Continuous, on-line DNA sequencing using oligodeoxynucleotide primers with multiple fluorophores

John A. Brumbaugh
*University of Nebraska - Lincoln*

Lyle R. Middendorf
*Molecular Biosystems*

Daniel L. Grone
*Molecular Biosystems*

J. L. Ruth
*Molecular Biosystems*

Follow this and additional works at: [https://digitalcommons.unl.edu/bioscifacpub](https://digitalcommons.unl.edu/bioscifacpub)

Part of the [Life Sciences Commons](https://digitalcommons.unl.edu/bioscifacpub)


[https://digitalcommons.unl.edu/bioscifacpub/147](https://digitalcommons.unl.edu/bioscifacpub/147)

This Article is brought to you for free and open access by the Papers in the Biological Sciences at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications in the Biological Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Continuous, on-line DNA sequencing using oligodeoxynucleotide primers with multiple fluorophores

(JohD A. Brumbaugh*, Lyte R. Middendorf†, Daniel L. Gronet‡, and J. L. Ruth§)

*School of Biological Sciences, University of Nebraska, Lincoln, NB 68588-0118; †LiCor, Inc., Lincoln, NB 68504; and §Molecular Biosystems, San Diego, CA 92121

Communicated by Myron K. Brakke, April 21, 1988 (received for review February 29, 1988)

ABSTRACT A method for sequencing DNA by using a dfluoresceinated primer and laser excitation is described. Dideoxy protocols have been determined that provide sequences for 600 bases starting with base 1 with <1% error in a single load. Electrophoresis is at 20 W and the bands are detected 24 cm from the bottom of the loading well with a scanning fluorescence detector. Bands are imaged on a TV screen in two dimensions. The sequences can be read from the TV screen manually or semiautomatically by using a simple software program. The system allows more bases to be read with a lower error rate than any other reported automated sequencing method.

Two general methods are used to sequence DNA. The chemical method, devised by Maxam and Gilbert (1), selectively cleaves DNA strands between specific bases. The enzymatic method, devised by Sanger et al. (2), enzymatically elongates primed DNA strands coupled with a stochastic termination of elongation caused by the incorporation of dideoxy analogs of each of the four nucleotides. The DNA strands are radioactively labeled. Electrophoresis followed by autoradiography gives a series of band patterns in four lanes that indicate the DNA sequence.

The need for rapid, reliable automated sequencing with low cost per sequenced base has been addressed (3–7). Several automated approaches to DNA sequencing have been reported using both radioactive detection (8, 9) and fluorescent detection (10–17).

The purpose of this report is to show an improved method for automated, continuous, on-line, real-time DNA sequencing (10–12). The method uses standard dideoxy reactions with a fluorescently tagged primer carrying two fluorophores. A fluorescence detection system located at a fixed distance from the loading wells records the bands as a two-dimensional image as they move past the detector.

MATERIALS AND METHODS

Synthesis of Fluorescently Labeled Primers. A deoxyuridine analog with a primary amine "linker arm" of 12 atoms attached at C-5 was synthesized as published (18, 19). Synthesis of the analog consists of derivatizing 2'-deoxyuridine through organometallic intermediates to 5-methylpropenoyl-2'-deoxyuridine. Reaction with dimethoxytritylchloride produces the corresponding 5'-dimethoxytrityl adduct. The methyl ester is hydrolyzed, activated, and reacted with an appropriately monoacetylated alkyl amine. After purification, the resultant linker arm nucleosides are converted to nucleotide analogs suitable for chemical oligonucleotide synthesis. The structure of the linker arm analog is shown in Fig. 1.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

A 19-base M13 primer (5'-dGGTTTTCAGTCACTGACG-3') was made that included two linker arm bases (T) at positions 5 and 12. The primer was synthesized by using modified phosphoridate chemistry and purified to electrophoretic and chromatographic homogeneity by reverse-phase HPLC (RP-HPLC; ref. 20).

To a solution of 50 nmol of 19-mer linker arm oligonucleotide in 25 μL of 500 mM sodium bicarbonate (pH 9.4) was added 20 μL of 300 mM FITC in dimethyl sulfoxide. The mixture was agitated at room temperature for 6 hr. The oligonucleotide was separated from free FITC by elution from a 1 × 30 cm Sephadex G-25 column with 20 mM ammonium acetate (pH 6), combining fractions in the first UV-absorbing peak. Analysis by analytical 20% PAGE indicated the reaction was complete, with fluorescent oligomer electrophoresing slower than nonfluoresceinated oligomer by the equivalent of 1 nucleotide unit. The FITC-oligomer was purified by preparative RP-HPLC using an 8.3-cm Perkin-Elmer Pecosphere C-8 (3 μm) column eluted with a linear gradient of 7–35 vol% acetonitrile in 100 mM triethylammonium acetate (pH 7.0) over 20 min at 1.0 ml/min. Analyses were by Waters model 490 multiple-wavelength absorbance detector with simultaneous detection at 260 nm and 495 nm. The product was concentrated and precipitated with ethanol to recover 33 nmol (65%) of fluoresceinated oligonucleotide. Products were homogeneous by PAGE and RP-HPLC, reacted with polynucleotide kinase, and had A260/A290 ratios of 2.08 as predicted for such oligomer–fluorophore conjugates at pH 8. The product was used directly for hybridization.

