Anal. Chem. **1996,** *68,* 936−939

Postelectrophoresis Capillary Scanning Method for DNA Sequencing

Suhyeon Kim, Hyun Ju Yoo, and Jong Hoon Hahn*

Department of Chemistry, Pohang University of Science and Technology, San 31 Hyoja Dong, Pohang, 790-784, South Korea

A postelectrophoresis capillary scanning method for increasing the throughput of DNA sequencing has been developed. The method features a spatially and temporally separated arrangement between capillary gel electrophoresis separation of DNA sequencing fragments and the visualization of the separation pattern. Fluorescently labeled DNA sequencing fragments of pBluescript SK(−) were produced in enzymatic chain-termination reactions and separated by capillary gel electrophoresis using a 110 cm fused silica capillary with a transparent polymer coating. A partial sequence from the DNA was then determined by repeated scanning of the capillary at a rate of 8.39 cm/s, with changing excitation wavelengths and spectral filter sets.

The Human Genome Project (HGP) aims at deciphering the complete instructions for making a human being. Since it essentially demands the sequencing of an entire genome, primary emphasis has been placed on developing a high-speed, high-throughput DNA sequencing method.[1] Recent large-scale sequencing that employs a random selection strategy and fluorescence-based automated sequencing instruments is approaching a throughput of 1 megabase per year.[2,3] This is, however, far behind the rate needed for the success of the HGP.

In prevalent DNA sequencing methods, nested sets of radioactively or fluorescently labeled DNA fragments are produced by enzymatic termination reactions and are separated by size with polyacrylamide slab−gel electrophoresis (SGE).[4−7] Replacing conventional SGE by capillary gel electrophoresis (CGE) has been considered to be a promising approach for enhancing the rate of DNA sequencing.[8−13] Because of the high surface-to-volume ratio of gel-filled capillaries, CGE allows the use of very high electric field strength and provides an increased sequencing rate. A sequencing rate 25 times faster than that possible with SGE has been demonstrated.[12] However, the throughput of DNA sequencing methods using CGE is about the same as that of SGE-based methods, because only one capillary can be analyzed at a time. Recent developments of highly multiplexed DNA sequencing based on parallel arrays of capillaries in CGE have transcended this limitation.[14−17]

Here we report a new, simple approach for increasing the throughput of CGE-based DNA sequencing. The method features a spatially and temporally separated arrangement between CGE separation of DNA sequencing fragments and the visualization of the separation pattern, which is very similar to conventional DNA sequencing where SGE is followed by autoradiography. After fluorescently labeled DNA fragments produced in enzymatic sequencing reactions are size-separated by CGE, the capillary containing all bands of the fragments is scanned longitudinally with a laser beam whose wavelength is suitable for the excitation of the fluorescent labels. The DNA sequence is then determined from the band pattern obtained by detecting fluorescence signals generated from the bands during the scan. By employing multiple capillaries and a high-voltage power supply with a large load capacity, sequencing reaction products of a great number of different DNA samples can be separated all at one time. The sequences of the DNA samples can be obtained by feeding the run capillaries into the capillary scanner one by one. Therefore, in this DNA sequencing method, the throughput is limited, in principle, by the scanning time per capillary. A similar technique of capillary scanning has been used in capillary isoelectric focusing.[18]

Our postelectrophoresis capillary scanning system is very simple in scanning capillaries and detecting fluorescence signals, consisting of a device for pulling a capillary, optics for guiding and focusing an excitation laser beam, and a part for collecting and detecting fluorescence. In addition to this simplicity, it enjoys a great advantage over the capillary array systems. A single unit of the capillary scanning system can be shared by many labora-

