ELSEVIER

Journal of Chromatography A, 960 (2002) 229–239

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Multiplexed capillary electrophoresis for DNA sequencing with ultra violet absorption detection

Wenwan Zhong, Edward S. Yeung*

Ames Laboratory–USDOE and Department of Chemistry, Iowa State University, 332 Wilhelm Hall, Ames, IA 50011, USA

Abstract

DNA sequencing is performed in a multiplexed capillary electrophoresis system by UV absorption detection. Four individual electropherograms are obtained by simultaneously running the unlabeled DNA products of the four ddNTP-terminated reactions in the capillary array. The sequence of the template used in the cycle-sequencing reaction can be determined by overlaying the four electropherograms. Two internal standards are employed to adjust for the variance in migration times among the capillaries. After applying the correction algorithm, base calling can be done at a high level of confidence. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: DNA sequencing; Capillary array electrophoresis; DNA

1. Introduction

Capillary gel electrophoresis (CGE) has become an important technique in DNA sequencing because of its high speed, high resolution, flexibility, and possibility of building an integrated and automated system. In order to decipher the DNA sequence, radioactive or fluorescent labeling of the DNA fragments created by Sanger's chain termination reaction [1] is required for applying standard detection methods. Since autoradiography is labor intensive and can pose safety concerns, laser-induced fluorescence (LIF) has replaced it as the main detection method in DNA analysis. Extensive research work has been done to develop better detection schemes in fluorescence-based DNA sequencing, such as four-channel or two-channel detection using four dyes and single-color, intensity-based detection [2–5]. Wavelength selection in the emission or excitation spectra and fluorescence lifetime measurements in the time domain or frequency domain are the major methods to discriminate among the different colored labels on the DNA fragments [6]. However, fluorescence detection has its drawbacks too. It requires expensive equipment and reagents. Furthermore, a fluorescence label affects the fractional size and charge of the DNA fragments, which results in mobility shifts among the labeled fragments in CGE [7]. Mobility shifts make data processing more complicated and lead to errors in base calling. Compared to LIF, UV absorption detection is not as sensitive, but the instrumental setup is much more simple and less expensive. It is easier to operate and maintain because of the use of a UV lamp rather than a laser system. No dye-label is required when using UV absorption detection in DNA analysis, since DNA has strong absorption at 254 nm. A 100 bp DNA has essentially 100 absorbers per fragment. Finally, mobility shift should not be a problem because no labels are present.

*Corresponding author. Tel.: +1-515-294-6342; fax: +1-515-294-0105.

E-mail address: yeung@ameslab.gov (E.S. Yeung).

0021-9673/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0021-9673(01)01393-0

Because there are no dye labels, the DNA products from the four individual termination reactions must be run at four different capillaries in order to assemble the sequence. This is analogous to radioactive labeling and infrared single-label sequencing in slab-gel sequencing. The use of four separate capillaries at a time means that a multiplexed capillary system is the only way to achieve high-speed, high-efficiency, and high-throughput DNA sequencing. A novel absorption detection method has been applied to multiplexed capillary electrophoresis in our group [8]. The system has proven to be very reliable and efficient in many applications, such as screening of enzyme activity, peptide mapping of proteins and genetic typing [9–11]. Here, we will demonstrate that the same instrumentation can be used in DNA sequencing. Because detection is based on UV absorption, sieving matrixes like polyacrylamide [12] or poly(vinyl pyrrolidone) [13] cannot be employed. Instead, we use a new dynamic sieving matrix based on the self-assembly of monomeric surfactants into large aggregates under certain conditions [14].

2. Materials and methods

2.1. Chemicals

All chemicals for preparing running buffer solutions were purchased from Sigma (St. Louis, MO). It contained 89 m*M* Tris, 50 m*M* TAPS, 20 m*M* histidine, 2 m*M* EDTA and 7 *M* urea in deionized water and was filtered with a 0.22 μm cellulose membrane filter from Corning (Corning, NY). The chemicals for cycle-sequencing buffer ($MgCl_2$ and Tris) were purchased from Fisher (Fair Lawn, New Jersey). The 10 bp DNA ladder was obtained from Life Technologies (Frederick, MD). The internal standards, 40 bp and 80 bp fragments, and selected cycle-sequencing primers were prepared at the Nucleic Acid Facility (Iowa State University, Ames, IA). The 323 bp template was prepared using reagents in PCR Core System II from Promega (Madison, WI) and its Positive Control PCR Protocol. The PCR product was purified using QIAquick PCR Purification Kit from Qiagen (Valencia, CA). ThermoSequenase (32 U/μl), dNTPs (100 m*M*) and ddNTPs (10 m*M*) were obtained from USB/Amersham Life Sciences (Arlington Heights, IL).