Sequencing Traits of Fluorescently Labeled Primers. Linker arm primer (with no fluorescent groups) was used in sequencing reactions with standard dideoxy protocol and [32P]dCTP (2). The test specimen was a 1-kilobase (kb) portion of pBR325 cloned into M13mp18 at the HindIII site and terminating at the EcoRI site (21). Similar reactions were done using the fluoresceinated primer to determine if it were a suitable substrate for Klenow primer extension.

Sample Preparation. DNA was prepared for automated sequencing by a modest scale-up of mini-preparations (22). Two hundred microliters of each M13 viral stock was added to 20 ml of a recently inoculated culture of JM109 and the

FIG. 1. General structure of the uracil analog with linker arm attached at C-5. n = 7; R₁ = fluorescein isothiocyanate (FITC); R₂ = deoxyribose.

Abbreviations: FITC, fluorescein isothiocyanate; RP-HPLC, reverse-phase HPLC; PMT, photomultiplier tube.
mixture was incubated on a shaker overnight. The following day the DNA was extracted (22). For convenience, the DNA was suspended to a concentration of >0.5 μg/μl.

This method provided enough DNA for several reactions. It was not necessary to prepare highly purified DNA through cesium chloride ultracentrifugation.

Dideoxy Reactions for Automated Sequencing. Five micrograms of template DNA (= 2 pmol) and 50 ng of difluoresceinated primer (=7.2 pmol) were hybridized in 12.5 μl of polymerase buffer. After hybridization, 1 μl of 0.1 M dithiothreitol, 3 μl of H2O, and 2.5 μl of the Klenow fragment of DNA polymerase I (6 units/μl) were added and mixed. One microliter of dNTP mix and 1 μl of ddNTP mix for each respective base type were added to 3 μl of the template/primer/Klenow mixture giving four reaction aliquots. The final concentration of dNTPs was 65.6 μM, except the reaction-specific deoxynucleotide concentration, which was 3.3 μM. The dideoxy concentrations were as follows: ddATP, 400 μM; ddTTP, 600 μM; ddGTP, 100 μM; ddCTP, 100 μM. The 7-deaza-2'-deoxy-GTP analog was used in place of dGTP. Each reaction aliquot was incubated at 30°C for 40 min and the reactions were stopped by adding 5 μl of stop buffer. This procedure is similar to standard dideoxy sequencing protocols (2) except that amounts of template, primer, Klenow fragment, and nucleotides are increased. No radioactivity is used.

Gel Electrophoresis. Prior to use, the plate (18 cm × 30.5 cm) next to the detector was treated with Sigmaclone (Sigma). The top cm of the other plate was treated with a binding silane, γ-methacryloxypropyltrimethoxysilane (ref. 23; Sigma). Two types of combs were used. The 8-well combs produced wells 5 mm × 7.6 mm (width × height) and allowed two samples to be run simultaneously. The 16-well combs produced wells 3.4 mm × 4.8 mm and allowed four samples to be run simultaneously.

Gel concentrations of 4%, 6%, and 8% acrylamide (5% bisacrylamide) with 8 mM urea were tested. The gel thickness for most runs was at 0.35 mm. The buffer was 133 mM Tris base, 44 mM boric acid, and 2.8 mM EDTA with the pH adjusted to 8.9 with NaOH at 50°C to minimize the reduction in dye fluorescence that occurs at lower pH values.

Electrophoresis commenced with a 1200-V prerun for 0.5-1.0 hr. Two thousand volts were applied immediately after loading to drive the samples quickly into the gel. The same was then electrophoresed at a constant power of 16-22 W (∼40 V/cm). A thermostatically controlled temperature plate placed against one side of the glass/gel sandwich was regulated at 50°C.

Detection, Imaging, and Data Analysis. The fluorescent detection and imaging system shown in Fig. 2 was used.

