ELSEVIER

Journal of Chromatography A, 680 (1994) 497–501

JOURNAL OF CHROMATOGRAPHY A

High-speed and high-accuracy DNA sequencing by capillary gel electrophoresis in a simple, low cost instrument Two-color peak-height encoded sequencing at 40°C

Hui Lu, Edgar Arriaga, Da Yong Chen, Norman J. Dovichi*

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

Abstract

A low-cost DNA sequencer was constructed based on a single helium–neon laser. The two-color peak-height encoded sequencing protocol, based on the use of T7 DNA polymerase in a manganese buffer, was used to generate samples. Two termination reactions were performed. In the first, a TAMRA (applied Biosystems)-labeled primer was extended in the presence of ddATP and ddCTP. The amounts of dideoxynucleotides were adjusted to produce a 3:1 peak height ratio. Similarly, a ROX (Applied Biosystems)-labeled primer was extended in the presence of ddGTP and ddTTP; the amounts of dideoxynucleotides was adjusted to produce a 3:1 peak height ratio. The pooled fragments were separated on a 4% T LongRanger gel operated at 39°C. Over 500 bases of sequence were generated in 50 min.

1. Introduction

Large-scale DNA sequencing projects, including the Human Genome Initiative, have spurred the development of fast-sequencing technology. One promising approach relies on the use of capillary gel electrophoresis to separate sequencing fragments generated by the Sanger dideoxynucleotide chain terminating reaction [1-13]. The thin capillary produces efficient heat transport, allowing the use of high electric fields. The high field both increases the separation rate and also improves resolution compared with conventional sequencing approaches. For example, we have demonstrated separation of fragments 250 bases long in 7 min at an electric field of 800 V/cm [9]. We have also demonstrated that use of an electric field of 200 V/cm produces a two-fold improvement in resolution compared with conventional electrophoresis performed at an electric field of 80 V/cm [14].

Instrumentation is important in DNA sequencing applications. It is important to have simple, robust, and inexpensive technology for widespread application. Conventional dye-labeled sequencing is based on the use of four different fluorescent labels run in a single lane in Applied Biosystems (ABI)'s and DuPont's versions, or on the use of a single fluorescent dye run in four lanes in Pharmacia or Licor's approaches [15–18]. The use of a single dye is not practical in capillary electrophoresis; multiple dyes seem to be ideally suited for the capillary system. However, the use of four dyes is not always simple. The ABI technology requires the use of two lasers and four spectral channels for detection.

* Corresponding author.

0021-9673/94/$07.00 © 1994 Elsevier Science B.V. All rights reserved
SSDI 0021-9673(94)00444-E

The DuPont technology used a single laser to excite the four dyes; however, the close spectral emission bands of the four dyes severely hindered resolution of the dyes.

In 1992, this group reported the use of two fluorescent dyes in DNA sequencing [10]. This instrumentation can be simple, requiring a single laser. The sequence accuracy produced by the technology is superior to sequence obtained by commercial instruments [19]. The peak-height technique relies on the use of T7 DNA polymerase in a manganese containing buffer; with this enzyme system, chain extension produces very uniform incorporation of nucleotides [20]. As a result, very uniform peak height is observed for a single dideoxynucleotide termination. One dye-labeled primer was used in a ddATP and ddCTP termination reaction; the second dye-labeled primer was used in the ddGTP and ddTTP termination reaction. The amounts of ddNTP were adjusted to produce peaks that varied in height; the best resolution seemed to be produced when the peaks for either of the dyes were adjusted to be 3:1 in height. We reported systems based on the combination of FAM and TAMRA (trade name for dye-labeled primer prepared by Applied Bioystems, Toronto, Canada) dyes that were excited by a combination of an argon ion and green helium–neon laser or by a single argon ion laser. Similarly, FAM- and JOE-labeled fragments were excited by a single argon ion laser.

In this paper, we report a simple DNA sequencer based on a single, low-cost helium–neon laser. This system contains about US$ 4000 in optical components. Sequence accuracy was 100% for fragments up to 517 bases in length. This paper also presents the first use of elevated temperatures with cross-linked gels for DNA sequencing by capillary gel electrophoresis; 50 min are required to obtain the sequence. While non-cross-linked polyacrylamide has been used at elevated temperatures, it has been thought that bubble formation would prevent the use of the cross-linked material at elevated temperatures. Most workers have covalently bound the polyacrylamide to the walls of the capillary along the entire length of the capillary. Expansion of the gel matrix at elevated temperatures generated stress in the gel, damaging its physical properties. We find that the simple expedient of covalently binding only a short portion of the gel to the capillary wall at the detection end of the capillary is sufficient to hold the gel stably in the capillary; no gel damage has been observed at elevated temperatures.

