DNA SEQUENCING BY CAPILLARY GEL ELECTROPHORESIS AND LASER-INDUCED FLUORESCENCE DETECTION

JIAN-ZHONG ZHANG, DA YONG CHEN, HEATHER HARKE, AND NORMAN J. DOVICH

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

Abstract - Detection limits of one to two thousand analyte molecules is produced by laser-induced fluorescence with the sheath flow cuvette for DNA sequencing. Two different detector systems are developed for the sequencing of real DNA samples: a single laser and single emission spectral channel system for the sequencing technique of Richardson and Tabor and a single laser excitation and dual emission spectral channel detector for the sequencing technique of DuPont. The best detection limits (1,400 molecules) are obtained with a 750- μ w helium-neon laser ($\lambda = 544$ nm), a cooled PMT, and tetramethylrhodamine labeled primer used with the Richardson/Tabor sequencing technique. Although better detection limits have been reported for the determination of highly dilute neat dye solution, these results for DNA sequencing are unprecedented for the analysis of real samples and are made possible by the extreme specificity of the Sanger chain-terminating reaction, the extreme resolution of capillary gel electrophoresis, and the extreme detection performance of laser induced fluorescence with the sheath flow cuvette.

Key Words DNA, laser-induced fluorescence, capillary gel electrophoresis, sheath flow cuvette

DNA sequencing is a fundamental process in the biological sciences. In 1977, Sanger's group reported an enzymatic method for DNA sequencing [1]. They generated a nested set of radioactively labeled DNA fragments in four reactions, one for each terminal dideoxynucleotide. The reaction products are separated by size in adjacent lanes of a high resolution polyacrylamide gel and detected by use of autoradiography. The sequence is interpreted from the pattern of alternating bands in the lanes corresponding to the terminal base of the fragment. Sequence determination by this classic technique is an important, albeit tedious and labor intensive, task.

An advance in sequencing technology occurred in 1986-88 when Smith and co-workers in Hood's laboratory, workers in Ansorge's group, Prober and co-workers at DuPont, and Kambara and coworkers at Hitachi reported DNA sequencers that replaced the radioactive labels and autoradiography in Sanger's method with fluorescent labels and laser-based detection [2-5]. In Smith's four spectral channel, single lane sequencer, four fluorescently labeled primers are associated with the terminating dideoxynucleotide through use of separate chain terminating reactions. Fluorescence, excited by two laser lines and detected in four spectral channels, is used to identify the terminal nucleotide. In Ansorge's single spectral channel, four lane sequencer, a single fluorescent label is used with each dideoxynucleotide chain terminating reaction and the products are run on separate lanes of a slab gel; sequence identification is similar to the classic autoradiography technique. In Prober's two spectral channel, single lane sequencer, one of four fluorophores is associated with the terminating dideoxynucleotide through use of distinctly labeled dideoxynucleotides in a single reaction mixture. Fluorescence, excited by a single laser line and detected in two spectral channels, is used to identify the terminal nucleotide. Kambara's instrument is similar to that of Ansorge's except that a viticon camera is used to record the fluorescence signal. Smith's, Ansorge's, Prober's and Kambara's fluorescence sequencers have been commercialized by Applied Biosystems, Pharmacia, DuPont, and Hitachi, respectively. As an example of typical performance, the Applied Biosystems model 373A instrument runs 24 lanes simultaneously on a slab gel to produce sequencing rates of 50

bases/hour/lane or 1,200 bases/hour/slab. Similar sequencing rates are produced by other instruments. Sequence may be determined, by use of computer algorithms, to about 450 bases, which is limited by the resolution of the gel. The separation efficiency may be characterized in terms of chromatographic resolution, defined as the peak spacing divided by four times the standard deviation of the Gaussian-shaped peaks; as shown by the data contained in a recent review, the resolution produced by the best slab gel seldom exceeds 1 [6].