(1) Smith, L. M. *Science* **1993**, *262*, 530−532.
(2) Hunkapiller, T.; Kaiser, R. J.; Koop, B. F.; Hood, L. *Science* **1991**, *254*, 59−67.
(3) Sulston, J.; Du, Z.; Thomas, K.; Wilson, R.; Hiller, H.; Staden, R.; Halloran, N.; Green, P.; Thierry-Mieg, J.; Qiu, L.; Dear, S.; Coulson, A.; Craxton, M.; Durbin, R.; Berks, M.; Metzstein, M.; Hawkins, T.; Ainscough, R.; Waterston, R. *Nature* **1992**, *356*, 37−41.
(4) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463−5467.
(5) Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. *Nature* **1986**, *321*, 674−679.
(6) Ansorge, W.; Sproat, B. S.; Stegemann, J.; Schwager, C. *J. Biochem. Biophys. Methods* **1986**, *13*, 315−317.
(7) Prober, J. M.; Trainor, G. L.; Dam, R. J.; Hobbs, F. W.; Robertson, C. W.; Zagursky, R. J.; Cocuzza, A. J.; Jensen, M. A.; Baumeister, K. *Science* **1987**, *238*, 336−341.
(8) Drossman, H.; Luckey, J. A.; Kostichka, A. J.; D'Cunha, J.; Smith, L. M. *Anal. Chem.* **1990**, *62*, 900−903.
(9) Swerdlow, H.; Gesteland, R. *Nucleic Acids Res.* **1990**, *18*, 1415−1419.
(10) Cohen, A. S.; Najarian, D. R.; Karger, B. L. *J. Chromatogr.* **1990**, *516*, 49−60.
(11) Luckey, J. A.; Drossman, H.; Kostichka, A. J.; Mead, D. A.; D'Cunha, J.; Norris, T. B.; Smith, L. M. *Nucleic Acids Res.* **1990**, *18*, 4417−4421.
(12) Swerdlow, H.; Zhang, J. Z.; Chen, D. Y.; Harke, H. R.; Grey, R.; Wu, S.; Dovichi, N. J.; Fuller, C. *Anal. Chem.* **1991**, *63*, 2835−2841.
(13) Chen, D. Y.; Harke, H. R.; Dovichi, N. J. *Nucleic Acids Res.* **1992**, *20*, 4873−4880.
(14) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 967−972.
(15) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 2149−2154.
(16) Kambara, H.; Takahashi, S. *Nature* **1993**, *361*, 565−566.
(17) Ueno, K.; Yeung, E. S. *Anal. Chem.* **1994**, *66*, 1424−1431.
(18) Wang, T.; Hartwick, R. A. *Anal. Chem.* **1992**, *64*, 1745−1747.

0003-2700/96/0368-0936$12.00/0

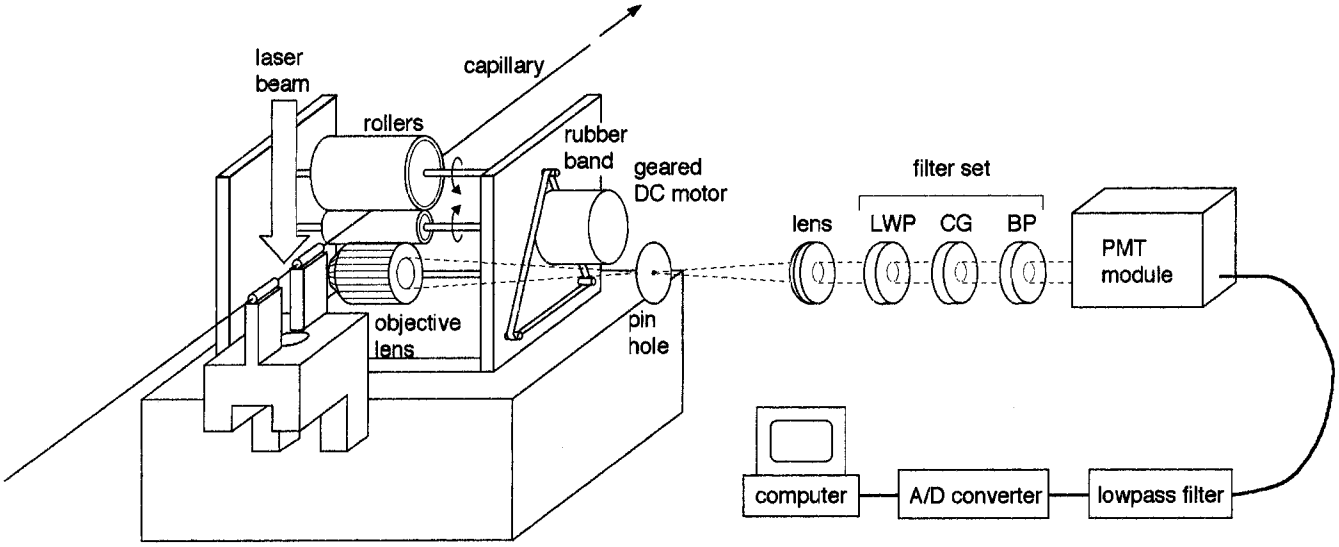


Figure 1. Schematic of the postelectrophoresis capillary scanning system. LWP, long-wave-pass filter; CG, colored glass filter; BP, band-pass filter.

