

Washington University School of Medicine

**Digital Commons@Becker**

---

Open Access Publications

---

9-1-1998

## **Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing**

H Zakeri

G Amparo

S. M. Chen

S Spurgeon

P. Y. Kwok

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

---

# Short Technical Reports

## SHORT TECHNICAL REPORTS

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

### Peak Height Pattern in Dichloro-Rhodamine and Energy Transfer Dye Terminator Sequencing

*BioTechniques 25:406-414 (September 1998)*

#### ABSTRACT

*Establishing the pattern in peak heights within local sequence contexts improves the accuracy of base calling and the identification of DNA sequence variations in dye-terminator cycle sequencing. We have systematically examined pairs of sequence-tagged sites (STSs) that vary at only a single nucleotide to determine how base changes influence the peak heights of neighboring bases in sequencing traces generated by two recently commercialized dye-terminator chemistries, the dichloro-rhodamine (dRhodamine) and the energy transfer (BigDye<sup>®</sup>) terminators. For sequencing traces generated with the dRhodamine terminators, the peak height of a particular base in 28 of 64 possible 3-base windows (44%) can be predicted by knowing just one or two bases 5' to the base in question. For those generated with the BigDye terminators, the peak height of a particular base in 23 of 64 possible 3-base windows (36%) can be predicted by knowing the local sequence context. When the peak heights are binned slightly differently, 75% (48 out of 64 cases) of the base peaks generated by both dRhodamine and BigDye terminators fall in the middle half, confirming that the peak patterns of these two new dye terminator chemistries are much more even than those found in the original rhodamine dye terminator sequences.*

#### INTRODUCTION

Among the available sequencing chemistries, cycle sequencing with thermostable DNA polymerases and dye-labeled dideoxy chain terminators is the most versatile because it eliminates the need for specially modified sequencing primers, and it requires only one sequencing reaction for each DNA sample (6,7). A major drawback of the dye-terminator chemistry has been the highly uneven peak patterns produced

that are dependent on local sequence context and the DNA polymerase used (3,4). Recently, two new classes of dye terminators, the dichloro-rhodamine (dRhodamine) dyes and energy transfer dichloro-rhodamine (BigDye) dyes, have become available commercially (2,5). These two new classes of dye terminators have improved spectral properties and give much more even peaks in cycle sequencing (5).

We have conducted a study of peak-height patterns for the dRhodamine dye and BigDye terminators in cycle sequencing. The results are compared to those obtained previously for the "original" dye terminators with AmpliTaq<sup>®</sup> DNA Polymerase, FS, which uses a set of rhodamine dyes. As shown below, the peak-pattern trends obtained with these new dye terminators are more predictable than those for the AmpliTaq, FS/rhodamine dye terminators. Knowledge of these trends increases the accuracy and confidence in identifying polymorphisms and mutations in a sequencing trace and makes it easier to edit sequences obtained using these chemistries.

#### MATERIALS AND METHODS

##### PCR Amplification and Purification of PCR Products

Human genomic DNA (16 ng) from individuals with known genotypes were amplified in 40- $\mu$ L reactions. The polymerase chain reaction (PCR) products were gel-purified and eluted into 50  $\mu$ L of water as previously described in detail (3,4). The purified PCR product was used directly as sequencing template without further characterization.

##### Taq-FS Cycle Sequencing with Dye-Labeled Terminators

Cycle sequencing was performed on the GeneAmp<sup>®</sup> PCR System 9600 (Perkin-Elmer, Norwalk, CT, USA) using the ABI PRISM<sup>™</sup> BigDye Terminator and ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq<sup>®</sup> DNA Polymerase, FS (PE Applied Biosystems, Foster City, CA, USA) according