2.2. Sequencing reaction

In order to generate enough quantities of sequencing fragments for UV detection, we used the cycle-sequencing protocol introduced by Cohen et al. [15]. The reaction mixture was combined in a microcentrifuge tube and put on ice: 200 pmol primer, 0.2 pmol template, 10 μl Tris pH 9 (250 m*M*), 10 μl $MgCl_2$ (50 m*M*), 10 μl dNTP mix (10 m*M*) and 32 U ThermoSequenase. Autoclaved and deionized water was added to obtain a total volume of 90 μl. A 20 μl amount of reaction mixture was added to each of the four 0.2 ml PCR reaction tubes (Molecular BioProducts, San Diego, CA) containing 1.25 μl of the appropriate ddNTP (1 m*M*). The samples were kept on ice before they were put onto the preheated block (95°C) of GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). The total cycle number is 200. Each cycle contains three consecutive steps: 95°C, 30 s; 52°C, 30 s; 72°C, 30 s. The product of cycle-sequencing reaction was purified by a spin column (Princeton Separation, Adelphia, NJ), and dried in vacuum. Before injection, the DNA samples were dissolved in 3 μl deionized water and transferred to a 96 well, 0.2 ml micro-tube plate (Marsh Biomedical Products, Rochester, NY), spiked with 10 pmol internal standards (40 bp and 80 bp DNA fragments). After heating the plate at 95°C for 3 min for denaturing, the sample plate was put onto ice for injection.

2.3. DNA separation

Sieving matrix was prepared by dissolving polyoxyethylene-6-cetyl-ether (Sigma) in the running buffer while gently heating and stirring [14]. Then the low viscosity gel was forced into a 24-capillary array from the ground end. Before injection, the matrix-filled capillary array was pre-run for 5 min at 32°C. Injection was performed at 2 kV for 2.5 min. During the run, the temperature was kept at 32°C. The running voltage of 8.8 kV was applied by a power supply from Glassman High Voltage (Whitehorse Station, NJ). After each run, the capillaries were regenerated by washing with 0.1 *M*

hydrochloric acid for a few minutes, then rinsed with deionized water for half an hour.

2.4. Instrumentation

The basic setup of the multiplexed capillary electrophoresis system is similar to that described in Ref. [8]. Thirty capillaries, 75 μm I.D. and 365 μm O.D., were packed side by side with 70 cm effective length and 90 cm total length. Of each capillary, 50 cm was enclosed in a water circulation system (Fisher Scientific, Pittsburgh, PA) to control the running temperature. A 254 nm mercury lamp was used for UV absorption detection. The transmitted light from the capillary array passed through an interference filter (Oriel) and a quartz lens (Nikon; focal length=105 mm; F=4.5). A linear photodiode array detector (PDA) (Hamamatsu model S5964, Hamamatsu, Japan) was used to collect data, and a National Instrument PCI E Series multifunction 16 bit I/O board was employed to transfer data to a computer (233 MHz Pentium, Packard Bell). The raw data sets were converted into single-diode electropherograms by a Labview program. Data treatment and analysis were performed using Microsoft Excel 97 and GRAMS/32 5.05 (Galactic Industries).

3. Results and discussion

3.1. Principles of normalization

Even though in a capillary array system, all the capillaries are run under exactly the same conditions (voltage, temperature, injection time and buffer pH), the surface chemistry and geometry of the capillaries are different. Also, the gel matrix cannot be exactly the same after being pushed into the capillaries. These variations can cause substantial variations in the migration times of DNA fragments, which precludes calling bases by simply overlapping the four individual electropherograms. We already demonstrated that the use of two internal standards provides normalization of migration times in micellar electrokinetic chromatography (MEKC) and capillary zone electrophoresis (CZE) [16]. This normalization method should also be useful in unlabeled DNA sequencing. Without labeling dyes, there should be no mobility shift among the Sanger fragments. Also, since sufficient denaturant (7 M urea) has been included in the buffer, compressions in GC-rich regions are minimized. Non-uniform migration times among different capillaries are therefore only caused by the variations mentioned above. So, we used two DNA fragments of known lengths to adjust the migration times of each capillary for base calling. Another reason that makes the internal standardization method suitable for CGE is that the relationship between migration time and base number is linear over a narrow range [17]. In entangled polymer solutions, the best model that describes DNA movement is ''the biased reptation with fluctuations'' (BRF) model [18]. According to this model, for small molecules (below a critical size), the mobility of the DNA fragment μ is inversely proportional to its size, represented by base number (''reptation without orientation'') [19]:

$$\mu/\mu_0 \sim 1/N \quad (1)$$

where μ_0 is the mobility in free solution. From the definition of electrophoretic mobility:

$$\mu = v/E = x/tE \quad (2)$$

where v is the average velocity, E is the externally applied electric field strength and x is the distance travel in time t [17]. We can thus change Eq. (1) to:

$$t \sim N \quad (3)$$

Based on the migration times of the two internal standards in each capillary, linear equations of $t \sim N$ for the corresponding electropherograms can be determined. Then, we can adjust the migration times of the DNA fragments in every capillary using one capillary as the migration time standard. After the normalization process, we can call the sequence of the DNA template according to the order of the adjusted migration times. For example, there are two different equations for the electropherograms of ddATP and ddCTP termination reactions:

$$N_A = a_1 t_A + b_1 \quad (4)$$

and

$$N_C = a_2 t_C + b_2 \quad (5)$$

Using Eq. (4) as the standard equation, we can

manipulate the terms in Eq. (5) to give the identical value for N_C:

$$N_C = a_1((a_2 t_C)/a_1 - (b_1 - b_2)/a_1) + b_1 \quad (6)$$

So the migration time of the C fragments can be normalized by:

$$t'_C = (a_2/a_1)t_C - (b_1 - b_2)/a_1 \quad (7)$$

3.2. Normalization of dsDNA migration times

First, we tested the normalization principle with dsDNA using 10 bp DNA ladder. The ladder was injected into selected capillaries without dilution. The raw data are shown in Fig. 1. Even though the samples were the same and all the capillaries were operated under the same conditions at the same time, the migration times of the same length fragment in different capillaries are different. Here, 4 or 5 capillaries were bundled together at the ground end to facilitate filling with gel with a 100 μl glass syringe. #2, #5, and #6 indicate the group of capillaries each numbered capillary belongs to. We can see that the migration times in the capillaries are very different no matter whether the capillaries are from the same bundle or not. Using the three largest peaks (10 bp, 100 bp and 330 bp) as internal standards, we can align these electropherograms (Fig. 2). Fig. 2A shows the result of using 10 bp and 100 bp fragments as standards to align the other peaks. After normalization, all the peaks of the same size fragments falling in the range of 10 bp to 100 bp have the same migration times. But for the peaks larger than 100 bp, the 330 bp must be used together with the 100 bp fragment as standards to achieve the proper result (Fig. 2B). This confirms that the linear relationship only fits in a narrow range.

3.3. DNA sequencing

The next experiment was for ssDNA, the Sanger fragments derived from cycle-sequencing reaction. The four chain-termination reactions created four sets of DNA fragments, corresponding to the four bases in DNA, A, C, G, and T. In order to obtain enough signal for UV absorption detection, we dissolved the dried sample in deionized water to implement stacking in injection. A layer of silicone oil (Life Technologies, Rockville, Maryland) was put on top of the vials to avoid evaporation during

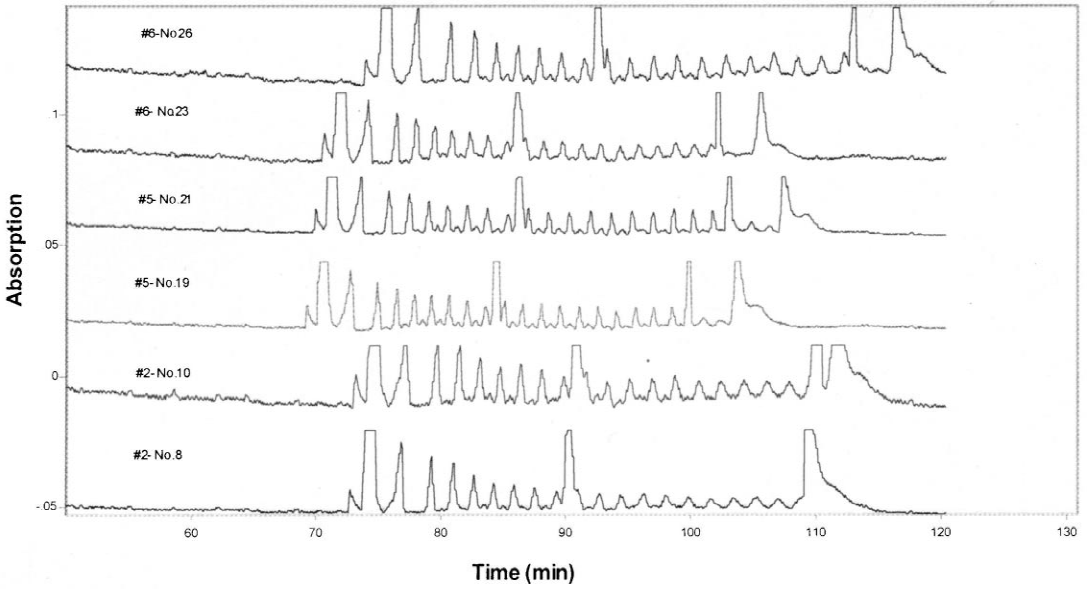

Fig. 1. Extracted UV electropherograms of 10 bp dsDNA ladder separation in the capillary array system. Capillaries No. 8, 10 are from bundle #2; No. 19, 21 are from bundle #5; No. 23, 26 are from bundle #6.