A 10-mW argon laser (model 2001-10SL, Cynics, Sunnyvale, CA) was used as an excitation source since FITC has an absorption peak at 490 nm (24). The 488-nm laser output radiation was filtered by a 1-nm bandpass filter centered at 488 nm (Corion D1-488-F) to eliminate stray 514.5-nm radiation from the laser.

The excitation radiation was chopped at 10 kHz by rotating a chopper wheel (5000 rpm, 128 slots per revolution). Borosilicate plates (3.2 mm thick, Schott Tempax) were primarily used due to their low inherent background fluorescence. Soda lime glass (3.2 mm thick, window glass) was also used successfully even though its background fluorescence was 5-10 times as much as that of borosilicate (unpublished observations).

The excitation radiation was reflected by a front surface mirror mounted on a movable microscope stage such that it entered the glass/gel sandwich at Brewster's angle (56°) to minimize reflected stray light and maximize the amount of energy imparted to the bands in the gel. The polarized laser beam was oriented so that the light vibrations were parallel to the plane of incidence. After being reflected by the mirror, but before entering the glass/gel sandwich, the excitation radiation was focused to a spot diameter of ∼30 μm by a 25.4-mm focal length lens (Melles Griot 01 LPX 037) located on the movable microscope stage.

The emitted radiation was collected by a microscope that was scanned back and forth across the glass/gel sandwich 24 cm down from the bottom of the loading wells. Fluorescent quenching was minimized by using a relatively rapid scan rate of 5 cm/sec. Scanning was accomplished by a stepper motor controlled lead screw that moved the microscope stage.

The objective lens was a 20-mm focal length aspheric lens, f/#1 (Rolyr Optics, no. 17.1055), that focused the emitted radiation at infinity. The emitted radiation was then filtered by a 20-nm bandpass filter centered at 520 nm (Corion S10-520-F), two sharp-cut orange filter glasses (Schott OG515, 3 mm thick), and a 550-nm short-pass filter (Corion LS-55-F) that filtered out the Raman background (585 nm) of the water (3400/cm Raman shift) in the gel. After filtering, the emitted radiation was focused by a 100-mm focal length lens (Melles Griot 01 LPX 177) through a spatial filter onto a photomultiplier tube (PMT; Hamamatsu R928). The PMT was operated at a voltage between –800 and –950 V.

The entire electrophoretic apparatus and optical scanner were housed in a light-proof enclosure to minimize stray room light as well as provide protection to the user against the laser light and high-voltage supplies. The PMT output current was amplified by a current amplifier (PAR model 181). A lock-in amplifier (PAR model 5101) was used to measure only the 10-kHz component of the PMT signal. A reference signal for the lock-in amplifier was generated by the chopper wheel. The signal was then filtered by an 8-pole Bessel filter (Frequency Devices model 902LF1) having a cutoff of 10 Hz and then digitized at 256 conversions per sec, giving 512 conversions for a 2-sec scan (10 cm).

The data were collected by an IBM XT computer that displayed in real time on a TV monitor each scan line. Each line contained 512 data points with the data displayed at 256 gray levels. The first scan line appeared at the bottom of the monitor with each succeeding scan line appearing above one another. After 512 lines were displayed, the video data (256 kilobytes) were stored in a 20-megabyte hard disk drive and the screen was blanked.

Each run could be programed to last for a given number of screens. After completion, a relay system turned off the high-voltage supplies and the laser. After the run was completed, the image data were recalled to the screen and semiautomatically converted to sequence data. This was facilitated through the use of a mouse (Microsoft). The left and right side of each lane was defined and the lane was.
assigned a base type adenine, guanine, cytosine, or thymine based upon the load. Four lanes were defined for each sample, with conversion accomplished one sample at a time. After lane definition a bar cursor could be moved horizontally from one lane to another and moved vertically a programmable distance corresponding to the band-to-band distance. When the bar cursor was positioned over the band in the sequence, a button on the mouse was depressed, which automatically wrote the base type in the sequence data file as well as visually marking the selected band. A cassette tape drive (ADIX model TD440) was used to archive video data after the sequence was determined.

RESULTS

Over 200,000 bases from >350 samples have been sequenced by using this method. In addition to the test specimen from pBR325, DNA segments from maize, a chloro virus, Xenopus, mouse, Drosophila, tobacco, and man have been sequenced.

The method appears very reliable and gives consistent sequences to between 500 and 600 bases starting with the first base past the primer. Table 1 shows the average cumulative reading errors for the test specimen. The average percent error for runs between 550 and 600 bases is 0.83%. Table 1 summarizes data obtained over an 11-month period. These error rates are based upon a comparison of our sequence data to the published sequence data for the cloned fragment of pBR325 (21). Our sample did have a point mutation at base 5696 where a thymine replaced the cytosine of the published sequence. Since this was consistent in all runs, we used the thymine as the normal sequence for our sample.

