Two-label peak-height encoded DNA sequencing by capillary gel electrophoresis: three examples

DaYong Chen, Heather R. Harke and Norman J. Dovichi*
Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

Received June 1, 1992; Revised and Accepted August 12, 1992

ABSTRACT
We report a modification to the peak-height encoded DNA sequencing technique of Tabor and Richardson. As in the original protocol, the sequencing reaction uses modified T7 polymerase with manganese rather than magnesium to produce very uniform incorporation of each dideoxynucleoside. To improve sequencing accuracy, two fluorescently labeled primers are employed in separate sequencing reactions. As an example, one sequencing reaction uses a FAM-labeled primer with dideoxyadenosine triphosphate and dideoxyctosine triphosphate; the concentrations of ddATP and ddCTP are adjusted to produce a 2:1 variation in the relative intensity of fragments. The second sequencing reaction uses a TAMRA labeled primer with dideoxythymidine triphosphate and dideoxyguanidine triphosphate; the concentrations of ddTTP and ddGTP are adjusted to produce a 2:1 variation in relative intensity of fragments. The pooled reaction products are separated by capillary gel electrophoresis and identified by one of three different detector systems. Use of a 2:1 peak height ratio typically produces a sequencing accuracy of 97.5% for the first 350 bases; a 3:1 peak height ratio improves accuracy to 99.5% for the first 400 bases. For these experiments, capillary electrophoresis is performed at an electric field of 200 V/cm; two to three hours are required to separate sequencing fragments up to 400 nucleotides in length.

INTRODUCTION
In 1989, Tabor and Richardson reported the effect of manganese ions on the incorporation of dideoxynucleoside triphosphate (ddNTP) by T7 polymerase (1). The manganese not only increased the incorporation rate of the ddNTP but also produced uniform termination of DNA sequencing reactions. In 1990, Ansorge and coworkers and Tabor and Richardson independently reported a DNA sequencing protocol based on the manganese-T7 polymerase reaction (2-3). Both groups reported use of a single sequencing reaction with one fluorescently labeled primer; adjustment of the concentration of ddNTP produces peak heights in an 8:4:2:1 ratio to encode the DNA sequence. Ansorge also reported the use of two sequencing reactions; ddCTP and ddTTP were present in the first while ddATP and ddGTP were present in the second. The products of the two sequencing reactions were separated in adjacent lanes of a polyacrylamide gel.

The single reaction technique offers advantages for primer walking applications: each sample requires only one primer and one reaction. Also, electrophoretic separation of the reaction products occurs in a single lane, increasing the sample throughput. Because the sample is separated in a single lane, the technique is useful in capillary gel electrophoresis, a technique that provides rapid separation of DNA sequencing fragments (4-12).

An 8:4:2:1 peak height ratio leads to poor accuracy in our hands (11); inevitably, fragments terminated in two of the ddNTP's will generate overlapping distributions of peak height. In this laboratory, the accuracy of the DNA sequence is typically 90% for fragments shorter than 250 bases. There are three sources of errors in sequence determination by this technique. First, the relative variation in peak heights associated with each ddNTP can be about 25% (2). This distribution in peak amplitude leads to errors in identification of the terminal nucleotide, particularly if the peak height ratio is not carefully adjusted. Second, the smallest peaks are frequently lost when sandwiched between two larger amplitude peaks. This problem becomes most severe for longer fragments where the electrophoretic resolution has degraded; it is difficult to sequence accurately fragments that are longer than 250 bases with the four level sequencing technique. Third, systematic errors occur due to the presence of ghost peaks associated with false priming, the finite processivity of the polymerase, and contaminant oligonucleotides present in the sample. These systematic errors are particularly important for the ddNTP present in lowest concentration; a small contribution from a ghost peak leads to the largest proportional error in peak amplitude. False priming and artifacts from contaminant oligonucleotides require that highly purified DNA be used for sequence determination.

While it is difficult to control a peak height ratio of 8:4:2:1, a peak height ratio of 2:1 is simple to maintain for any two of the ddNTPs. This paper describes a modification to the peak height encoded technique that is similar to that contained in Ansorge's original paper; we perform two sequencing reactions, each reaction containing a different fluorescently labeled primer.

* To whom correspondence should be addressed
and two ddNTP's. The pooled reaction products are separated on a single capillary gel. This procedure provides the accuracy inherent in a two-level discrimination and the convenience of single column separation with the accuracy of multiple labels, albeit at the expense of performing two sequencing reactions.