2. Experimental

2.1. Instrument

Fig. 1 presents a schematic diagram of the instrument; the instrument is similar to the device used for analysis of DuPont labeled DNA sequencing samples and also one of the instruments used for two-color peak-height encoded sequencing of FAM- and JOE-labeled sequencing fragments [8,10].

In this instrument, a 1.2-mW green helium–

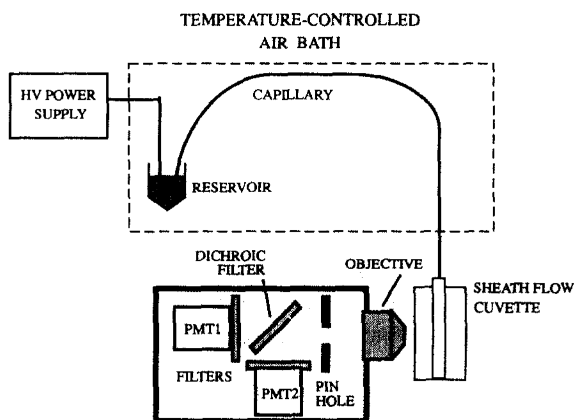

Fig. 1. Schematic diagram for the low-cost, high-speed DNA sequencer. The temperature-controlled air bath is denoted by the dashed square in the figure; the box has a safety interlock that disconnects the high-voltage power supply when the door is opened. Fluorescence is excited by a helium–neon laser that is focused at right angles to the cuvette; the excitation channel is not shown in this figure. Fluorescence is collected by the microscope objective, imaged onto a pinhole, split into two spectral channels with the dichroic filter, spectrally isolated with two bandpass filters, and detected with two photomultiplier tubes (PMT). The signal from the two PMTs is sent to a computer for digitization and processing.

neon laser beam ($\lambda = 543.5$ nm, Melles Griot, Nepean, Canada) is reflected from two front surface mirrors and focused with a $6.3\times$ objective (Melles Griot) into the locally constructed sheath flow cuvette. The mirrors are held in mirror mounts equipped with two-axes tilt adjustment. The microscope objective is held in a three-axes translation stage for alignment. The sheath-flow cuvette has 1-mm thick quartz windows and a 200-μm square flow chamber. The cuvette is similar to those used in flow cytometry. The cuvette is also held in a three-axes translation stage. The stainless-steel body of the cuvette is held at ground potential. The sheath stream is provided by siphon flow from a modified wash bottle. The sheath stream is the same as that used for the electrophoresis gel.

Fluorescence is collected at right angles with a $60\times$, 0.70 numerical aperture microscope objective (Universe Kogaku, Japan) and imaged onto a pinhole that is matched in size to the image of the illumined sample stream. The collection optic and pinhole are held rigidly in space with no adjustments. After passing through the pinhole, the fluorescence is split into two spectral channels with a dichroic beam splitter (Model 590DRLP|45° XF40, Omega Optical, Brattleboro, VT, USA). Reflected light is passed through a spectral bandpass filter (Model 580DF40, Omega) and transmitted light is passed through a second spectral bandpass filter (Model 630DF30, Omega). Each spectral channel is equipped with a R1477 photomultiplier tube (Hamamatsu, Middlesex, NJ, USA). The photomultiplier tube is operated at typically -1100 V. The output of the photomultiplier tubes is conditioned with an RC (resistor–capacitor) circuit and recorded with a Macintosh computer.

The capillary was placed within a temperature-controlled box; a companion manuscript describes the temperature controller and the behavior of DNA as a function of temperature in sequencing gels operated at high field [21].

2.2. Sample preparation

Samples were prepared as described previously [10]. A (-21) universal primer was used for the sequencing reaction. A TAMRA-labeled primer was used with ddATP- and ddCTP-terminated sample; the amount of dideoxynucleotide was adjusted to produce a nominal 3:1 peak height ratio, $A > C$. A ROX (trade name for dye-labeled primer prepared by Applied Biosystems)-labeled primer was used with ddGTP- and ddTTP-terminated sample; the amount of dideoxynucleotide was adjusted to produce a nominal 3:1 peak-height ratio, $G > T$.