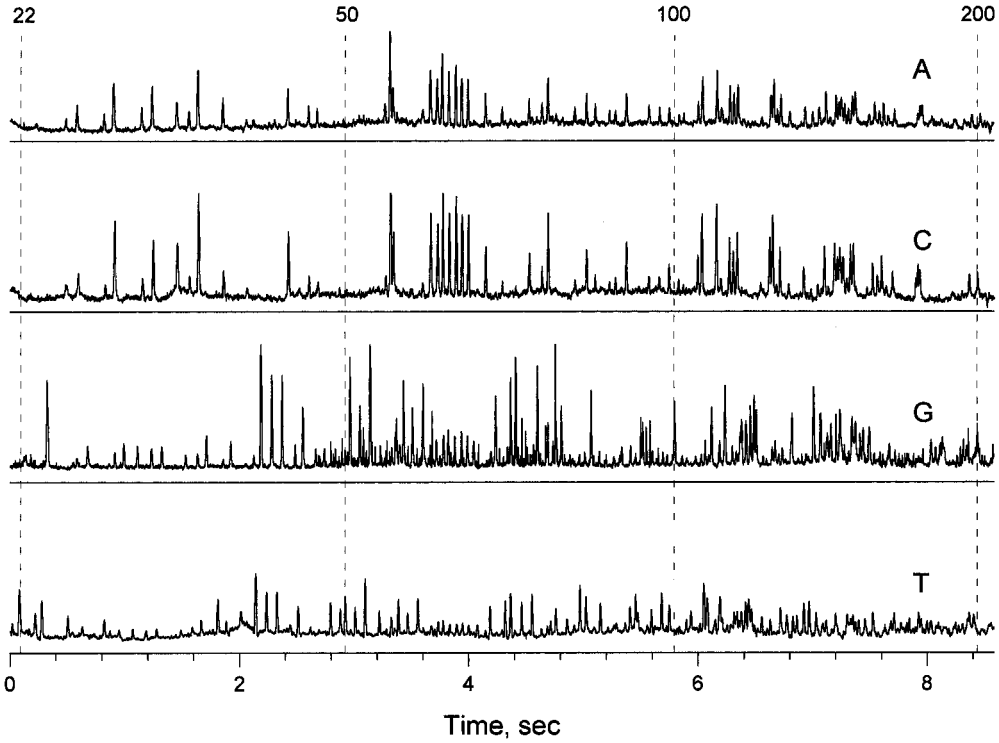


Figure 2. Four electropherograms obtained by repeated scanning of a 110 cm capillary containing separated bands of DNA fragments of pBluescript SK(−). The numbers on the top indicate the base numbers.

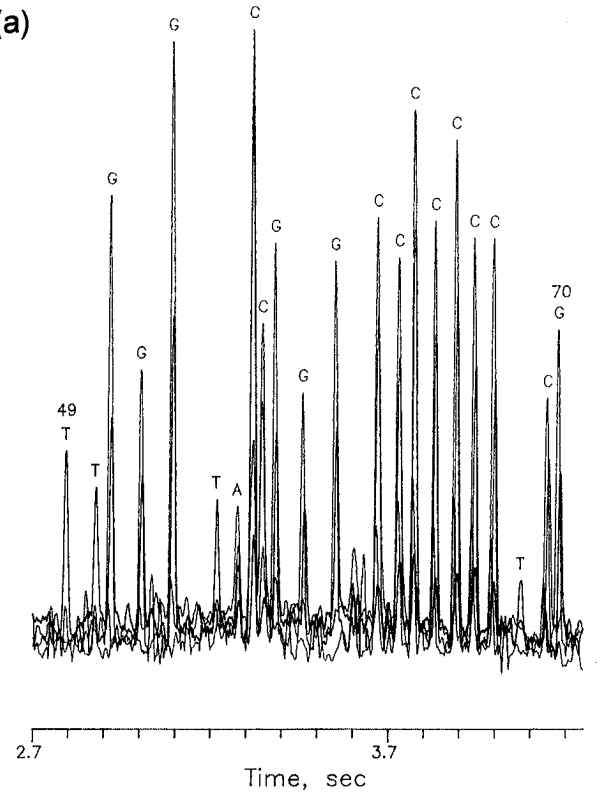
tories, each of which owns a very simple apparatus for electrophoretic separations of DNA sequencing fragments using multiple capillaries. This advantage also gives us versatility in choosing electrophoresis conditions, capillary lengths, and number of capillaries per electrophoresis run, whereas in capillary array electrophoresis (CAE), these conditions should be fixed for one batch of capillaries. However, in our previous work,[19] a scan speed of only 0.58 cm/s was obtained using a preliminary capillary scanning setup. The DNA sequencing throughput corresponding to this rate is at least 6 times lower than that CAE can achieve.[15] We have increased the scan rate of the postelectrophoresis capillary scanning system by equipping it with a digitized data acquisition system and achieved a 15-fold enhancement in scan rate. A multicolor detection capability has also been provided to the system by employing replaceable spectral filter sets in fluorescence collection.