**EXPERIMENTAL PROCEDURE**

**Instrumental design—Electrophoresis**

The DNA sequencing capillary electrophoresis systems are similar to previous reports from this laboratory (7, 9, 11—13). The polyimide coated, fused silica capillary is 50-μm inner diameter, 190-μm outer diameter, and typically 35-cm long. Gels are prepared in 5-mL aliquots from carefully degassed mixtures of acrylamide and bisacrylamide (4% T, 5% C), 1× TBE, and 7M urea. Polymerization is initiated by addition of 2-μl of TEMED and 20-μl of 10% ammonium persulfate. The gel solution is injected into the capillary by use of a syringe. To prevent deformation of the gel into the detection cuvette, 7-methacryloyloxypropyltrimethoxysilane is used to bind covalently the gel to the last ~2-cm of the capillary wall. Although polymerization appears complete in 30 minutes, the capillaries are typically stored overnight before use.

A Plexiglas box equipped with a safety interlock holds the injection end of the capillary. The other end of the capillary is inserted into the flow chamber of a locally constructed sheath flow cuvette. The cuvette has a 200-μm square flow chamber with either 1 or 2-mm thick quartz windows.

**Detectors**

Three different fluorescence detectors are described for the two-color peak-height encoded sequencing technique. In the first fluorescence detector, figure 1, a 6-mW argon ion laser beam (λ = 488 nm) is aligned to be parallel with a 1-mW helium-neon laser beam (λ = 543.5 nm) thorough use of a dichroic filter. The parallel beams are focused with a 5× microscope objective into the sheath flow cuvette (13). The helium-neon laser beam is brought to a focus about 100-μm below the tip of the capillary; the argon ion laser beam is focused about 200-μm downstream from the first laser beam. Two 0.70 NA, 60× microscope objectives (Universe Kogaku model 60X-LWD) collect fluorescence at right angles to the beams and cuvette. The fluorescence excited by the helium neon laser is collected with the first objective, imaged onto a 1.5 mm diameter pinhole, passed through a bandpass interference filter with 600 nm center wavelength and 50 nm bandwidth, and detected with a R1477 photomultiplier tube (PMT) operated at −15°C. Fluorescence excited by the argon ion laser is collected with a second objective, imaged onto a 1.5 mm diameter pinhole, passed through a bandpass interference filter with 518 nm center wavelength and 25-nm bandwidth, and detected with a R1477 PMT operated at room temperature. The output from each PMT is conditioned with a simple low-pass electronic filter with 0.01-s time constant and digitized by a National Instruments A/D board in a Macintosh Ilii computer. The data are treated with a Gaussian-shaped digital filter before presentation. This 32 point convolution filter is built from a Gaussian peak with a 0.3 second standard deviation; the filter introduces minimal band-broadening. The sheath flow is provided by a simple siphon based on 7-cm difference in height between the sheath flow reservoir and the waste collection vial, which produces approximately 0.02 mL/hr flow rate.

The second detector, figure 2, is identical to the first, except that fluorescence from both dyes is excited by a 20-mW argon ion laser at right angles to the beams and cuvette. The argon ion laser is imaged onto a 1.5 mm diameter pinhole, passed through a bandpass interference filter with 600 nm center wavelength and 50 nm bandwidth, and detected with a R1477 photomultiplier tube (PMT) operated at −15°C. Fluorescence excited by the argon ion laser is collected with a second objective, imaged onto a 1.5 mm diameter pinhole, passed through a bandpass interference filter with 518 nm center wavelength and 25-nm bandwidth, and detected with a R1477 PMT operated at room temperature. The output from each PMT is conditioned with a simple low-pass electronic filter with 0.01-s time constant and digitized by a National Instruments A/D board in a Macintosh Ilii computer. The data are treated with a Gaussian-shaped digital filter before presentation. This 32 point convolution filter is built from a Gaussian peak with a 0.3 second standard deviation; the filter introduces minimal band-broadening. The sheath flow is provided by a simple siphon based on 7-cm difference in height between the sheath flow reservoir and the waste collection vial, which produces approximately 0.02 mL/hr flow rate.

The second detector, figure 2, is identical to the first, except that fluorescence from both dyes is excited by a 20-mW argon ion laser beam.
ion laser (λ = 488 nm) and both collection optics are focused on the illuminated sample.