The sequencing gel was 4% T[1] LongRanger in a 35 cm $\times$ 20 μm I.D. $\times$ 150 μm O.D. capillary, prepared in 7 *M* urea and in $1\times$ TBE buffer (0.54 g Tris, 0.275 g boric acid and 0.100 mmol disodium EDTA, diluted to 50 ml with deionized water). The capillary was operated at 39°C. Injection was for 35 s at an electric field of 200 V/cm.

2.3. Data processing

The transmitted spectral channel detects fluorescence from both TAMRA and ROX. The reflected spectral channel detects fluorescence from only TAMRA. To simplify data interpretation, the signals in the two spectral channels were normalized to the same height. The reflected signal was then subtracted from the transmitted channel; the result is a spectral channel that only contains signal from ROX.

3. Results and discussion

Fig. 2 presents the sequencing electropherogram produced by this instrument. The solid curve is the fluorescence signal generated by subtracting the reflected channel signal from the transmitted channel signal; this channel detects fluorescence from ROX-labeled fragments. Large peaks in this channel correspond to G; small peaks correspond to T. The dashed curve is the fluorescence signal generated by the re-

[1] T = (g acrylamide + g N,N′-methylenebisacrylamide)/100 ml solution.

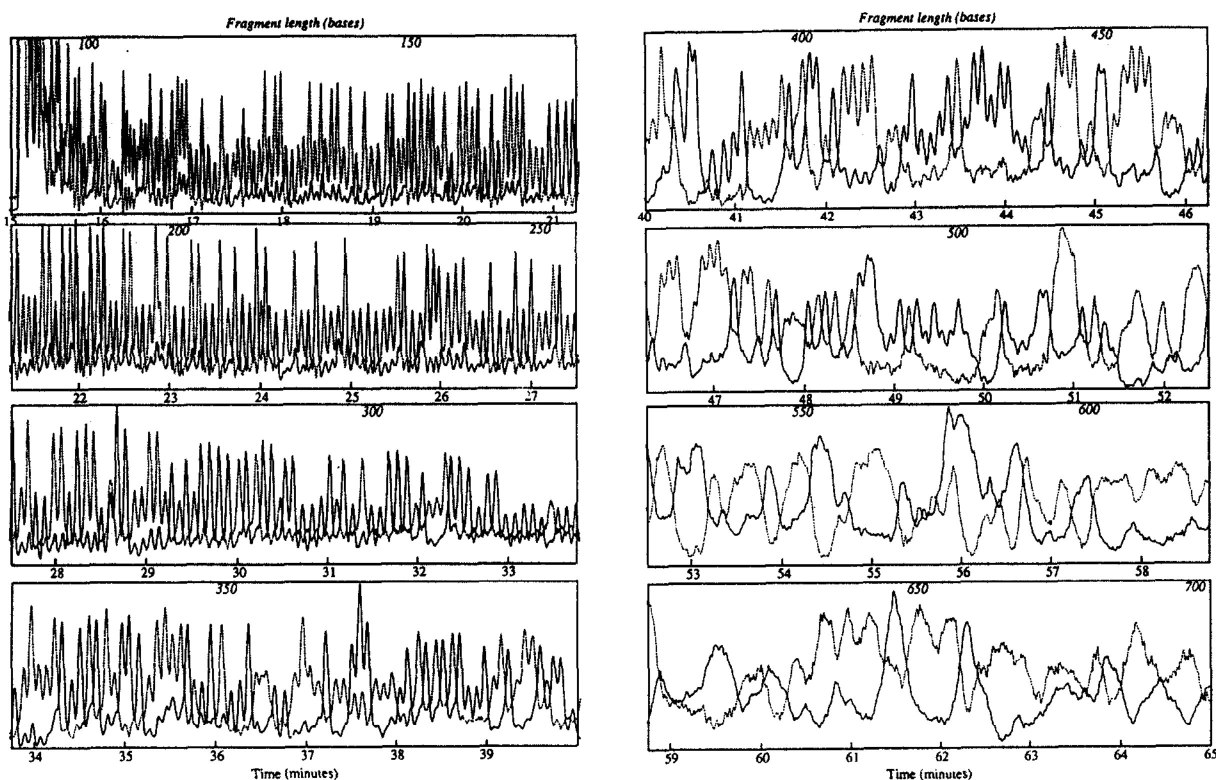

Fig. 2. High-speed DNA sequencing of M13mp18. The two-color peak-height encoded signal is denoted by the two curves. The solid curve corresponds to the fluorescence signal generated by TAMRA-labeled fragments and the dashed curve corresponds to the signal generated by the ROX-labeled fragments. The numbers at the top of each panel are the fragment lengths, in bases, while time is denoted below each panel.

flected channel; this channel detects fluorescence from TAMRA-labeled fragments. Large peaks in this channel are A; small peaks are C.