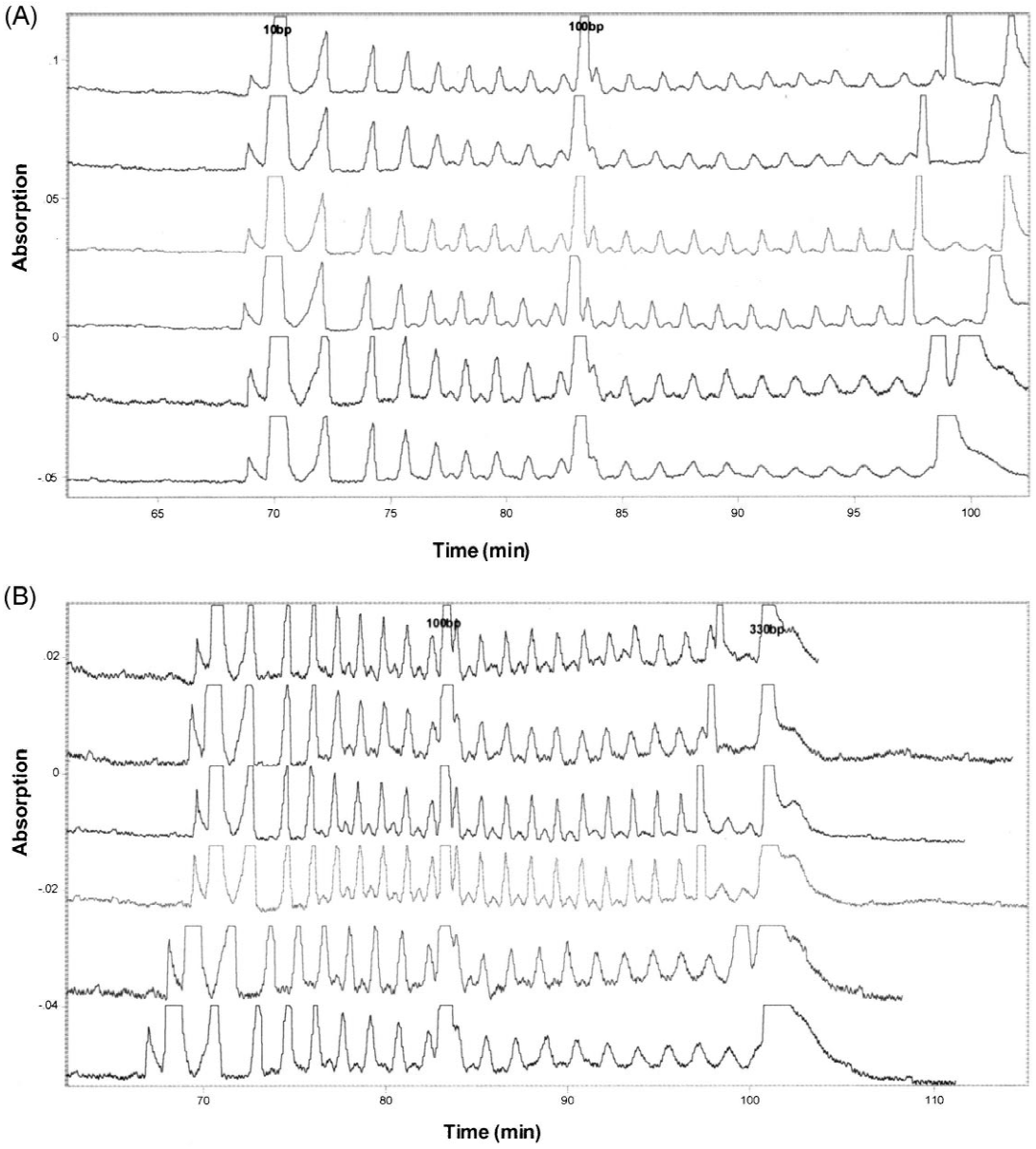

Fig. 2. The 10 bp ladder electropherograms of Fig. 1 after migration time adjustment. Peaks of same size fragments in different capillaries show up at the same adjusted times. (A) 10 bp and 100 bp fragments were used as internal standards in normalization. (B) 100 bp and 330 bp fragments were used as internal standards in normalization.

heating. It is known that the efficiency of CGE separation of oligonucleotides dissolved in water or other low ionic strength solvents is affected by the injection field strength and duration [20]. To achieve high resolution, a low injection field and a long injection time should be utilized [21]. We found that 2 kV injection voltage and 2.5 min injection time worked best in our experiment. The longer injection time did not degrade the separation performance because of stacking. This is confirmed by examining