The third detector, figure 3, was originally designed for the DuPont Genesis DNA sequencing system (11). Fluorescence is excited by a 30-mW argon ion laser (λ = 488 nm) and collected by a 0.60 NA, 32× microscope objective (Leitz/Wild model 2569–1130). The fluorescence is imaged onto a 0.75-mm pinhole. A dichroic filter splits the fluorescence into two spectral channels. The fluorescence transmitted by the dichroic filter passes through a bandpass interference filter, centered at 550 nm and with 35 nm bandwidth, and is detected with a R1477 PMT. The fluorescence reflected from the dichroic filter passes through a second bandpass interference filter, centered at 515 nm and with 25 nm bandwidth, and is detected with a R1477 PMT. The output from each PMT is conditioned with a simple low-pass electronic filter with 0.5-s time constant and digitized by a National Instruments A/D board in a Macintosh Ilisi computer. The data are treated with a Gaussian-shaped filter before presentation. The sheath flow is provided by a syringe pump operating at a flow rate of 0.08 mL/hour.

Sample preparation

Three samples are used. In each case, the sequencing reaction is carried out in 40 mM MOPS buffer, pH 7.5, 50 mM NaCl, 10 mM MnCl₂, and 15 mM sodium isocitrate. For the first sample, 8 pmol of FAM labeled primer (Applied Biosystems -21M13 FAM) is annealed to 10 µg of M13mp18 single stranded DNA at 65°C for 2 min. followed by slow cooling. A mixture of deoxy- and dideoxynucleoside triphosphates is added to give a ratio of dideoxy nucleoside triphosphate of greater than 200:1. The ratios of dideoxynucleosides are adjusted to yield a nominal peak height ratio of 2:1 for T and G. After the mixture is warmed to 37°C, 6 units of Sequenase Version 2.0 and 0.006 units of pyrophosphatase are added. Incubation continues at 37°C for 30 min., after which the DNA is precipitated with ethanol (2).

Identical experimental conditions are used with a TAMRA-labeled primer (Applied Biosystems -21M13 TAMRA) to yield a nominal peak height ratio of 2:1 for A and C. The samples are resuspended in a 49:1 mixture of formamide-EDTA. Three microliter aliquots are taken from the resuspended samples, heated to 95°C for two minutes, and injected onto the capillary by applying a 200 V/cm electric field for 20 seconds.

Sequence determination

In each case, sequence is interpreted by eye from the smoothed electrophoresis data. The data are plotted in 20 minute intervals with approximately 60 peaks per plot. Two lines are drawn across the plot corresponding to the discrimination level for each fluorescent dye. For fragments between 350 and 400 bases in length, small amplitude fragments occasionally suffer from overlap with neighboring peaks. Interpretation of these small amplitude peaks is usually simple, particularly for samples prepared with a 3:1 peak height ratio.

RESULTS AND DISCUSSION

Two laser-Two collection optic system

The first optical system minimizes cross-talk between fragments labeled with the two dyes. A helium-neon laser, operating at 543.5 nm, efficiently excites TAMRA labeled fragments; FAM labeled fragments are transparent at this wavelength. An argon ion laser, operating at 488 nm, efficiently excites FAM labeled fragments; the molar absorptivity of TAMRA is an order of magnitude less than FAM at this wavelength. In this detection system, the helium-neon laser beam was focused about 200-µm upstream from the argon ion laser beam to minimize photobleaching effects (13). Fluorescence of the TAMRA-labeled fragments is collected by an optical system tuned to the emission maximum of that dye. Fluorescence of the FAM-labeled fragments is collected by an optical system that transmits fluorescence of FAM but blocks the low level fluorescence of TAMRA.

Figure 4a presents data from the two-laser system. Here, A and C terminated fragments are present in a 2:1 ratio in the solid trace while T and G are present in a 2:1 ratio in the dashed trace. The two traces are offset for presentation. The region from 200 to 250 bases, top panel of figure 4b, demonstrates the quality of the data produced by this instrument. Tall peaks in the dashed channel are A, short peaks in the dashed channel are C, tall peaks in the solid channel are T, and short peaks in the solid channel are G. For fragments of this length, the resolution between adjacent peaks ranges from 1.0 to 1.5 and the theoretical plate count is 4.6 million for a fragment with 200 nucleotides. The bottom panel in figure 4b shows the portion of the electropherogram corresponding to fragments ranging from 300 to 350 bases in length. Here, resolution ranges from 0.5 to 0.75; degradation in resolution for longer fragments is associated with increased band-broadening. The theoretical plate count was 2.1 million plates for fragments 334 bases in length. Molecular diffusion dominates band-broadening under these experimental conditions (12). For fragments much longer than 300 bases, peak overlap often swamp the low amplitude C and G peaks. However, the pattern produced by the larger peaks may be identified with the known sequence for fragments up to 750 bases in length.