Increased resolution and sequencing rates are produced by operating the gels at high electric field strength. Unfortunately, the finite resistance of the separation buffer leads to unacceptable heating of conventional 200- to 400- μm thick gels at electric field strengths much greater than 50 V cm^{-1} . Gel filled capillaries have attracted interest because their high surface-to-volume ratio provides excellent heat transport properties, allowing use of very high electric field strength. Typical capillaries are 50 to 75 μm inner diameter and 25 to 100 cm long, although early work on RNA separations used cellulose fibers of 10 μm diameter and 25 mm length [7-8]. More recently, Hjertén's and Karger's groups have reported the use of gel filled capillaries for protein separations [9-10]. Karger's group and others have reported the use of gel filled capillaries for separation of oligonucleotide standards with detection by ultraviolet absorbance [11-14]. An important issue is the performance of the gel filled capillaries under high electric field strength; patents have been issued on the use of a bifunctional silane reagent to bind the acrylamide to the wall of the capillary to improve the gel stability [15-17].

Capillary gel electrophoresis has been applied to the separation of fluorescently labeled DNA sequencing fragments. Swerdlow and Gesteland reported the separation of a single dideoxynucleotide reaction mixture with a single spectral channel laser-induced fluorescence detector [18]. Resolution of the capillary based system was superior to that of a slab gel by a factor of 2.4. Drossman and co-workers in Smith's laboratory reported the separation of a single reaction mixture in gel-filled capillaries [19]. When operated at 400 V/cm , the system produced sequencing rates of 1,000 bases/hour after elution of the primer, a roughly twenty five times higher sequencing rate than produced by conventional slab gel electrophoresis. A subsequent report from Smith's laboratory described the use of capillary gel electrophoresis for the separation of the reaction products of the four spectral channel, single lane sequencing system [20]. Resolution was between a factor of 1.5 to 2 times superior to slab-gel data. Karger's laboratory reported separation of DNA fragments, labeled with a single fluorophore, at 350 V/cm yielding sequencing rates of 450 bases/hour; base line resolution was obtained for fragments up to 330 nucleotides in length [21].

Laser-induced fluorescence detection in slab-gel electrophoresis requires low detection limits, typically on the order of 1 to 100 attomoles per band [2]. However, the detection requirements in capillary gel electrophoresis are more severe. Based on a comparison of the cross-section of a 50- μm capillary with the cross-section of a loading well in a slab-gel, Smith has estimated that the sample loading in capillary gel electrophoresis will be on the order of 1 to 10 attomoles of DNA per band [19]. Detection limits in the zeptomole range (1 zeptomole = 10^{-21} mole = 600 molecules) are necessary for accurate DNA sequence determination by capillary gel electrophoresis. Detector design is complicated further by the requirement that the detection volume be less than 1 nL to preserve the separation efficiency of the nano-scale electrophoresis technique. Swerdlow and Gesteland's system used an on-column laser-induced fluorescence detector with detection limits of 200 zmol of fluorescently labeled product [18]. Smith and co-workers described the use of one- and four-spectral channel on-column fluorescence detectors; both the single- and four-spectral channel detectors produced detection limits of about 100 zmol of labeled fragment [19, 20]; Karger's group has reported similar detection limits for their single spectral channel detector [12]. This group has reported a post-column single spectral channel fluorescence detector in the gel electrophoresis separation of the products of a single chain terminating reaction; detection limits were 10 zmol [21].

TWO SPECTRAL CHANNEL SEQUENCING. The DuPont sequencing technique uses succinylfluorescein dyes to label the four dideoxynucleotides. These dyes have relatively closely spaced absorbance and emission bands. As a result, a single excitation wavelength, 488 nm, is used to excite fluorescence from all dyes. Emission is distributed between two spectral channels centered at 510 and 540 nm. The ratio of the fluorescence intensity in the two spectral channels is used to identify the terminating dideoxynucleotide.

Our two-spectral channel capillary gel electrophoresis detector is based on a post-column sheath-flow detector to reduce the background signal (Fig. 1). A 5-mw argon ion laser beam ($\lambda = 488 \text{ nm}$) was focused with a $4 \times$ microscope objective about 200- μm below the exit of the capillary in a sheath flow cuvette with 200- μm square flow chamber and 2-mm thick windows [10]. Fluorescence was collected at right angles with a $32\times$, 0.6 numerical aperture microscope objective and imaged onto a 0.75-mm diameter pinhole. A single interference filter (525 nm

center wavelength, 40 nm band pass) transmitted fluorescence. A dichroic filter, centered at 525 nm, was used to direct fluorescence in two spectral channels to two R1477 photomultiplier tubes

operated at 1200 V; the transmitted light was centered at 535 nm (red) and the reflected light was centered at 515 nm. The signals were passed through a 1-Hz low pass filter, digitized at 10 Hz with a computer, and passed through a quadratic-cubic polynomial filter.