**Table 1. Effect of 5' Bases on Peak Heights of the 3' Base in dRhodamine Dye Terminator Sequences**

Base String <sup>a</sup>	Mean Peak Height (mm) <sup>b</sup>	Range (mm)	Cases Analyzed	Small Peaks (%) <sup>c</sup>	Average Peaks (%) <sup>d</sup>	Large Peaks (%) <sup>e</sup>
1. <b>ACA</b>	19.5 ± 3.4	14–28	28		27 (96%)	1 (4%)
2. <b>ACC</b>	15.7 ± 2.3	13–21	21		20 (95%)	1 (5%)
3. <b>AGC</b>	20.5 ± 3.5	15–29	30		28 (93%)	2 (7%)
4. <b>AGG</b>	36.0 ± 4.5	26–40	31		1 (3%)	30 (97%)
5. <b>ATA</b>	17.4 ± 2.2	14–23	35		35 (100%)	
6. <b>ATT</b>	20.4 ± 4.3	13–34	41		38 (93%)	3 (7%)
7. <b>CAG</b>	9.6 ± 2.3	5–16	40	36 (90%)	4 (10%)	
8. <b>CCA</b>	21.9 ± 2.7	17–30	31		30 (97%)	1 (3%)
9. <b>CCC</b>	18.4 ± 2.3	13–23	31	1 (3%)	30 (97%)	
10. <b>CCT</b>	16.8 ± 2.7	12–24	43	3 (7%)	40 (93%)	
11. <b>CTA</b>	21.6 ± 2.7	16–26	24		24 (100%)	
12. <b>CTG</b>	21.5 ± 4.4	13–34	36		33 (92%)	3 (8%)
13. <b>GCA</b>	20.2 ± 3.2	13–26	29		29 (100%)	
14. <b>GCC</b>	16.1 ± 1.9	11–21	26	2 (8%)	24 (92%)	
15. <b>GCG</b>	10.0 ± 2.1	7–14	12	11 (92%)	1 (8%)	
16. <b>GGT</b>	20.8 ± 3.8	15–29	20		18 (90%)	2 (10%)
17. <b>GTC</b>	23.7 ± 3.5	18–32	20		17 (85%)	3 (15%)
18. <b>GTT</b>	20.0 ± 4.0	14–29	26		24 (92%)	2 (8%)
19. <b>TAC</b>	11.0 ± 1.5	8–14	27	26 (96%)	1 (4%)	
20. <b>TAG</b>	7.2 ± 1.9	3–12	28	28 (100%)		
21. <b>TCA</b>	21.7 ± 3.4	14–31	37		34 (92%)	3 (8%)
22. <b>TCG</b>	9.3 ± 2.4	6–14	12	11 (92%)	1 (8%)	
23. <b>TGA</b>	19.5 ± 3.3	12–29	39	1 (3%)	37 (94%)	1 (3%)
24. <b>TGC</b>	19.6 ± 2.3	16–24	24		24 (100%)	
25. <b>TGG</b>	21.3 ± 4.1	15–33	46		42 (91%)	4 (9%)
26. <b>TTA</b>	21.3 ± 2.9	15–29	44		42 (95%)	2 (5%)
27. <b>TTC</b>	23.5 ± 3.5	15–34	39		33 (85%)	6 (15%)
28. <b>TTT</b>	20.6 ± 3.6	14–31	61		57 (93%)	4 (7%)

<sup>a</sup>The 3' base is in bold type.

<sup>b</sup>Mean peak height followed by standard deviation.

<sup>c</sup>The cases for a particular window where the 3' base has a peak height of ≤1/3 of full scale (≤13 mm, designated as small).

<sup>d</sup>The cases for a particular window where the 3' base has a peak height of 1/3–2/3 of full scale (14–26 mm, designated as average).

<sup>e</sup>The cases for a particular window where the 3' base has a peak height of ≥2/3 of full scale (≥27 mm, designated as large).

# Short Technical Reports

Table 2. Effect of 5' Bases on Peak Heights of the 3' Base in BigDye Terminator Sequences