The sequence for fragments ranging from 70 to 350 bases in length contained five errors, producing an overall sequencing accuracy of 97.5%. In our system, three errors were associated with a missing C (bases 196, 198, and 326); in each case a minor compression caused the C to be lost as a shoulder on an adjacent peak of neighboring bases. The third detector, figure 3, was originally designed for the DuPont Genesis DNA sequencing system (11). Fluorescence is excited by a 30-mW argon ion laser (λ = 488 nm) and collected by a 0.60 NA, 32× microscope objective (Leitz/Wild model 2569–1130). The fluorescence is imaged onto a 0.75-mm pinhole. A dichroic filter splits the fluorescence into two spectral channels. The fluorescence transmitted by the dichroic filter passes through a bandpass interference filter, centered at 550 nm and with 35 nm bandwidth, and is detected with a R1477 PMT. The fluorescence reflected from the dichroic filter passes through a second bandpass interference filter, centered at 515 nm and with 25 nm bandwidth, and is detected with a R1477 PMT. The output from each PMT is conditioned with a simple low-pass electronic filter with 0.5-s time constant and digitized by a National Instruments A/D board in a Macintosh Ilisi computer. The data are treated with a Gaussian-shaped filter before presentation. The sheath flow is provided by a syringe pump operating at a flow rate of 0.08 mL/hour.

Sample preparation

Three samples are used. In each case, the sequencing reaction is carried out in 40 mM MOPS buffer, pH 7.5, 50 mM NaCl, 10 mM MnCl₂, and 15 mM sodium isocitrate. For the first sample, 8 pmol of FAM labeled primer (Applied Biosystems -21M13 FAM) is annealed to 10 µg of M13mp18 single stranded DNA at 65°C for 2 min. followed by slow cooling. A mixture of deoxy- and dideoxynucleoside triphosphates is added to give a ratio of dideoxy nucleoside triphosphate of greater than 200:1. The ratios of dideoxynucleosides are adjusted to yield a nominal peak height ratio of 2:1 for T and G. After the mixture is warmed to 37°C, 6 units of Sequenase Version 2.0 and 0.006 units of pyrophosphatase are added. Incubation continues at 37°C for 30 min., after which the DNA is precipitated with ethanol (2).

Identical experimental conditions are used with a TAMRA-labeled primer (Applied Biosystems -21M13 TAMRA) to yield a nominal peak height ratio of 2:1 for A and C. The samples are resuspended in a 49:1 mixture of formamide-EDTA. Three microliter aliquots are taken from the resuspended samples, heated to 95°C for two minutes, and injected onto the capillary by applying a 200 V/cm electric field for 20 seconds.

Sequence determination

In each case, sequence is interpreted by eye from the smoothed electrophoresis data. The data are plotted in 20 minute intervals with approximately 60 peaks per plot. Two lines are drawn across the plot corresponding to the discrimination level for each fluorescent dye. For fragments between 350 and 400 bases in length, small amplitude fragments occasionally suffer from overlap with neighboring peaks. Interpretation of these small amplitude peaks is usually simple, particularly for samples prepared with a 3:1 peak height ratio.

RESULTS AND DISCUSSION

Two laser-Two collection optic system

The first optical system minimizes cross-talk between fragments labeled with the two dyes. A helium-neon laser, operating at 543.5 nm, efficiently excites TAMRA labeled fragments; FAM labeled fragments are transparent at this wavelength. An argon ion laser, operating at 488 nm, efficiently excites FAM labeled fragments; the molar absorptivity of TAMRA is an order of magnitude less than FAM at this wavelength. In this detection system, the helium-neon laser beam was focused about 200-µm upstream from the argon ion laser beam to minimize photobleaching effects (13). Fluorescence of the TAMRA-labeled fragments is collected by an optical system tuned to the emission maximum of that dye. Fluorescence of the FAM-labeled fragments is collected by an optical system that transmits fluorescence of FAM but blocks the low level fluorescence of TAMRA.