In this detector, similar to that used in flow cytometry, the sample passes as a narrow stream in the center of a square flow chamber, surrounded by a sheath stream consisting of conducting buffer, typically 1X TBE. The high optical quality quartz windows of the cuvette produce much less light scatter than does an on-column detector. We have reported the use of a single spectral channel fluorescence detector based on the cuvette for capillary zone electrophoresis separation of zeptomole

quantities of fluorescently labeled amino acids and for capillary gel electrophoresis separation of zeptomole quantities of the products of a single dideoxynucleotide reaction [23-25]; Keller's group has reported high sensitivity fluorescence detectors based on the cuvette [26-28].

The detection limit for this instrument is 20 zmol for a labeled dideoxynucleotide triphosphate and 5 zmol (3,000 molecules) for a 100-mer fluorescein-labeled oligonucleotide. The difference in detection limit between the monomer and the 100-mer is related to the very low sheath flow rate employed to transport analyte from the capillary to the illumination region. During the transit time, analyte can diffuse an appreciable amount, decreasing its effective concentration. Larger fragments undergo less diffusion and are more concentrated in the probe volume. As a result, the detection limit for the system improves for larger DNA fragments, which are produced in sequencing reactions at lower concentration compared with early eluting fragments.

A typical separation on a 4%T gel, at 465 V/cm, of the fragments from an M13mp18 template yields sequencing rates of 1,000 bases/hour; early termination of the sequence is an artifact of the reaction. In the region from 60 to 100 nucleotides, the peaks are nearly base line resolved (Fig. 2). The sample was prepared from a DuPont Genesis 2000 protocol using 3ug M13mp18 single stranded DNA, 15 ng -40 17-mer M13 primer, and 1 µl sequenase (US Biochemicals) in a standard reaction mix. They were ethanol precipitated, washed, and resuspended in 3 µl of a 49:1 mixture of formamide:0.5M EDTA at pH 8.0. The 34 cm long, 50-µm inner diameter, 190-µm outer diameter capillary was filled with 4%T, 5%C acrylamide gel that was covalently bound to the capillary wall through use of γ-methacryloxypropyltrimethoxy silane [6]. The sample was injected at 100 V/cm for 30 seconds; after injection, the sample was replaced with a fresh vial of 1× TBE. The electrophoresis continued for 1 min at 100 V/cm. The capillary was then trimmed at the injection end by 1-mm and the voltage was increased in 1 kV increments over a 14 minute period to a total electric field strength of 465 V/cm. The sheath stream was 1× TBE at a flow rate of 0.16 ml/hr. Time is measured from the injection; the numbers at the top of the figure correspond to the nucleotide length including the universal 17-mer primer. The known sequence is printed above. The solid trace corresponds to emission at wavelengths longer than 525 nm and the dashed trace corresponds to emission at wavelengths shorter than 525 nm.

ONE SPECTRAL CHANNEL SEQUENCING. In 1990, Richardson and Tabor, and independently Ansorge, reported a sequencing technique based on a single fluorophore; by varying the amount of dideoxynucleotide in the reaction mixture, each base is identified with a particular fluorescence peak height during separation in a single lane of an acrylamide gel [29-30]. This technique relies on uniform labeling of the reaction products through use of the manganese/T7

polymerase reaction. In the work that has been published, a fluorescein labeled primer is excited with an argon ion laser at 488 nm; emission is detected in a single spectral band.

In our single spectral channel sequencer, a TAMRA-labeled primer is excited by a green helium-neon laser ($\lambda = 543.5 \text{ nm}$) (Fig 3). The laser beam was focused with a 4 $\times$ microscope objective about 200- μm below the exit of the capillary. Fluorescence was collected at right angles with an 18 $\times$, 0.45 numerical aperture objective and imaged onto a 0.8-mm diameter pinhole. A single interference filter (590 nm center, 40 nm band pass) was used to block scattered laser light. Fluorescence was detected with an R1477 photomultiplier tube,

cooled to -25°C . The photomultiplier tube output was passed through a 1-s RC low-pass filter and digitized by a computer. The signal is digitized at 2 Hz and passed through a quadratic-cubic polynomial filter before display.