Base String <sup>a</sup>	Mean Peak Height (mm) <sup>b</sup>	Range (mm)	Cases Analyzed	Small Peaks (%) <sup>c</sup>	Average Peaks (%) <sup>d</sup>	Large Peaks (%) <sup>e</sup>
1. AAA	18.0 ± 3.0	12–27	60	3 (5%)	57 (95%)	
2. ACA	21.1 ± 4.4	14–31	30	4 (13%)	26 (87%)	
3. ACC	21.6 ± 4.1	14–28	20		19 (95%)	1 (5%)
4. ATA	18.0 ± 2.6	14–25	36		36 (100%)	
5. ATT	16.4 ± 3.3	11–25	39	5 (13%)	34 (87%)	
6. CAA	16.1 ± 2.6	10–21	24	3 (12%)	21 (88%)	
7. CCA	20.8 ± 3.0	15–28	33		31 (94%)	2 (6%)
8. CGA	20.6 ± 4.6	14–28	14		13 (93%)	1 (7%)
9. CGG	19.5 ± 5.1	13–31	15		13 (87%)	2 (13%)
10. CTA	21.8 ± 4.2	14–33	23		22 (96%)	1 (4%)
11. CTC	11.2 ± 2.8	6–20	42	37 (88%)	5 (12%)	
12. CTT	17.3 ± 4.1	11–29	52	5 (10%)	46 (88%)	1 (2%)
13. GAA	19.4 ± 3.6	13–28	33	1 (3%)	31 (94%)	1 (3%)
14. GCA	18.5 ± 3.6	12–28	31	2 (6%)	28 (91%)	1 (3%)
15. GGA	21.2 ± 4.3	14–32	25		22 (88%)	3 (12%)
16. GGC	22.2 ± 4.0	14–31	35		30 (86%)	5 (14%)
17. GGG	17.2 ± 4.0	11–28	30	3 (10%)	26 (87%)	1 (3%)
18. GTG	21.6 ± 3.8	15–31	32		29 (91%)	3 (9%)
19. TAA	18.6 ± 3.8	12–27	39	4 (10%)	35 (90%)	
20. TCC	20.8 ± 4.2	14–31	30		27 (90%)	3 (10%)
21. TGA	21.5 ± 4.2	13–29	41	1 (2%)	39 (96%)	1 (2%)
22. TGG	20.5 ± 4.8	12–34	45	1 (2%)	40 (89%)	4 (9%)
23. TTA	21.8 ± 4.5	12–33	42	1 (2%)	36 (86%)	5 (12%)

<sup>a–e</sup>See Table 1 legend.

to the manufacturer's directions. Briefly, the gel-purified DNA (10.4 µL) was added to a MicroAmp<sup>®</sup> reaction tube (Perkin-Elmer) containing 3.2 pmol of sequencing primer and 8.0 µL of premixture (containing buffer, dNTPs, dye-labeled ddNTPs and *Taq*-FS/pyrophosphatase). After the initial denaturation at 96°C for 2 min, the reaction mixture was incubated for 25 cycles at 96°C for 15 s, 50°C for 1 s and 60°C for 4 min. Excess dye-labeled terminators were removed from the extension products by spin column purification (CentriSep<sup>™</sup> spin columns; Princeton Separations, Adelphia, NJ, USA) according to the manufacturer's directions. [Note: The excess dyes can

also be removed successfully by ethanol-precipitation, which avoids the labor and expense associated with spin-column purification.] Once separated, the extension products were evaporated to dryness under reduced pressure (SpeedVac<sup>®</sup>; Savant instruments, Holbrook, NY, USA). Each sample was resuspended in 3 µL (for dRhodamine dye terminator) or 12 µL (for BigDye terminators) of loading buffer (5:1 deionized formamide/25 mM EDTA, pH 8.0, 50 mg/mL blue dextran) and heated for 5 min at 90°C. Aliquots (1.6 µL) of the extension products were loaded onto a Model 377 DNA Sequencer (PE Applied Biosystems) according to the manufacturer's instructions.

## Analysis of Sequencing Traces

Homozygous individuals containing each of the alternative alleles for 35 different polymorphic sequence-tagged sites (pSTSs: 70 sequencing traces altogether for each class of dye terminators) were sequenced using the dRhodamine dye terminator and the BigDye terminator. These traces represent all the possible 3-base windows (64 in total), where the middle base in the window is the only variant base in each pair of the sequences. In addition, 20 sets of control sequencing traces that were produced using the dRhodamine dye terminator and the BigDye terminator chemistries were also analyzed to