Figure 4a presents data from the two-laser system. Here, A and C terminated fragments are present in a 2:1 ratio in the solid trace while T and G are present in a 2:1 ratio in the dashed trace. The two traces are offset for presentation. The region from 200 to 250 bases, top panel of figure 4b, demonstrates the quality of the data produced by this instrument. Tall peaks in the dashed channel are A, short peaks in the dashed channel are C, tall peaks in the solid channel are T, and short peaks in the solid channel are G. For fragments of this length, the resolution between adjacent peaks ranges from 1.0 to 1.5 and the theoretical plate count is 4.6 million for a fragment with 200 nucleotides. The bottom panel in figure 4b shows the portion of the electropherogram corresponding to fragments ranging from 300 to 350 bases in length. Here, resolution ranges from 0.5 to 0.75; degradation in resolution for longer fragments is associated with increased band-broadening. The theoretical plate count was 2.1 million plates for fragments 334 bases in length. Molecular diffusion dominates band-broadening under these experimental conditions (12). For fragments much longer than 300 bases, peak overlap often swamp the low amplitude C and G peaks. However, the pattern produced by the larger peaks may be identified with the known sequence for fragments up to 750 bases in length.

The sequence for fragments ranging from 70 to 350 bases in length contained five errors, producing an overall sequencing accuracy of 97.5%. In our system, three errors were associated with a missing C (bases 196, 198, and 326); in each case a minor compression caused the C to be lost as a shoulder on an adjacent peak of neighboring bases.
A peak. Operation of the capillaries above room temperature or addition of formamide to the sequencing gel minimizes compressions (14). Regression analysis should prove useful in detection of these poorly resolved bases. In the other two cases (bases 281, 282), a T was called for a G. These errors arise from inaccurate setting of a discrimination level for the G-T peaks. Although careful interpretation of the peak heights might lead to improved discrimination, a large difference in peak heights

Figure 4. Sequencing data for M13mpl8 generated with the two-laser two-collection optic peak height encoded sequencing method. FAM labeled primer is used to generate fragments terminated in A and C in a 2:1 peak height ratio. TAMRA labeled primer is used to generate fragments terminated in T and G in a 2:1 peak height ratio. Fluorescence collected in the red spectral channel is presented in the solid trace and fluorescence collected in the blue spectral channel is presented as the dashed trace. Separation took place at an electric field of 200 V/cm. Figure 4a. Entire separation. The sample was injected at $T = 0$. Figure 4b presents data for fragments from 200 to 250 bases in length (top panel) and from 300 to 350 bases in length (bottom panel).
Figure 5. Sequencing data for M13mp18 generated with the two-laser one-collection optic peak height encoded sequencing method. FAM labeled primer is used to generate fragments terminated in A and C in a 3:1 peak height ratio. TAMRA labeled primer is used to generate fragments terminated in T and G in a 3:1 peak height ratio. Fluorescence collected in the red spectral channel is presented in the solid trace and fluorescence collected in the blue spectral channel is presented as the dashed trace. Separation took place at an electric field of 200 V/cm. Figure 5a. Entire separation. The sample was injected at T = 0. Figure 5b presents data for fragments from 200 to 250 bases in length (top panel) and from 300 to 350 bases in length (bottom panel). Figure 5c plots the blue spectral data as the dashed trace and the difference between the red and blue data as the solid trace for fragments ranging from 300 to 350 bases in length.
simplifies decisions on peak identity. As shown below, the latter errors can be minimized by use of a 3:1 peak height ratio rather than a 2:1 peak height ratio.

Regression analysis appears necessary for accurate determination of sequence for fragments longer than 400 bases. A particular advantage of this two-laser system comes in designing an algorithm for peak identification. One of the more critical tasks in interpreting the electrophoretic data is determination of the signal baseline. Because the signals from the two dyes do not interfere, the signal drops to the baseline whenever more than a few consecutive bases appear in the other spectral channel.

One laser—Two collection optic system

The two-laser system is experimentally complicated. To simplify the system, a single argon ion laser beam can excite fluorescence from both FAM and TAMRA. Two microscope objectives image the single fluorescent spot into two spectral channels equipped with appropriate spectral filter and photomultiplier tube. Because the absorbance cross-section for TAMRA is much less than the absorbance of FAM at 488 nm, a relatively high laser power should be used. The higher laser power is sufficient to photobleach FAM but not TAMRA. The signal amplitude from the two dyes is similar because of this photobleaching.