As in the two spectral channel sequencer, a sheath flow cuvette is used as a post-column fluorescence detector to minimize background light scatter. Further reduction in the background signal comes from the relatively long excitation wavelength and low power excitation beam (750 μW). In fact, the major contribution to background signal often is dark current produced by the photomultiplier tube. This contribution to the background signal is minimized by cooling the photomultiplier tube to -25°C .

The standard deviation in the background signal for this system corresponds to 700 ymol (1 yoctomole = 1 ymol = 10^{-24} mol) of labeled primer introduced onto the capillary; detection limits are, by definition, a factor of three higher (1,200 molecules). As with the two-spectral channel detector, detection limits improve for higher molecular weight fragments. These detection limits are associated with the excellent spectral properties of the TAMRA-fluorophore and reflect the simple detector design allowed by the single spectral channel sequencing technique.

This single spectral channel sequencing protocol, when applied to M13mp18, produces compressions in our standard capillary sequencing gels at room temperature. To minimize compressions, 30% by volume formamide was incorporated in the 6%T, 5%C, 7M urea gel for separation of the sequencing products. Fig. 4 presents the elution of fragments ranging from 60 to 100 nucleotides in length. The sequencing reaction was carried out in 40 mM MOPS buffer, pH 7.5, 50 mM NaCl, 10 mM MgCl_2 , and

15 mM sodium isocitrate. Dye labeled primer (Applied Biosystem 21M13 TAMRA, 1.6 pM) was annealed to 1 μg of M13mp18 single stranded DNA at 65°C for 2 min followed by slow cooling. A mixture of deoxy- and dideoxynucleoside triphosphates was added to give an average nucleotide ratio (dNTP/ddNTP) of 1200:1 with 7-deaza-2'-deoxyguanosine-5'-triphosphate used in place of dGTP. The ratios of nucleotides were adjusted to yield a nominal peak height ratio of 8:4:2:1 for A, C, G, T. The mixture was warmed to 37°C and 6 units of Sequenase Version 2.0 and 0.006 units of pyrophosphatase were added. Incubation continued at 37°C for 30 min, after which the DNA was precipitated with ethanol [29]. The 37 cm long, 50- μm inner diameter, 190- μm outer diameter capillary was filled with 6%T, 5%C, 30% formamide, 7M urea gel that was covalently bound to the capillary wall through use of γ -methacryloxypropyltrimethoxysilane. The sample was injected at 200 V/cm for 30 seconds; after injection, the sample was replaced with a fresh vial of 1 $\times$ TBE. The capillary tip was not trimmed after injection and separation proceeded at 200 V/cm. The sheath stream was 1 $\times$ TBE at a flow rate of 0.16 ml/hr. The numbers at the top of the figure correspond to the nucleotide length including the -21 18-mer primer.

The data, obtained at modest electric field strength, 200 V/cm, suffers from low migration rate, 80 bases/hour, and reflects the low mobility of DNA in the formamide-urea gels. The peaks across this region are nearly base-line resolved; the resolution ranges from 1.2 to 1.8, with an

average of about 1.5. This recipe reduced greatly the compressions but yielded a relatively slow separation rate. The sequencing accuracy of this gel appears to be poor, particularly with respect to C and G discrimination, reflecting difficulties in the reaction chemistry. Improved performance should be possible with elimination of ghost peaks and optimization of the dideoxynucleotide ratios.

CONCLUSIONS

A post-column detector produces lower background signal and superior detection limits than an on-column fluorescence detector. In the best case, the detector performance is dominated by dark-current; cooling the PMT produces a factor of two improvement in detection limit. Photobleaching is important for these fluorescent dyes; microwatt power lasers produce outstanding detection limits.

Several additional issues must be addressed in the development of capillary gel electrophoresis for high speed DNA sequencing. Either gel stability must be increased or the cost and time associated with capillary replacement must be reduced so that unattended, automated runs are possible. It will be necessary to operate several capillaries simultaneously to produce enhanced sequencing rates for large-scale sequencing projects; this laboratory has proposed the development of a 32-capillary instrument to produce sequencing rates approaching 64,000 bases/hour. Automated base calling algorithms must be incorporated into the sequencing instrumentation. Last, the instrumentation must be evaluated with a large number of samples to assess and improve the sequencer accuracy.