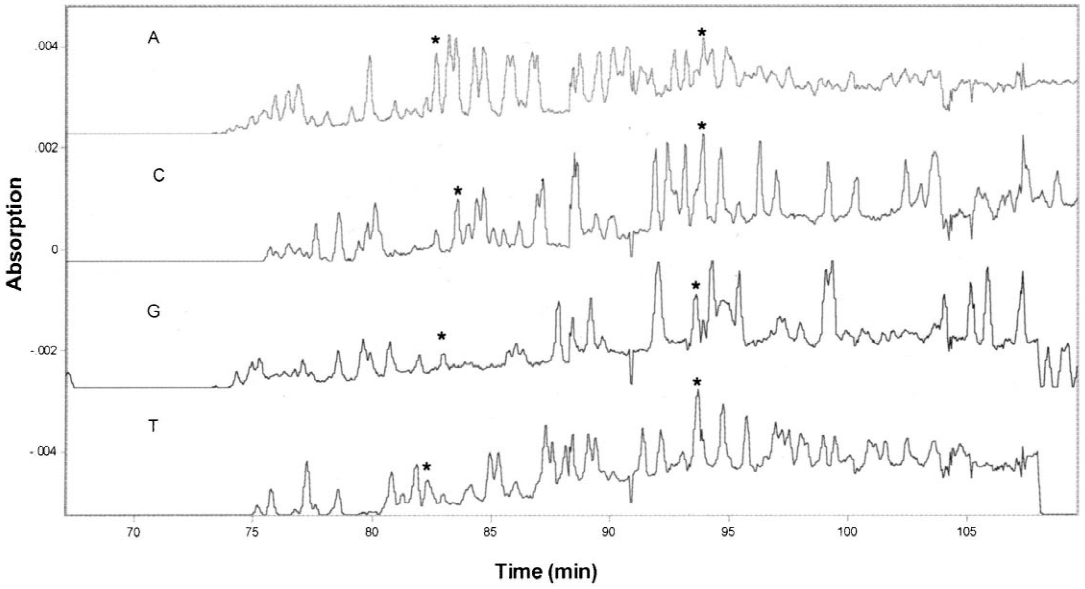

Fig. 3. Electropherograms of four individual sequencing reactions. The peaks with ‘‘*’’ on top are the internal standards at 40 bp and at 80 bp.

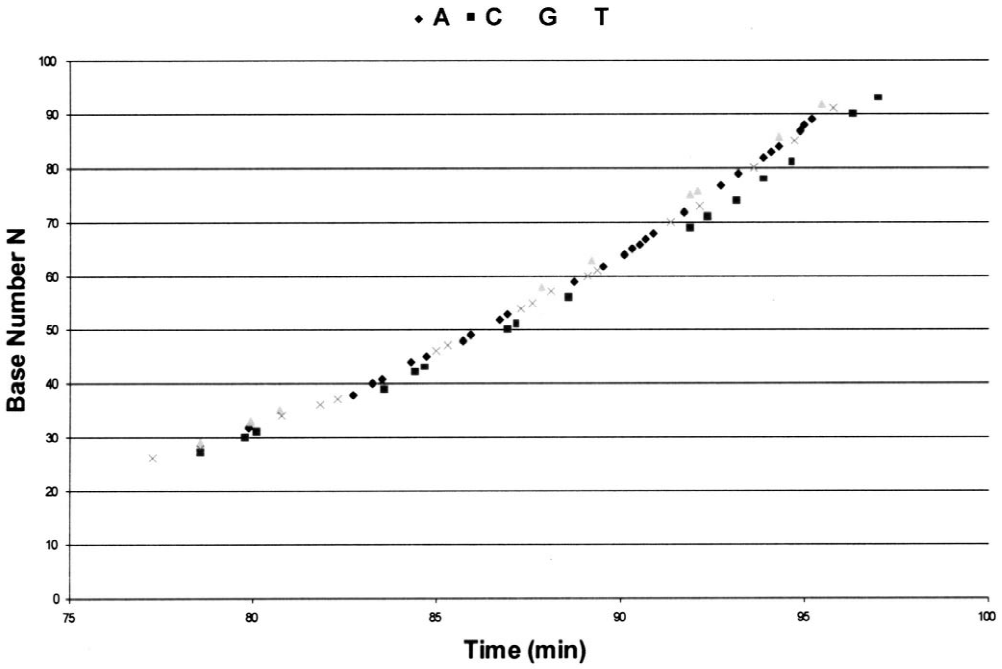

Fig. 4. Migration times plotted as a function of base number from the original electropherograms in Fig. 3.