It is more difficult to obtain an estimation of sequencing accuracy when the sequences being analyzed have not been determined. Table 2 shows a comparison of sequence data obtained from separate isolations of maize alcohol dehydrogenase fragments. Comparisons were made within a given sample but derived from separate DNA isolations and separate polymerase reactions. It can be seen from Table 2 that only seven single reading differences to 500 bases were found in 11 samples, about 0.13% average variation.

The errors were of two types. Some errors were due to conditions involving the enzymatic and chemical reactions, such as bands appearing in more than one lane at a given position or bands missing at a given position. Such errors were not due to the detection system and would be present when manual methods are used. The second type of error occurs near the end of a sequence when the band separations become less distinct and difficult to interpret.

Sequencing Traits of Fluorescently Labeled Primers. Standard autoradiographic techniques showed that the linker arms did not interfere with biological activity and allowed sequence data to be generated in a conventional manner (Fig. 3). The bands containing the difluoresceinated primer migrated more slowly in the gel than the bands produced using the "linker arm only" primer but yielded a normal pattern (Fig. 3). The difluoresceinated primer, therefore, served as a good substrate for Klenow primer extension and thus was suitable for fluorescent DNA sequencing.

Reactions for Automated Sequencing. To prime 5 µg of template DNA (=2 pmol), 50 ng of primer (=7.2 pmol) was used, a molar ratio of 1:3.6. This provides enough product for three to six fluorescent sequencing runs (100--250 ng of template per well or 400--1000 ng per sample). The concentrations of the deoxy and deoxy nucleotide pools were adjusted such that bands had nearly even brightness over the entire run. The concentrations described in Materials and Methods provided broad coverage for moderately (A + T)- to moderately (G + C)-rich samples in the range from 1 to 600 bases.

Gel Electrophoresis. Eight percent gels gave excellent well morphology and well-defined bands but produced sequence data rather slowly. Four percent gels, on the other hand, produced sequence data very fast but gave poor well morphology, which caused interpretation problems. Six percent gels gave good well morphology and resolution, producing data at about 1.0 base per min per four lanes and were used routinely.

Electrophoresis was done by using constant power. Tests were conducted in the range between 16 and 22 W. There was a slight resolution advantage in interpreting data produced at 16 W over that of higher wattages, but this did not compensate for the slower speed at which data were produced. We chose to use 20--22 W.

Detection. Several factors enhanced fluorescent detection that enabled 500--600 bases to be sequenced in a single load. The long time constant of background fluorescence was discriminated from signal fluorescence by chopping the laser beam at 10 kHz. This was especially needed when soda lime glass was used. Focusing the laser beam allowed for the resolution of closer band-to-band intervals, which decrease as a function of electrophoresis time (unpublished observa-

<table>
<thead>
<tr>
<th>Table 2. Comparison of sequence data obtained from separate isolations of maize alcohol dehydrogenase fragments read to 500 bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone designation</td>
</tr>
<tr>
<td>18N</td>
</tr>
<tr>
<td>18Q</td>
</tr>
<tr>
<td>19D</td>
</tr>
<tr>
<td>19N</td>
</tr>
</tbody>
</table>

Table 1. Average cumulative reading errors using test specimen (fragment of pBR325)

<table>
<thead>
<tr>
<th>Bases</th>
<th>Average number of errors*</th>
<th>n</th>
<th>Average % errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1--300</td>
<td>0.34</td>
<td>120</td>
<td>0.11</td>
</tr>
<tr>
<td>To 400</td>
<td>0.84</td>
<td>116</td>
<td>0.25</td>
</tr>
<tr>
<td>To 450</td>
<td>1.74</td>
<td>107</td>
<td>0.39</td>
</tr>
<tr>
<td>To 500</td>
<td>3.15</td>
<td>86</td>
<td>0.63</td>
</tr>
<tr>
<td>To 550</td>
<td>4.62</td>
<td>58</td>
<td>0.84</td>
</tr>
<tr>
<td>To 600</td>
<td>5.25</td>
<td>28</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Gel concentrations of 6% acrylamide/8 M urea were used at 50°C. Samples were read using the semiautomated software described in the text. n, Number of trials.

*Average number of errors per sample analyzed to base number indicated.