# Short Technical Reports

Table 3. 3-Base Windows with Peak Height Distribution Outside of Middle Range in dRhodamine and BigDye Sequences

dRhodamine			BigDye		
Base String	<1/2 Full-Scale (%)	>1/2 Full-Scale (%)	Base String	<1/2 Full-Scale (%)	>1/2 Full-Scale (%)
1. AAG	96	4	1. ACT	7	93
2. ACG	100	0	2. AGA	17	83
3. AGG	0	100	3. AGG	6	94
4. AGT	22	78	4. CAT	6	94
5. CAA	100	0	5. CCG	100	0
6. CAG	100	0	6. CCT	11	89
7. CCG	100	0	7. CGC	20	80
8. CGG	14	86	8. CTC	100	0
9. GAG	100	0	9. GAC	5	95
10. GCG	100	0	10. GCT	3	97
11. GGA	32	68	11. TAC	8	92
12. TAC	100	0	12. TAG	85	15
13. TAG	100	0	13. TAT	5	95
14. TAT	100	0	14. TCG	100	0
15. TCG	100	0	15. TGC	46	54
16. TGA	100	0	16. TTG	32	68

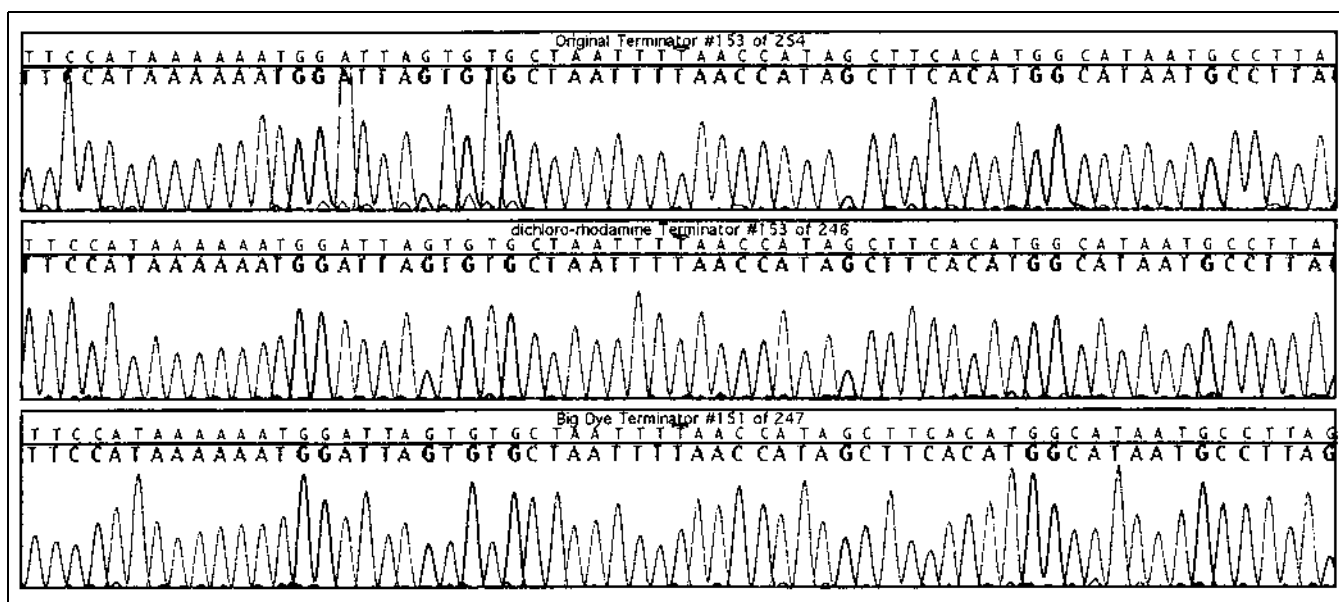


Figure 1. Peak pattern comparison for the three dye-terminator chemistries with AmpliTaq DNA polymerase, FS. The sequencing traces of a 62-base window DNA sequence generated from the same PCR product using the original rhodamine dye (top panel), the dRhodamine dye (middle panel) and the BigDye (bottom panel) terminators are displayed here. The rhodamine dye terminators give the characteristic small G peaks following A peaks, with off-scale A and T peaks following GG and TG peaks, respectively. The dRhodamine dye terminators and the BigDye terminators give much more even peaks, with very few extremely small peaks or off-scale peaks.