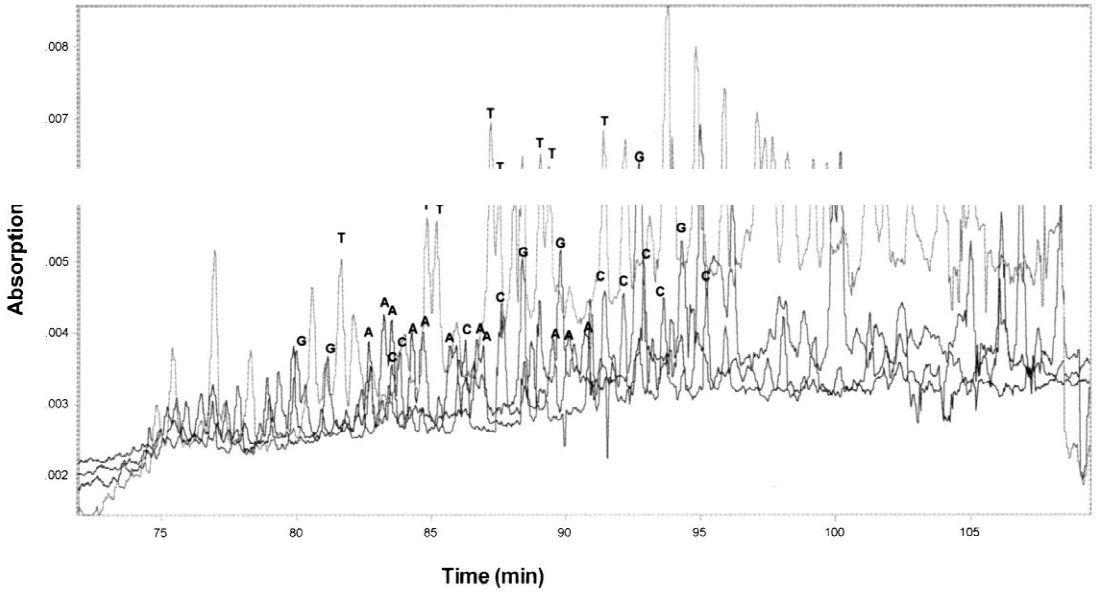

Fig. 5. Overlay of the four individual UV electropherograms after normalization of Fig. 3 using 40 bp and 80 bp fragments as internal standards.

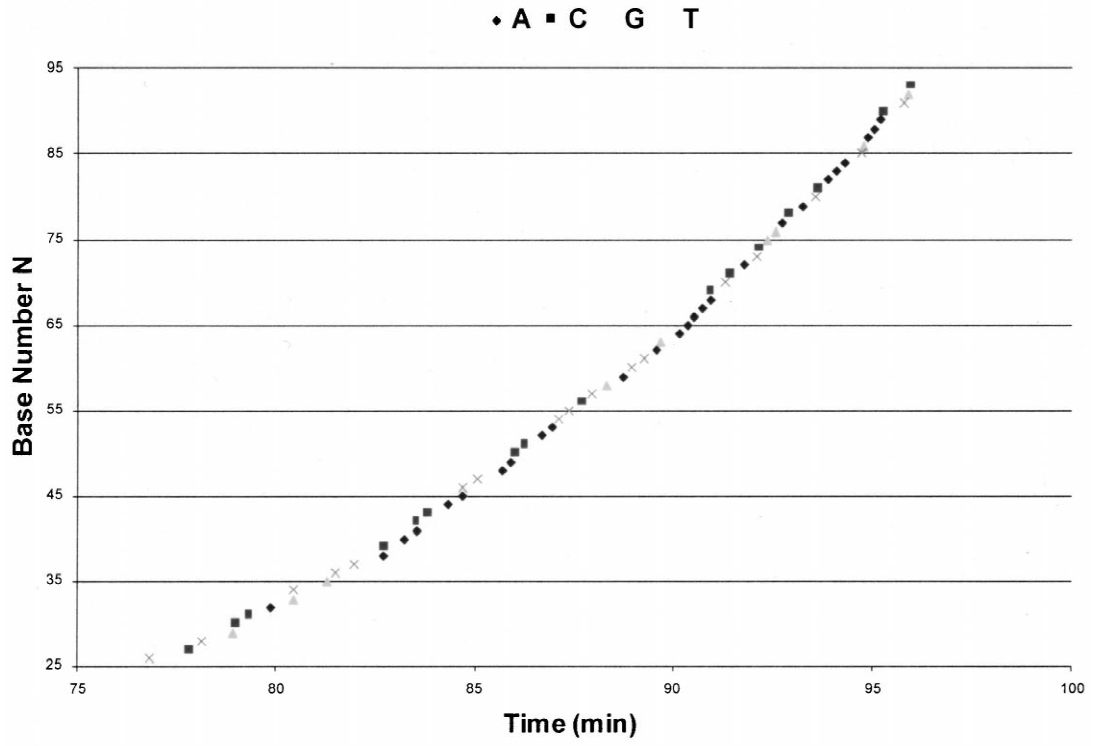

Fig. 6. Adjusted migration times plotted as a function of base number.