To improve sequencing accuracy, a 3:1 peak height ratio was used to generate the sequencing sample shown in figure 5. Figure 5a presents data for the entire separation. Figure 5b presents a more detailed view of the data. The top panel presents data taken for fragments ranging from 200 to 250 bases in length. Fragments that produce a large signal in both spectral channels are A, fragments with small signals in both spectral channels are C, peaks with large signals in the solid channel are T, and peaks with small signals in the solid channel are G. Resolution ranges from 1.0 to 1.5 for these fragments. Two minor compressions are observed in this data (peaks 223-224 and 225-226). These compressions are not observed in the data of figure 4 and may represent incomplete denaturing of the sample before injection. The bottom panel of figure 5b presents sequence for fragments ranging from 300 to 350 bases in length. Resolution ranges from 0.5 to 0.75. As the resolution between adjacent peaks degrades, sequence determination becomes more problematic, particularly for the case where a low amplitude peak (C or G) is sandwiched between two high amplitude peaks.

Subtraction of the second channel from the first channel improves data interpretation. Figure 5c presents the subtracted data along with the raw data for channel 2 for fragments ranging from 300 to 350 bases in length. Several poorly resolved peaks in figure 5b are clearly resolved in the subtracted data; in particular, peaks 310, 331, 335, 338, and 346 are difficult to distinguish by eye in the raw data of figure 5b but are clearly resolved as G's in the processed data of figure 5c. This processed data produces excellent sequencing accuracy. In the range of 70 to 350 bases, there are two errors: a C is missing at base 198 due to a minor compression where a G is sandwiched between two A's, a G is missing at base 291 due to overlap with a T. For fragments in this size range, sequencing accuracy exceeds 99%.

Accuracy degrades for longer fragments: for bases 354—356 an A is called for CCC, a C is missing at base 361, a G is missing at base 368, a C is missing at base 386, a C is missing at base 395, and a C is missing at base 399; for fragments ranging from 70 to 400 bases, the accuracy exceeds 97%. In this data set, most errors are associated with missing bases: there are five missing C's and one missing G. Bases ending in C and G produce low amplitude peaks that are easily swamped by higher amplitude neighboring peaks.

One laser—One collection optic system

We have reported a capillary gel electrophoresis system based on the DuPont fluorescently labeled dideoxynucleosides (11). The DuPont system, which is no longer manufactured, excites fluorescence with a single argon ion laser beam and a single microscopic optic collects fluorescence. A simple spectrophotometer is constructed from a dichroic filter that reflects light of wavelength shorter than 530 nm and transmits longer wavelengths. Spectral bandpass filters isolate fluorescence in the 515nm and 550 nm regions. Two fluorescently labeled primers (FAM and JOE from Applied Biosystems) are available that match the spectral filters used in the DuPont sequencer.

We have used these dyes in the modified peak amplitude sequencing technique. Preliminary work, with a peak height ratio of 2:1 for G-T and A-C, generated an average error rate of 3% for fragments less than 340 bases in length. Poor height discrimination produced these errors, particularly for fragments terminated by G and T. To improve discrimination, a 3:1 peak height ratio was used to generate the sequencing sample shown in figure 6. A 41 cm long capillary was used for this data; as a result, the sequencing rate is less than the earlier figures.

To demonstrate sequencing for samples other than M13mp18, an A-T rich insert was used for this experiment. Compressions are uncommon with this sample. Figure 6a presents data for the entire separation. Sequence may be determined accurately for the first 400 bases. Figure 6b presents a more detailed view of the data. The top panel presents data taken for fragments ranging from 200 to 250 bases in length. Fragments that produce a similar amplitude signal in both the dashed channel and the solid channel are either G or T; the larger peaks in this set correspond to fragments terminated in T and the smaller peaks correspond to fragments terminated in G. Fragments that produce a large signal only in the solid channel are either A or C. Fragments in this set that produce large peaks are A and fragments that produce smaller peaks are C. Resolution ranges from 1.0 to 1.5 for these fragments. The bottom panel of figure 6b presents sequence for fragments ranging from 300 to 350 bases in length. Resolution ranges from 0.5 to 0.75. The sequence contains two errors for fragments ranging from 25 bases to 430 bases in length: a G is read for a C at base 387 and a C is missing at base 395. Sequencing accuracy exceeds 99.5% for fragments in this size range. In this sample, both errors are associated with fragments ending in C. The small peaks produced by this fragment tend to be misinterpreted as both the resolution and signal-to-noise ratio degrade.