The sequence was called blind; later comparison with the consensus sequence revealed that the sequence was determined with no errors for fragments shorter than 517 bases. Only 51 min are required to generate this amount of sequence. This sequencing rate is an order of magnitude faster than that produced by commercial sequencers. Furthermore, this sequence read length and accuracy are superior to those produced by commercial sequencers. Comparison with the consensus sequence allowed identification of peaks for fragments up to 700 bases in length. We anticipate that deconvolution algorithms will allow generation of more sequence information per run.

These sequencing results are superior to all previous capillary gel electrophoresis systems. This is one of the longest read length produced by a capillary gel system. This is the highest-accuracy capillary system. Last, this is by far the fastest separation of long sequencing fragments by a capillary system.

This system is very inexpensive. We use a low cost helium–neon laser (ca. US$ 2500), and we estimate that the total cost of the optical components in the device is US$ 4000. This very-low-cost instrumentation, along with the excellent sequencing results, should make this system quite attractive to small-scale sequencing facilities. We have developed a multiple capillary version of the system; this multiple capillary instrument will be very useful for larger-scale sequencing applications.

Acknowledgements

This work was supported in part by the Department of Energy (DOE)–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 (NSERC) and the Department of Chemistry of the University of Alberta. N.J.D. acknowledges a Steacie fellowship from NSERC.

References

[1] H. Swerdlow and R. Gesteland, *Nucl. Acids Res.*, 18 (1990) 1415.
[2] H. Drossman, J.A. Luckey, A.J. Kostichka, J. D'Cunha and L.M. Smith, *Anal. Chem.*, 62 (1990) 900.
[3] A.S. Cohen, D.R. Najarian and B.L. Karger, *J. Chromatogr.*, 516 (1990) 49.
[4] H. Swerdlow, S. Wu, H. Harke and N.J. Dovichi, *J. Chromatogr.*, 516 (1990) 61.
[5] J.A. Luckey, H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris and L.M. Smith, *Nucl. Acid Res.*, 18 (1990) 4417.
[6] D.Y. Chen, H.P. Swerdlow, H.R. Harke, J.Z. Zhang and N.J. Dovichi, *J. Chromatogr.*, 559 (1991) 237.
[7] A.E. Karger, J.M. Harris and R.F. Gesteland, *Nucl. Acids Res.*, 19 (1991) 4955.
[8] H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S. Wu, N.J. Dovichi and C. Fuller, *Anal. Chem.*, 63 (1991) 2835.
[9] M.J. Rocheleau and N.J. Dovichi, *J. Microcolumn Sep.*, 4 (1992) 449.
[10] D.Y. Chen, H.R. Harke and N.J. Dovichi, *Nucl. Acids Res.*, 20 (1992) 4873.
[11] S.L. Pentoney, K.D. Konrad and W. Kaye, *Electrophoresis*, 13 (1992) 467.
[12] X.C. Huang, M.A Quesada and R.A. Mathies, *Anal. Chem.*, 64 (1992) 2149.
[13] M.C. Ruiz-Martinez, J. Berka A. Belenkii, F. Foret, A.W. Miller and B.L. Karger, *Anal. Chem.*, 65 (1993) 2851.
[14] N.J. Dovichi, in J.P. Landers (Editor), *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, FL, 1993, Ch. 14.
[15] 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 (1986) 674.
[16] 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 (1987) 336.
[17] W. Ansorge, B.S. Sproat, J. Stegemann and C. Schwager, *J. Biochem. Biophys. Methods*, 13 (1986) 315.
[18] L.R. Middendorf, J.C. Bruce, R.C. Bruce, R.D. Eckles, D.L. Grone, S.C. Roemer, G.D. Sloniker, D.L. Steffens, S.L. Sutter, J.A. Brumgaugh and G. Patonay, *Electrophoresis*, 13 (1992) 487.
[19] S. Tabor and C.C. Richardson, *J. Biol. Chem.*, 265 (1990) 8322.
[20] S. Bay, H. Starke, J. Elliott and N. Dovichi, *Proc. SPIE—Int. Soc. Opt. Eng.*, 1891 (1993) 8.
[21] H. Lu, E. Arriaga, D.Y. Chen, D. Figeys and N.J. Dovichi, *J. Chromatogr. A*, 680 (1994) 503.