Table 1
Adjusted migration times (300 data points = 1 min), the corresponding base number, and the read sequence obtained from Fig. 6. There are no base-calling errors from 26 bp to 93 bp

Base No.	A(t)	C(t')	G(t')	T(t')	Sequence
26				23044.8	T
27		23343.18			C
28				23443.01	T
29			23669.29		G
30		23696.49			C
31		23796.85			C
32	23957.9				A
33			24130.79		G
34				24131.26	T
35			24376.79		G
36				24455.01	T
37				24595.22	T
38	24804.5				A
39		24816.53			C
40	24963.9				A
41	25055.5				A
42		25060.04			C
43		25145.64			C
44	25288.5				A
45	25403.5				A
46				25412.33	T
47				25523.66	T
48	25704.5				A
49	25773.5				A
50		25808.86			C
51		25879.7			C
52	26009.5				A
53	26078.5				A
54				26132.64	T
55				26222.16	T
56		26305.72			C
57				26388.89	T
58			26489.61		G
59	26626.5				A
60				26690.83	T
61				26780.96	T
62	26864.9				A
63			26898.78		G
64	27040.5				A
65	27100				A
66	27160				A
67	27211.5				A
68	27275				A
69		27279.75			C
70				27393.22	T
71		27426.74			C
72	27526.9				A
73				27628.37	T
74		27650.57			C
75			27699.77		G

Table 1. Continued

Base No.	A(t)	C(t')	G(t')	T(t')	Sequence
76			27765.24		G
77	27816.9				A
78		27865.05			C
79	27969.7				A
80				28082.4	T
81		28094.3			C
82	28169.4				A
83	28230				A
84	28295.9				A
85				28414.95	T
86			28423.98		G
87	28466.5				A
88	28502				A
89	28562.9				A
90		28579.84			C
91				28738.19	T
92			28756.18		G
93		28788.91			C

the resolution among the small DNA fragments, which would have been affected the most by electrokinetic injection.

Fig. 3 shows the four electropherograms of A, C, G, T for pGEM DNA in four capillaries of the array. The known base number N was plotted as a function of migration time t (Fig. 4). In an individual capillary, the base number is proportional to the migration time t. However, we cannot obtain the right order of bases from Fig. 4 because of migration variations among capillaries. Some fragments can be off by 5 bp. Two internal standards, 40 bp and 80 bp DNA fragments, were co-injected with the DNA samples. Based on the migration times and base

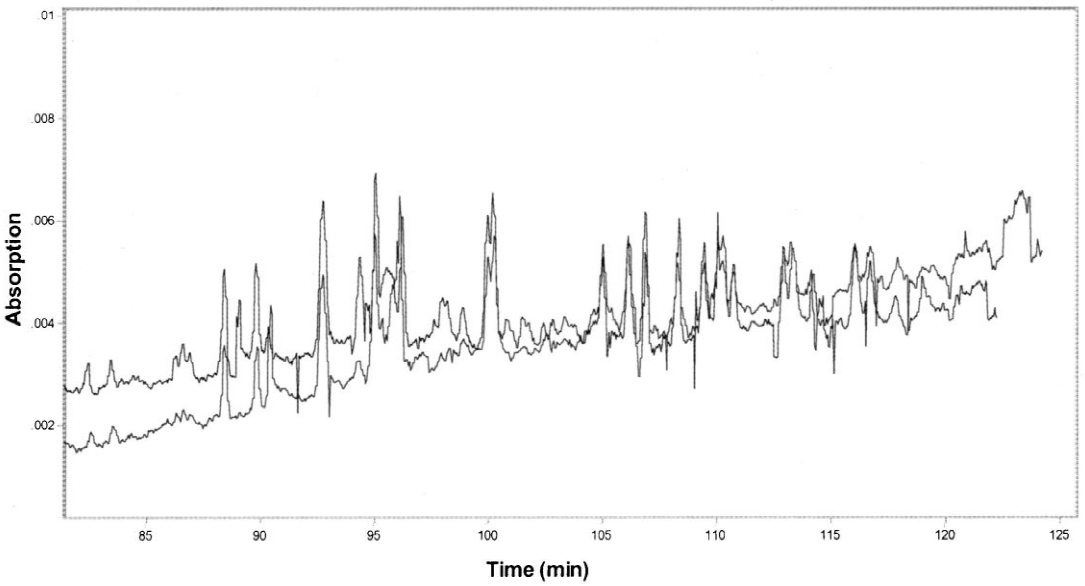

Fig. 7. Overlay of two G electropherograms. The peaks without matching peaks in the other electropherogram are ghost peaks.

number of these internal standards, all the $t \sim N$ equations of the capillaries were determined. After normalization of the migration times, the four electropherograms were aligned to call the sequence of the template (Fig. 5). Fig. 6 shows more clearly that after normalization, all the peaks are on the same line, and the sequence can be read directly from the data in these two figures (Table 1). In the worst case, there exists only a 0.5 bp error. The standard fragments added prevent base calling at those specific locations. However, staggered sizes can be used to span a large normalization range and to recover any missing information.