CONCLUSIONS

There is an optimum ratio of peak heights for this sequencing technique. If the ratio is near 1, then errors arise because of insufficient discrimination between peak heights. If the ratio is larger than 3, then errors are made late in the run for low amplitude peaks. It appears that a ratio of 2.5 to 3 is ideal in obtaining high sequencing accuracy for at least 400 bases. Improved signal-to-noise and computer algorithms will inevitably produce superior results for longer fragments.
Figure 6. Sequencing data for an MMTV-insert in M13mp18 generated with the one-laser one-collection optic peak height encoded sequencing method. FAM-labeled primer is used to generate fragments terminated in A and C in a 3:1 peak height ratio. JOE-labeled primer is used to generate fragments terminated in T and G in a 3:1 peak height ratio. Fluorescence collected in the red spectral channel is presented in the solid trace and fluorescence collected in the blue spectral channel is presented as the dashed trace. Separation took place at an electric field of 200 V/cm. Figure 6a. Entire separation. The sample was injected at T = 0. Figure 6b presents data for fragments from 200 to 250 bases in length (top panel) and from 300 to 350 bases in length (bottom panel).

Sequencing accuracy was determined by visual inspection of the data by an undergraduate and a graduate student in this laboratory. Our best data was obtained with an A-T rich sample that tends not to form compressions; sequencing accuracy exceeded 99% for fragments up to 400 bases in length. Our typical M13mp18 data shows 99% accuracy for fragments ranging from 70 to 350 bases in length. This sequencing accuracy should be compared with that reported by Karger, Harris, and
Gesteland, who sequenced M13mp18 with both an ABI 373A slab-gel sequencer and a capillary gel electrophoresis system based on four dye-labeled primers (10). They reported an accuracy of 97% for both the ABI sequencer and the capillary gel electrophoresis system for fragments up to 350 bases in length.

The two-color peak height encoded system produces accuracy equal to or greater than conventional fluorescence based sequencing. However, only two fluorescently labeled primers are required for the sequencing reaction, compared with four labels for the ABI sequencer. As a result, the two-color peak height encoded system could be cost effective for primer walking experiments, particularly when the rapid throughput of capillary electrophoresis outweighs disadvantages of preparing two labeled primers. The two-color sequencing protocol suffers from one significant disadvantage: the technique requires use of T7 polymerase. Until a thermally stable polymerase is discovered that produces uniform incorporation of dideoxynucleosides, cycle sequencing and PCR-based sequencing will remain impractical with the peak-height encoded sequencing techniques.

We have presented systems designed for use with the following pairs of Applied Biosystems primers: FAM—TAMRA and JOE—FAM. The former optical system also works well with JOE—ROX, JOE—TAMRA, and FAM—ROX. Any of these reaction pairs could be separated with an ABI sequencer. The JOE—FAM system can be separated with the DuPont Genesis sequencer. The peak-height encoded technique may be employed on commercial sequencers with appropriate software modification.

One last system can be considered based on TAMRA—ROX. These dyes are excited efficiently by the low cost green helium neon laser; the use of a single collection optic and a dichroic filter to split the emission spectrum to two photodetectors would result in a low cost and compact detector.

ACKNOWLEDGEMENTS

This work was supported in part by the Department of Energy-Human Genome Initiative (USA) grant number DE-FGO2-91ER61123. Support by DOE does not constitute an endorsement of the views expressed in this article. This work was also supported by the Natural Sciences and Engineering Research Council of Canada, the Department of Chemistry of the University of Alberta, Pharmacia Inc. and Waters Division of Millipore Inc. HRH acknowledges a predoctoral fellowship from the Alberta Heritage Foundation for Medical Research. NJD acknowledges a Steacie fellowship from the Natural Sciences and Engineering Research Council of Canada. Carl Fuller of United States Biochemical Corporation kindly supplied the sequencing sample used in figures 4 and 5. John Elliott of the Department of Medical Microbiology and Infectious Diseases of the University of Alberta kindly provided the template used to generate the sequencing data of figure 6. Sue Bay and Daniel Figeys of the University of Alberta called the sequence for this data.

REFERENCES