# Short Technical Reports

extend and confirm these trends observed in the allelic pairs of sequencing traces to provide a significantly larger sample of each of the 3-base windows. Control sequencing traces were normalized by reanalyzing 3500 data points of the sequencing read using the ABI 377 Analysis Software (PE Applied Biosystems) to produce identical spacing between the traces. To ensure that there is as much uniformity as possible, measurements were taken only from the middle portion (bases 100–200) of the trace. This strategy avoids potential problems that could

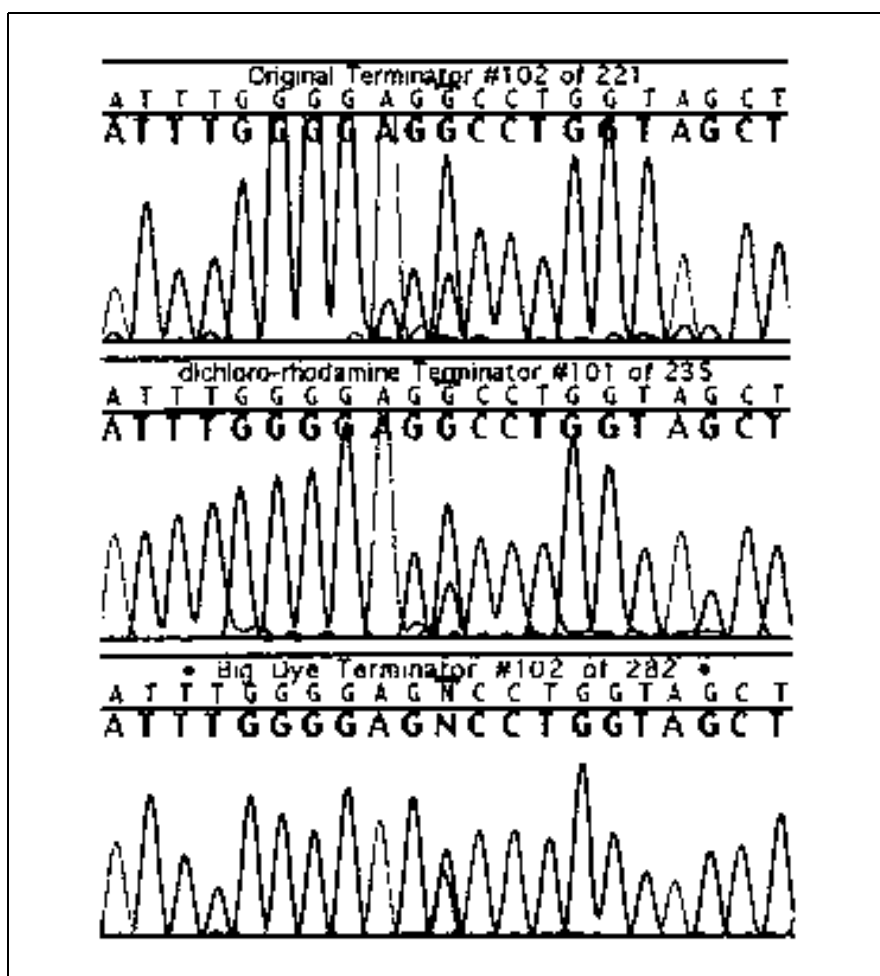
arise by comparing the weaker peaks toward the end of the sequencing run to the stronger peaks at the beginning of the sequencing run. For each of the 64 possible 3-base windows, the 3'-most base in each window was measured in the context of the two bases 5' to it. In the sequencing traces of this study, a full-scale peak measures 40 mm and represents a strong signal that reaches the top of the panel of the printed sequencing trace. Peaks that are >40 mm in height are off-scale and are represented by peaks truncated at 40 mm. When a 3'-base peak was found to be

equal to or less than 1/3 of full-scale ( $\leq 13$  mm), it was defined as a small peak. Similarly, when a 3'-base peak was found to be equal to or greater than 2/3 of full-scale ( $\geq 27$  mm), it was defined as a large peak. When a 3'-base peak was found to be between 1/3 to 2/3 of full-scale (14–26 mm), it was defined as an average peak. Another way to tabulate the results was to divide the peak heights into quartiles, with 10-mm intervals for our 40-mm, full-scale span. In this case, peaks that measured from 10–30 mm were considered to be in the middle half, while those <20 mm were in the smaller half and those >20 mm were in the larger half.

