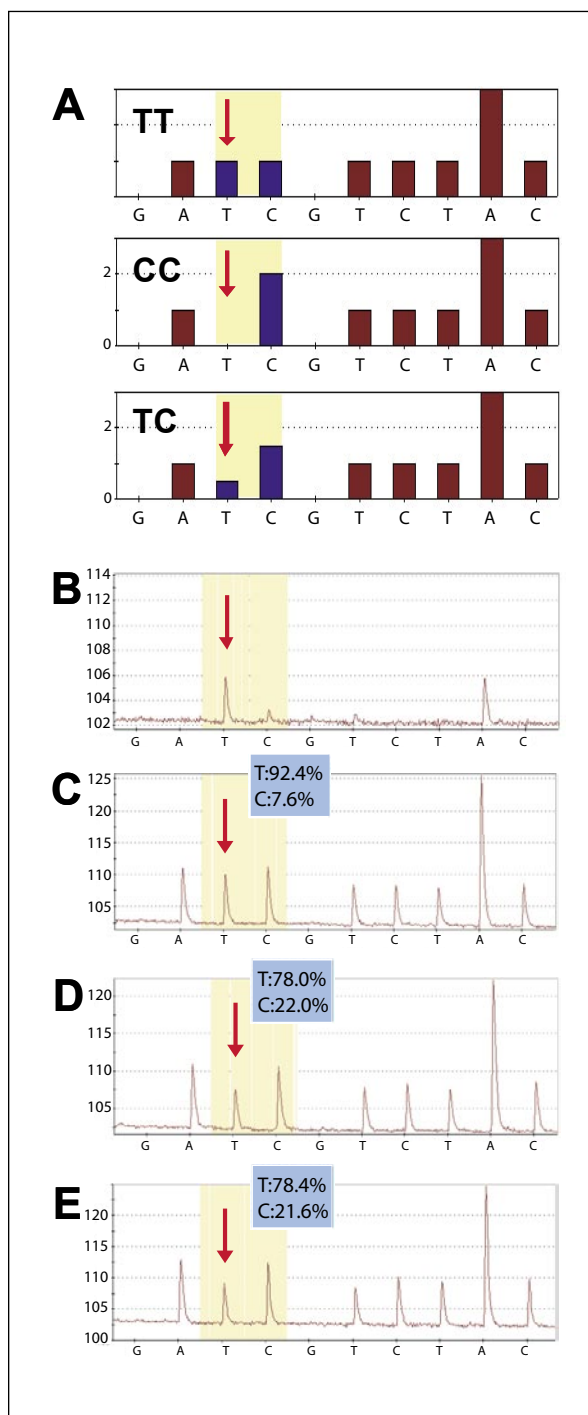


Figure 3. Determination of allele frequencies of the A2M gene single nucleotide polymorphism (SNP) rs226379 in pooled genomic DNA. The sequence around SNP rs226379 was analyzed by the Pyrosequencing of a single sample containing pooled genomic DNA. (A) Theoretical histograms for the genotypes determined by SNP rs226379. The positions of the bases informative for the SNP are highlighted. (B) Pyrogram of a blank control experiment of single-stranded template without sequencing primer. (C) Pyrogram obtained using the standard protocol. (D) Pyrogram obtained using the α -2-macroglobulin (A2M) *bOligo* (blocking oligonucleotide). (E) Pyrogram using ddCMP-modified template. Red arrows indicate allele-specific base. The blue insets show the allele frequencies calculated from the peak heights.

the SNP under investigation to prevent the formation of any secondary structures.

On the other hand, TdT treatment is universally applicable. However, it needs additional hands-on time for incubation. Nevertheless, as a major advantage, it will block the 3' end of any DNA or oligonucleotide in the reaction mixture. Thus, the TdT treatment will avoid even the falsifying background signals generated by incomplete or recombinant PCR products (13) in pooled genomic or cDNA samples, occurring even if the *bOligo* is added because these molecules lack the 3' end targeted by the *bOligo*. In such cases, and in particular for bisulfite-treated DNA, which is hampered by its lower complexity, the TdT treatment will be the method of choice.

In conclusion, the two methods presented for blocking of ssDNA 3' ends are simple and reliable approaches for avoiding misinterpretation in genotyping, quantifying allelic imbalance or allele frequencies, and sequencing when using Pyrosequencing.

REFERENCES

1. Ehn, M., A. Ahmadian, P. Nilsson, J. Lundberg, and S. Hober. 2002. Escherichia coli single-stranded DNA-binding protein, a molecular tool for improved sequence quality in pyrosequencing. *Electrophoresis* 23:3289-3299.
2. Nordstrom, T., A. Alderborn, and P. Nyren. 2002. Method for one-step preparation of double-stranded DNA template applicable for use

- with Pyrosequencing technology. *J. Biochem. Biophys. Methods* 52:71-82.
3. Ronaghi, M., B. Pettersson, M. Uhlen, and P. Nyren. 1998. PCR-introduced loop structure as primer in DNA sequencing. *BioTechniques* 25:876-884.
4. Nordstrom, T., K. Nourizad, M. Ronaghi, and P. Nyren. 2000. Method enabling pyrosequencing on double-stranded DNA. *Anal. Biochem.* 282:186-193.
5. Andreasson, H., A. Asp, A. Alderborn, U. Gyllenstein, and M. Allen. 2002. Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. *BioTechniques* 32:124-133.
6. Garcia, C.A., A. Ahmadian, B. Gharizadeh, J. Lundberg, M. Ronaghi, and P. Nyren. 2000. Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* 253:249-257.
7. Nordstrom, T., B. Gharizadeh, N. Pourmand, P. Nyren, and M. Ronaghi. 2001. Method enabling fast partial sequencing of cDNA clones. *Anal. Biochem.* 292:266-271.
8. Gharizadeh, B., M. Ghaderi, D. Donnelly, B. Amini, K.L. Wallin, and P. Nyren. 2003. Multiple-primer DNA sequencing method. *Electrophoresis* 24:1145-1151.
9. Rickert, A.M., A. Premstaller, C. Gebhardt, and P.J. Oefner. 2002. Genotyping of SNPs in a polyploid genome by pyrosequencing. *BioTechniques* 32:592-600.
10. Ronaghi, M. 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286:282-288.
11. Bhattacharya, S., M.V. Botuyan, F. Hsu, X. Shan, A.I. Arunkumar, C.H. Arrowsmith, A.M. Edwards, and W.J. Chazin. 2002. Characterization of binding-induced changes in dynamics suggests a model for sequence-nonspecific binding of ssDNA by replication protein A. *Protein Sci.* 11:2316-2325.
12. Yan, H., W. Yuan, V.E. Velculescu, B. Vogelstein, and K.W. Kinzler. 2002. Allelic variation in human gene expression. *Science* 297:1143.
13. Judo, M.S., A.B. Wedel, and C. Wilson. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* 26:1819-1825.

Received 19 February 2004; accepted 1 April 2004.

Address correspondence to Michael Utting, Institute of Molecular Biotechnology Jena, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany. e-mail: mutting@imb-jena.de