ACKNOWLEDGMENTS

This work was funded by the Natural Sciences and Engineering Research Council of Canada, the Department of Chemistry of the University of Alberta, Pharmacia Inc., Waters Division of Millipore Inc., and US Biochemical Corporation. H.H. acknowledges a predoctoral fellowship from the Alberta Heritage Foundation for Medical Research.

REFERENCES

1. F. Sanger, S. Nicklen, A.R. Coulson, *Proc. Nat. Acad. Sci.* **74**, 5463 (1977).
2. L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, and L.E. Hood, *Nature* **321**, 674 (1986).
3. W. Ansorge, B.S. Sproat, J. Stegemann, C. Schwager, *J. Biochem. Biophys. Meth.* **13**, 315 (1986).
4. J.M. Prober, G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen, and K. Baumeister, *Science* **238**, 336 (1987).
5. H. Kambara, T. Nishikawa, Y. Katayama, and T. Yamaguchi, *Biotechnology* **6**, 816, (1988).
6. G.L. Trainor, *Anal. Chem.* **62**, 418 (1990).
7. J.E. Edström, *Nature* **172**, 809 (1953).
8. J.E. Edström, *Biochimica Biophysica Acta* **22**, 378 (1956).
9. S. Hjertén, *J. Chromatogr.* **270**, 1 (1983).
10. A.S. Cohen and B.L. Karger, *J. Chromatogr.* **397**, 409 (1987).
11. A.S. Cohen, D.R. Najarian, A. Paulus, A. Guttman, J.A. Smith, and B.L. Karger, *Proc. Nat. Acad. Sci.* **85**, 9660 (1988).
12. A. Guttman, A.S. Cohen, D.N. Heiger, and B.L. Karger, *Anal. Chem.* **62**, 137 (1990).
13. A. Paulus, E. Gassmann, M.J. Field, *Electrophoresis* **11**, 702 (1990).
14. H.F. Yin, J.A. Lux, and G. Schomburg, *J. High Resolution Chromatography* **13**, 624 (1990).
15. P.F. Bente and J. Myerson, U.S. Pat. 4,810,456 (1989).
16. B.L. Karger and A.S. Cohen, U.S. Pat. 4,865,706 (1989).
17. B.L. Karger and A.S. Cohen, U.S. Pat. 4,865,707 (1989).
18. H. Swerdlow and R. Gesteland, *Nucleic Acids Res.* **18**, 1415 (1990).
19. H. Drossman, J.A. Luckey, A.J. Kostichka, J. D'Cunha, and L.M. Smith, *Anal. Chem.* **62**, 900 (1990).
20. J.A. Luckey, H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris, and L.M. Smith, *Nucleic Acids Res.* **18**, 4417 (1990).
21. A.S. Cohen, D.R. Najarian, and B.L. Karger, *J. Chromatogr.* **516**, 49 (1990).
22. H. Swerdlow, S. Wu, H. Harke, and N.J. Dovichi, *J. Chromatogr.* **516**, 61 (1990).
23. S. Wu and N.J. Dovichi, *J. Chromatogr.* **480**, 141 (1989).
24. Y.F. Cheng and N.J. Dovichi, *Science* **242**, 562 (1988).
25. Y.F. Cheng, S. Wu, D.Y. Chen, and N.J. Dovichi, *Anal. Chem.* **62**, 496 (1990).
26. N.J. Dovichi, J.C. Martin, J.H. Jett, R.A. Keller, *Science* **219**, 845 (1983).
27. N.J. Dovichi, J.C. Martin, J.H. Jett, M. Trkula, R.A. Keller, *Anal. Chem.* **56**, 348 (1984).
28. D.C. Nguyen, R.A. Keller, J.H. Jett, J.C. Martin, *Anal. Chem.* **59**, 2158 (1987).
29. S. Tabor and C.C. Richardson, *J. Biological Chemistry* **265**, 8322 (1990).
30. W. Ansorge, J. Zimmermann, C. Schwager, J. Stegemann, H. Erfle, and H. Voss, *Nucleic Acids Res.* **18**, 3419 (1990).