## RESULTS AND DISCUSSION

Analysis of all 3-base windows within the polymorphic STSs (64 windows altogether) revealed several trends that were confirmed by analyzing similar 3-base windows (>20 independent observations of each window except in windows that are less frequently found in the genome, such as NCG, CGN and GAC) in 20 control sequencing traces. The peak patterns produced with BigDye and dRhodamine terminators were highly reproducible and predictable in specific sequence contexts. In this study, if the peak heights of  $\geq 85\%$  of all the examples of a 3-base window found among the control sequences fell in a particular range, that window was considered to have a consistent and predictable peak height.

In 28 (44%) out of the 64 possible 3-base windows among the sequences generated with dRhodamine terminators, by knowing the identity of the two bases 5' to it, one could predict the peak height of the 3'-base consistently as being a small ( $\leq 1/3$  full scale), average ( $>1/3$  but  $<2/3$  full scale) or large ( $\geq 2/3$  full scale) (Table 1). Similarly, one could predict the peak height of a particular base in 23 (36%) of the cases when BigDye terminators were used (Table 2). If one were to expand these arbitrarily assigned bins slightly to include the middle half of the peak heights ( $>1/4$  to  $<3/4$  of full scale), 75% (48 out of 64 possible windows) of the 3' base in both the dRhodamine and the BigDye traces fell within this



**Figure 2. Heterozygote detection by DNA sequencing.** The sequencing traces of a 21-base window DNA sequence containing a heterozygous (G/C) base generated from the same PCR product using the original rhodamine dye (top panel), the dRhodamine dye (middle panel) and the BigDye (bottom panel) terminators are shown. Because the G peak is substantially larger than the C peak in this particular case for the sequences derived from the original rhodamine dye terminator chemistry, the base-calling software is unable to detect the underlying C peak at the polymorphic site. In the dRhodamine dye-terminator sequence, the G peak and the C peak are more similar in size, but are still too different in size for the base-calling software to detect the smaller C peak. The G and C peaks are so similar in height for the sequence generated by BigDye terminator chemistry that the base-calling software is able to call the base at the polymorphic site as an N.

range. This confirms the general impression that these terminators gave much more even peaks. Furthermore, as shown in Table 3, the 3' base in 14 out of the 16 remaining windows among the dRhodamine sequences and 12 out of the 16 remaining windows among the BigDye sequences could be classified as consistently  $<1/2$  full-scale (the smaller half) or  $>1/2$  full-scale (the larger half). It is interesting to note that while 12 of the dRhodamine windows predict peak heights consistently in the smaller half (NAG, NCG, CAA, TAC, TAT and TGA), only 4 of the BigDye windows predict peak heights in the smaller half (CCG, CTC, TAG and TCG). In contrast, only 2 dRhodamine windows predict peak heights in the larger half (AGG and CGG), while 8 BigDye windows do so (ACT, AGG, CAT, CCT, GAC, GCT, TAC and TAT).

Other obvious trends of these two new classes of dye terminators include: (i) very few of the peaks are off-scale, (ii) all the variable windows (AGT and GGA among the dRhodamine sequences; AGA, CGC, TGC and TTG among the BigDye sequences) predict negligible occurrence of peaks less than  $1/4$  full-scale (data not shown) and (iii) the baseline noise was less prominent than that found in the original dye terminators. Figure 1 shows examples of the sequencing traces generated with these terminators as compared with those with the original rhodamine dye terminators. The top panel shows a 62-base section of a sequencing trace generated by the original rhodamine dye terminators. It demonstrates several common peak pattern features of this chemistry, namely, extremely small G peaks in the TAG string, extremely large A peaks in the GGA string and T peaks in the TGT string that are actually off-scale (4). The middle panel shows the same section of sequence generated by the dRhodamine dye terminators. While the G peak in the TAG string is still small, it is twice as tall as the corresponding G peak in the original rhodamine dye-terminator sequence. Furthermore, there are no off-scale peaks in the sequence at all. The bottom panel shows the same section of sequence generated by the BigDye terminators. The peak heights are generally more even in size and, as with the dRho-

damine dye terminators, there are no off-scale peaks.