EXPERIMENTAL SECTION

Experimental details of this work are nearly the same as those of our previous work.[19] Therefore, only the improvements and modifications are described in detail. Figure 1 shows the capillary scanning system newly equipped with a replaceable spectral filter set and a digitized data acquisition system. A fused silica capillary with a transparent polymer coating (100 μm i.d., 375 μm o.d.) (TSU100375, Polymicro Technologies, Phoenix, AZ), which con-

(19) Kim, S.; Yoo, H. J.; Nam, H.-G.; Hahn, J. H. *Chromatographia* **1995**, *40*, 345−349.

(a)



(b)

M13 primer

TGTAAAACGA CGGCCAGTGA	ATTGTAATAC	GACTCACTAT	40
AGGGCGAATT GGGTACCGGG	CCCCCCCTCG	AGGTCGACGG	80
TATCGATAAG CTTGATATCG	AATTCCTGCA	GCCCGGGGGA	120
TCCACTAGTT CTAGAGCGGC	CGCCACCGCG	GTGGAGCTCC	160
AGCTTTTGTT CCCTTTAGTG	AGGGTTAATT	TCGAGCTTGG	200

Figure 3. (a) Part of the electropherogram obtained by overlapping the four electropherograms in Figure 2. (b) Partial sequence from pBluescript SK(−) DNA. The first boxed part of the sequence corresponds to −21M13 primer. The sequence determined in (a) is equal to the sequence in the second boxed part.

tains fluorescently labeled DNA fragments separated by CGE, is inserted through two aligned poly(ether ether ketone) (PEEK) tubings and then lightly clamped between two aluminum rollers. A heat−shrinkable plastic film evenly covers the surface of each roller. The capillary is pulled through the system by rotating the lower roller with a geared DC motor. The upper roller gently presses down the lower one with rubber bands.

An argon ion laser is used as the excitation source for laser-induced fluorescence detection. The laser light is focused onto a moving capillary at the middle of the line connecting the two PEEK tubings. The fluorescence generated from DNA fragments is collected at right angles with the excitation laser beam by using a microscope objective and focused onto a pinhole. The spatially filtered light is collimated with a plano-convex lens of 100 mm focal length (PLCX-25.4/77.3UV, CVI Laser, Albuquerque, NM) and then filtered with a spectral filter set which has a defined transmittance profile. A filter set consists of a long-wave-pass filter, a colored glass filter, and a band-pass filter. Long-wave-pass filters and colored glass filters are employed to block the scattered light emanating from capillary walls. Since the excitation wavelength used in the determination of cytosine (C) or adenine (A) sequence is 488 nm, a long-wave-pass filter (LWP-510-1.00, CVI Laser) and a colored glass filter (OG-515, CVI Laser) with cut-on wavelengths of 510 and 515 nm are used, respectively. In the guanine (G) or thymine (T) scan, a long-wave-pass filter (LWP-550-1.00, CVI Laser) and a colored glass filter (OG-550, CVI Laser) with the same cut-on wavelength of 550 nm are used to block the scattered light of the 514 nm excitation laser beam. Band-pass filters 10 nm in width, centered at 540, 560, 580, and 610 nm (F10-540-1, F10-560-1, F10-580-1, and F10-610-1, respectively, CVI Laser), selectively transmit the fluorescence generated from C, A, G, and T fragments to a photomultiplier (PMT) detector, respectively. High-frequency noise in the output of the PMT module is filtered out using a homemade second-order Bessel low-pass filter (cut-off frequency, 250 Hz). The signal is then digitized at 450 Hz using an IBM PC 386-compatible computer equipped with a 12 bit analog-to-digital converter (PCL-812PG, Advantech, Taiwan).

Chain-terminated pBluescript SK(−) DNA sequencing fragments were generated using a Sequenase 2.0 sequencing kit (United States Biochemical Corp., Cleveland, OH) and fluorescently labeled −21M13 primers. FAM-, JOE-, TAMRA-, and ROX-tagged primers (Applied Biosystems, Foster City, CA) were used in sequencing reactions for C, A, G, and T base fragments, respectively. Capillaries filled with 9% T, 0% C polyacrylamide gel were prepared using the procedure developed by Karger and co-workers.[20−22] The DNA sequencing fragments were separated by CGE with a high voltage of 250 V/cm for about 50 min.