The protocol we used compensates for the amount of product, by using short template, high concentration of primer and high number of reaction cycles. So, the read length is short in this experiment. Especially for the A reaction, no peaks can be seen in the electropherogram after 89 bp. For the C, G, T reactions, even though the read length was up to 150 bp, we can only call the right sequence from 26 bp to 93 bp because of the limited useful range of two internal standards. Clearly, such short sequences will not allow the present scheme to compete with current instrumentation for genomic sequencing. However, for diagnosis [9] or antisense characterization [17], read lengths of 100 bp are adequate. The present scheme therefore offers an alternative to mass spectrometric analysis of short fragments [22]. The fact that capillary arrays are eventually scalable to 384 or even 1536 formats means that using 4 lanes at a time is not unreasonable. The unusually large number of amplification cycles implies a longer sample preparation time. However, that is always performed off-line in an automated system and has little effect on the throughput.

In addition to studying the four electropherograms for four different termination reactions, we also studied the electropherograms for a single termination reaction. Fig. 7 shows the result of overlaying two electropherograms of two G reactions after normalization. The G sample was run with the other four A, C, G, T samples (Fig. 6) in the same array. After migration time correction, the peaks for the same length fragments were aligned exactly. We thus confirmed that, after migration time correction, there is no mobility shift among the capillaries. We found some small ghost peaks which only showed up in one of the electropherograms, presumably due to the loss of fidelity of the cycle-sequencing reaction.

4. Concluding remarks

This experiment demonstrated that DNA sequencing could be performed in a multiplexed capillary electrophoresis system with UV absorption detection using internal standards. Because the products of four ddNTP reactions are run at four different capillaries, no additional information is needed to distinguish them in detection. Recently, a commercial version of the 96-capillary absorption instrument has become available [23]. It makes the detection scheme much simpler and more straightforward than autoradiography or LIF detection, and also lowers the operation cost. By using two internal standards, about 100 bp can be read. For longer reads, we expect that an additional one or two internal standards will suffice.

Acknowledgements

The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No.W-7405-Eng-82. This work was supported by the Director of Science, Office of Biological and Environmental Research, and by the National Institutes of Health.

References

[1] F. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. (USA) 74 (1977) 5463.
[2] K. Ueno, E.S. Yeung, Anal. Chem. 66 (1994) 1424.
[3] S. Takahashi, K. Murakami, T. Anazawa, H. Kambara, Anal. Chem. 66 (1994) 1021.
[4] R.A. Mathies, X.C. Huang, M.A. Quesada, Anal. Chem. 64 (1992) 2149.
[5] H. Kambara, S. Takahashi, Nature 361 (1993) 565.
[6] E.S. Yeung, Q. Li, in: M.A. Khaledi (Ed.), High Performance Capillary Electrophoresis, Vol. 146, John Wiley & Sons, 1998, Chapter 22.
[7] H. Tan, E.S. Yeung, Electrophoresis 18 (1997) 2893.
[8] X. Gong, E.S. Yeung, Anal. Chem. 71 (1999) 4989.
[9] X. Gong, E.S. Yeung, J. Chromatogr. B 741 (2000) 15.

[10] S.H. Kang, X. Gong, E.S. Yeung, Anal. Chem. 72 (2000) 3014.
[11] L. Ma, X. Gong, E.S. Yeung, Anal. Chem. 72 (2000) 3383.
[12] S. Carson, A.S. Cohen, A. Belenkii, M.C. Ruiz-Martinez, J. Berka, B.L. Karger, Anal. Chem. 65 (1993) 3219.
[13] Q. Gao, E.S. Yeung, Anal. Chem. 70 (1998) 1382.
[14] W. Wei, E.S. Yeung, Anal. Chem. 73 (2001) 1776.
[15] D. Froim, C.E. Hopkins, A. Belenky, A.S. Cohen, Nucl. Acids Res. 25 (1997) 4219.
[16] G. Xue, H.-M. Pang, E.S. Yeung, Anal. Chem. 71 (1999) 2642.
[17] A. Belenky, D.L. Smisek, A.S. Cohen, J. Chromatogr. A 700 (1995) 137.
[18] T. Duke, J.L. Viovy, A.N. Sememov, Biopolymers 34 (1994) 239.
[19] C. Heller, Electrophoresis 20 (1999) 1962.
[20] D. Demorest, R.J. Dubrow, J. Chromatogr. 559 (1991) 43.
[21] O. Salas-Solano, M.C. Ruiz-Martinez, E. Carrilho, L. Kotler, B.L. Karger, Anal. Chem. 70 (1998) 1528.
[22] Z. Fei, T. Ono, L.M. Smith, Nucl. Acids Res. 26 (1998) 2827.
[23] www.combisep.com