As reported previously, ten 3-base strings among the 64 possible cases give consistently small peaks in sequencing traces generated by the original rhodamine dye terminators (4). In contrast, there were 5 cases with small peaks found among the sequencing traces derived from the dRhodamine dye terminators and only 1 case with small peaks among those from the BigDye terminators. Moreover, while 75% of the peaks consistently fall within the middle half of the peak height range in dRhodamine dye and BigDye terminator sequences, only 44% (28 cases) of the original rhodamine dye terminator sequences fall within the middle half.

An even peak pattern improves the accuracy of base assignment by the base-calling software and makes it easier to spot heterozygous bases in a sequence. Figure 2 shows a AGG to AGC substitution polymorphism found in a heterozygous individual. Because of the significant difference in peak heights between the G and C peaks in the original rhodamine and dRhodamine terminator sequencing traces, the heterozygous C peak is much smaller than the heterozygous G peak, and the base-calling software is unable to recognize the presence of the small heterozygous C peak. In contrast, the two heterozygous peaks in the BigDye terminator sequence are so similar in height that it is designated N, making it an easy task to find the polymorphic base in the sequence. Although peaks in BigDye terminator sequences are more similar in height than those in original rhodamine and dRhodamine terminator sequences in general, heterozygous peaks of disparate heights are still found in a small number of sequence contexts, making it a challenge for the standard base-calling software to identify the heterozygous bases.

In sequencing pSTSs with the rhodamine dye terminators with the AmpliTaq DNA polymerase, we previously found that in addition to the peak height changes at the polymorphic site, changes in the height of bases immediately 3' to the polymorphic site were useful in providing additional evidence for identifying heterozygous bases in a

# Short Technical Reports

sequencing trace (1). Our current study shows that the more even peak pattern of dRhodamine and BigDye terminator sequencing traces makes it easier to identify heterozygous bases without having to rely on changes in the peak height of the 3' base. The improved attributes of these new dye terminators improve the accuracy of the base-calling programs and enhance the power of computer programs designed for automated identification of DNA sequence variations based on peak height variations.

**H. Zakeri, G. Amparo<sup>1</sup>, S.-M. Chen<sup>1</sup>, S. Spurgeon<sup>1</sup> and P.-Y. Kwok**

*Washington University  
School of Medicine  
St. Louis, MO*

*<sup>1</sup>PE Applied Biosystems  
Foster City, CA, USA*

## REFERENCES

1. Kwok, P.-Y., C. Carlson, T.D. Yager, W. Ankener and D.A. Nickerson. 1994. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* 23:138-144.
2. Lee, L.G., S.L. Spurgeon, C.R. Heiner, S.C. Benson, B.B. Rosenblum, S.M. Menchen, R.J. Graham, A. Constantinescu, K.G. Upadhy and J.M. Cassel. 1997. New energy transfer dyes for DNA sequencing. *Nucleic Acids Res.* 25:2816-2822.
3. Parker, L.T., Q. Deng., H. Zakeri, C. Carlson, D.A. Nickerson and P.-Y. Kwok. 1995. Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry. *BioTechniques* 19:116-121.
4. Parker, L.T., H. Zakeri, Q. Deng, S. Spurgeon, P.-Y. Kwok and D.A. Nickerson. 1996. AmpliTaq<sup>®</sup> DNA Polymerase, FS Dye-Terminator sequencing: analysis of peak height patterns. *BioTechniques* 21:694-699.
5. Rosenblum, B.B., L.G. Lee, S.L. Spurgeon, S.H. Khan, S.M. Menchen, C.R. Heiner and S.M. Chen. 1997. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* 25:4500-4504.
6. Rosenthal, A. and D.S. Charnock-Jones. 1992. New protocols for DNA sequencing with dye terminators. *DNA Seq.* 3:61-64.
7. Tracy, T.E. and L.S. Mulcahy. 1991. A simple method for direct automated sequencing of PCR fragments. *BioTechniques* 11:68-75.

*We thank P. Taillon-Miller for polymorphism information, J. Lee, J. Putzel, K. Zakeri and Q. Li for technical assistance, and L.T. Parker for useful discussion. This work is supported in part by grants from the National Institutes of Health. Address correspondence to Dr. Pui-Yan Kwok, Division of Dermatology, Washington University School of Medicine, 660 S. Euclid Ave., Box 8123, St. Louis, MO 63110, USA. Internet: kwok@im.wustl.edu*

Received 6 April 1998; accepted 22 May 1998.