RESULTS AND DISCUSSION

Figure 2 shows four electropherograms of A, C, G, and T base fragments of pBluescript SK(−) DNA. The electropherograms were obtained by separating a mixture of A, C, G, and T base fragments of the DNA using a 110 cm capillary, followed by scanning the capillary four times using a proper excitation wavelength and filter set for each scan. The fragments with single base length difference are well resolved up to 150 nucleotides. The numbers of theoretical plates are 4.5×10^6 (at base 22) and 1.9×10^6 (at base 100) over effective column lengths of 104 and 58 cm, respectively. It takes 12.5 s to scan the whole length of the capillary. Therefore, the scan rate is 8.39 cm/s, which is about 15 times faster than the previous result of 0.58 cm/s.[19] The relative standard deviation in scan rate is 0.44%. The dynamic noise due to fluctuations in capillary position during a scan is about twice as big as static noise. The background level during a scan is also 4−16% higher than that in static condition. We have compared static background levels at both excitation wavelengths for the transparent coated capillaries used in this work with those for uncoated fused silica capillaries prepared by removing standard polyimide coatings and observed no significant difference in the background level, which confirms that the transparent polymer coatings are practically nonfluorescent under our experimental conditions. In general, significant photobleaching has not been observed up to the 10th scan of a capillary.

A partial sequence from pBluescript SK(−) DNA was determined by identifying the color of the highest intensity at each nucleotide position in the electropherograms in Figure 2. Figure 3a shows a combined partial electropherogram from base 49 to

(20) Cohen, A. S.; Najarian, D. R.; Paulus, A.; Guttman, A.; Smith, J. A.; Karger, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9660−9663.
(21) Guttman, A.; Cohen, A. S.; Heiger, D. N.; Karger, B. L. *Anal. Chem.* **1990**, *62*, 137−141.
(22) Heiger, D. N.; Cohen, A. S.; Karger, B. L. *J. Chromatogr.* **1990**, *516*, 33−48.

base 70, which demonstrates that the sequence determined agrees well with the known sequence (the second boxed part in Figure 3b). The sequence was determined up to 200 bases. For over 200 bases, very low signal-to-noise ratio (S/N) significantly interferes with a clear determination. Rather disappointing S/N seems to be due to both the low sensitivity of the detection system and the dynamic noise. Figure 3a illustrates a serious effect of the low S/N on the determination of a DNA sequence. There are two ghost peaks between the base 59 and base 60 peaks, which are comparable to the base 68 peak in height. We could exclude the two peaks in base calling, because of their abnormality in relative color−intensity ratios among combined peaks. The poor read lengths of Figure 2 result from not only the low S/N but also from our lack of experience in DNA sequencing reactions and gel preparation. There is also an intrinsic factor that limits read lengths. In our postelectrophoresis method, the location of each fragment band is proportional not to its base number, like in the cases of CAE and CGE with standard on-column detection schemes, but to the logarithm of the base number.

At the present scan rate, a length of about 300 m can be scanned in 1 h. However, since mounting a capillary on the scanner between two sequential scans and changing spectral filter sets also take time, the number of 1 m long capillaries that can be scanned in 1 h would be considerably less than 300, unless a fast capillary feeder and an automated filter changer are provided. Other drawbacks of the present scheme of our capillary scanning method, such as the inability to determine the four-base sequence with a single scan and the time needed for the prerequisite electrophoresis, should be also considered in estimating a sequencing rate. The present scan rate is limited by the speed of the cheap DC motor and the slow sampling rate of the analog-to-digital converter (450 samples/s). We expect that it would not be technically very difficult to achieve a scan rate of over 20 cm/s by employing faster systems for capillary scanning and data acquisition. It has been reported that CAE can analyze 100 capillaries in 1 h.[17] Therefore, a capillary scanning system with an increased scan rate, a fast capillary feeder, and simultaneous four-color detectability will compete well with CAE in the number of capillaries that can be analyzed in unit time.

This work has demonstrated the feasibility of the postelectrophoresis capillary scanning technique for a high-throughput DNA sequencing method. Since the technique is still in an early stage, many improvements should be made to make it more practical. For example, the present detection scheme, limited to single-color detection per scan, must be extended to a simultaneous multicolor detection system for the determination of four-base sequence with a single scan. Enhancements in sensitivity and readability are also required.

ACKNOWLEDGMENT

The authors are grateful to the Ministry of Education (South Korea), the Biotech 2000 Program (directed by the Ministry of Science and Technology, South Korea), and the Center for Biofunctional Molecules (POSTECH, South Korea) for the financial support of this work.

Received for review August 10, 1995. Accepted November 30, 1995.[⊗]

AC9508107

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1996.