Fig. 3. Sequence autoradiogram of M13mp18 with a pBR325 segment inserted at the HindIII site. The left four lanes show the pattern produced by the difluoresceinated primer; the right four lanes were produced by using unfluoresceinated primer. The arrow indicates the region of the HindIII site (AAGCTT). The difluoresceinated primer produced discrete bands that migrated more slowly than unfluoresceinated primer due to the molecular weight of the fluoresceins.
DISCUSSION

The primary reason for scaling up the amount of DNA produced and reacted was to eliminate the necessity of repeating isolations and reactions. Clearly less DNA can be used since three to six runs can be made per reaction. The lower limits have not yet been determined but would appear by calculation to be in the 1-μg range (see Results). Use of the 7-deazaguanine analog eliminated most compressions, although it was still a problem with poly(CG) linkers (26, 27).

Smith et al. (13) found that primers with attached dyes migrated at different rates than those that did not have attached dyes; in this respect their results were similar to ours. Smith et al. (16) also used four different dyes and varied the linkers to compensate for differing mobility effects (28), but the migration of the four primers still varied approximately one-fourth of a base due to the effects of the different fluorophores. The four-dye single-lane format (16, 28) can be compared with our single-dye four-lane format. The four-dye single-lane format requires four independent hybridizations followed by four independent elongation reactions. The single-dye four-lane format requires just a single hybridization followed by four independent reactions. Prober et al. (17) use fluorescently tagged dideoxy analogs and do reactions in one tube.

Ansorge et al. (14, 15), Smith et al. (16), Prober et al. (17), and Connell et al. (28) have only single reporter groups per DNA fragment. This limits the amount of signal. The difluoresceinated primer increases the amount of signal, which increases sensitivity. If needed, additional fluorescins could be added to further increase sensitivity; we have tested up to three fluorescins per primer but did not need the added sensitivity.

Ansorge and Barker (29) demonstrated that for longer fragments 4% acrylamide gave better resolution than 6%. If 4% or 5% gels can be made with good well morphology, the number of readable bases in a single load may be extended even further.

A comparison of the various detection systems is interesting. Chopping the laser signal to discriminate between background and sample fluorescence as described here is unique. Ansorge et al. (14, 15), Prober et al. (17), and our method filter out the Raman shift. The limits of detection determined by Smith et al. (16) were 0.1–1 fmol. Our method detects <0.054 fmol. Ansorge et al. (15) state that their system can detect as few as 3 amol, whereas Prober et al. (17) state that they are capable of detecting 1–10 amol. Detection limits are a function of the power of the laser used as an excitation source. Higher sensitivity should be attained with a 40- to 100-mW laser instead of the 10-mW laser used in this system (14, 16). Why Ansorge et al. (15) and Prober et al. (17) have not reported longer sequencing results is unclear.

The imaging system that we use produces a two-dimensional array that looks like an autoradiogram. Ansorge et al. (14, 15), Smith et al. (16), Nagai et al. (9), and Prober et al. (17) produce a single dimensional curve for each band.

Beck and Pohl (8) were able to resolve between 500 and 600 bases. Connell et al. (28) resolve about 300 bases with 95–98% accuracy. Ansorge et al. (14, 15) were capable of resolving 400 bases and Nagai et al. (9) report a 95% accuracy in the range of 150–300 bases. Prober et al. (17) sequence 300–400 bases per sample. Our method has a lower error rate due, in part, to the two-dimensional imaging of the results.

On several occasions we have sequenced clones oriented in opposite directions with inserts of from 800 to 1100 bases. In these instances, we have had homologous overlapping sequences, which indicates the accuracy of our method. Clones up to 500 bases have been sequenced through the 5' and 3' cloning sites.
A method for sequencing DNA using a difluoresceinated primer and laser excitation is described in this paper. The sequencing protocols have been empirically determined to provide base reading from 1 to 600 bases with <1% error in a single load. This finding is different from that reported by Connell et al. (28), who stated that 400 bases was the upper limit for a single reaction.

In this method of automated, continuous DNA sequencing, the electrophoretic gel is scanned at a fixed location on the gel back and forth across the gel. This provides a horizontal input for the creation of a two-dimensional image. The vertical input is provided by the movement of bands vertically in the gel, and as such bands cross the scanning location over time, a picture is constructed line by line. The full two-dimensional appearance of the bands and background is displayed. This allows for human interpretation to utilize all aspects of the brain's image processing capabilities. In addition, it allows for a better computerized, non-human attempt at decoding the image information through the use of two-dimensional filtering, neural network analysis, and other computerized image analysis. Information that is necessary for proper interpretation is not lost as in the case of comparing four one-dimensional graphs or curves.

We thank Mim Sawtell for her careful preparation of the manuscript and Dr. K. W. Lee for photographic assistance. The research was supported by funds from Li-Cor, Inc., and reagents from Molecular Biosystems, Inc.