High speed DNA sequencing by capillary electrophoresis

John A.Luckey, Howard Drossman, Anthony J.Kostichka, David A.Mead, Jonathan D'Cunha, Tracy B.Norris and Lloyd M.Smith*

Department of Chemistry, University of Wisconsin, Madison, WI 53706, USA

Received May 17, 1990; Revised and Accepted June 15, 1990

ABSTRACT

A major challenge of the Human Genome Initiative is the development of a rapid, accurate, and efficient DNA sequencing technology. A major limitation of current technology is the relatively long time required to perform the gel electrophoretic separations of DNA fragments produced in the sequencing reactions. We demonstrate here that it is possible to increase the speed of sequence analysis by over an order of magnitude by performing the electrophoresis and detection in ultra thin capillary gels. An instrument which utilizes these high speed separations to simultaneously analyze many samples will constitute a second generation automated DNA sequencer suitable for large-scale sequence analysis.

INTRODUCTION

Automated DNA sequencers using laser-based fluorescence detection in conjunction with conventional slab gel electrophoresis have theoretical capacities of about 10,000 bases of raw sequence data per day (1). Even if these instruments were operated continuously their throughput is insufficient to permit cost-effective sequence analysis on a genomic scale (1,2). This throughput is determined both by the number of samples that can be analyzed at one time, and by the speed of the separation process. The separation speed is limited in slab gel electrophoresis by the magnitude of the electric field which can be applied to the gel without excessive heat production. Recently it has been shown that if electrophoretic separations are performed in very small diameter capillaries (typically 50–100 μm i.d.), a much greater field may be applied due to the reduced Joule heating. Separations performed by this method, known as capillary electrophoresis (CE), are generally characterized by extremely high speeds and resolutions (3–6). This paper describes an instrument developed for automated DNA sequencing by capillary gel electrophoresis, in which separation speed is increased by up to 14 fold over conventional methodology. The ability to obtain such high speeds in DNA sequence analysis opens the possibility of developing a second generation of automated sequencing instruments an order of magnitude more powerful than those presently in existence.

EXPERIMENTAL PROCEDURES

Instrument Design

The capillary electrophoresis apparatus employed has been described previously (7). In order to perform fluorescence-based DNA sequence analysis in capillaries, a sensitive detector capable of simultaneous measurements of fluorescence at four wavelengths was constructed (Figure 1). Briefly, a ten milliwatt beam from an argon ion laser operated in multiline mode is focused onto the capillary in a 50 μm spot to fill the inner diameter of the tube. The emitted fluorescence is collected at right angles with a microscope objective, spatially filtered to remove scattered excitation light, and split into four parts of approximately equal intensity using beamsplitters. Each of the four parts is passed through a different 10 nm bandpass interference filter (Omega Optical Inc., Brattleboro, VT) to select a wavelength region of interest, and the fluorescence is measured with a photomultiplier tube. The signal from each photomultiplier tube is filtered with a one hertz low pass filter (third order, Bessel), digitized every 100 milliseconds on a 12 bit AD converter (Real-time Devices, State College, PA), and stored in an IBM AT compatible microcomputer. The data is processed after acquisition by use of a digital filtering algorithm and multicomponent analysis similar to that described previously (8,9).

Sequencing Procedure

The sequencing strategy employed is the multiple fluorophore approach which we have described previously (10,11). In this approach the oligonucleotide primer used in enzymatic sequence analysis is covalently attached to one of four different fluorophores. Each of the four reactions is performed using a different fluorophore-conjugated primer, and the reactions are combined and loaded together onto a denaturing polyacrylamide gel. The sequence is determined from the order in which the different fluorophore-labeled DNA fragments pass through a high-sensitivity fluorescence detector positioned near the bottom of the gel. Sequencing reaction mixtures (4.5 μl) contained 70 fmoles of single stranded M13mp19, 70 fmoles of dye primer, 5 mM MgCl_2 , 5 mM Tris-HCl (pH 8.5), 0.025 units of Bst DNA polymerase (BioRad), 35 μM of the three deoxynucleotides in excess, 2 μM of the limiting deoxynucleotide, and 6 μM ddGTP, 15 μM ddATP, 31 μM ddTTP, 5 μM ddCTP for each of the four respective reactions. Mixtures were incubated at 65°C

* To whom correspondence should be addressed

for 5 minutes, EDTA was added to 20 μ M, and the reactions were combined and precipitated by adding NH_4OAc to 4 M and ethanol to 50%. The pellet was resuspended in 5 μ l formamide and one μ l 50 μ M EDTA, and heated at 90°C for 3 minutes prior to injecting. The fluorophores employed were the standard set of four dyes (known as FAM, JOE, TAMRA, and ROX,

reference 12) available commercially from Applied Biosystems. The dye-coupled primers were synthesized by the general procedures described in reference 12 and Applied Biosystems technical literature, in which a six carbon spacer is utilized with the FAM and JOE dyes, and a two carbon spacer with the TAMRA and ROX dyes. Gel-filled capillaries were prepared as described (7).

RESULTS

The performance of this system was tested by sequence analysis of the bacteriophage vector M13mp19. Mixtures of the four DNA sequencing reactions labelled with different fluorophores were applied to a 50 μ m i.d. 375 μ m o.d. denaturing polyacrylamide gel-filled capillary by electrokinetic injection (7) and separated by electrophoresis. Representative data are presented in Figures 2A, and an expanded display of the region from base 306 to base 333 is shown in Figure 2B. The correlation of the peaks to the known sequence of this template (Figure 2C) is exact from base 80 to base 358. The discrepancies observed for fragments shorter than 80 bases are due to two compressions (at $\text{C}_{78}\text{G}_{79}$ and $\text{G}_{68}\text{C}_{69}\text{A}_{70}$) which are also observed in conventional sequence analysis of this template, and to the low resolving power of this gel composition for these short DNA fragments. Excluding the compressions, readable sequence is obtained from base 53 on, which is the 35th nucleotide of sequence added to the primer. The expanded plot in Figure 2B shows the excellent resolution which is obtained even for fragments over 300 bases in length differing in size by a single base.

Table I presents a summary of data obtained from multiple capillary and slab gel analyses performed under a variety of conditions. The retention time, resolution, and column efficiency obtained for three different length DNA fragments was measured.

Figure 1. Optical system employed for multiple wavelength fluorescence detection in capillary electrophoresis.

Figure 2. DNA sequence data obtained in the capillary electrophoretic analysis of M13mp19. The conditions employed are those given in the first column of Table I. Red (top trace), T reaction; orange (second trace), G reaction; green (third trace), A reaction; blue (fourth trace), C reaction. A) 80 minutes of data are shown, with peaks evident out to about base 360 in the sequence. The three T peaks identified in Table I are indicated with arrows. B) Expanded display of the boxed region in A, with the baselines for the four traces superimposed. C) The sequence for M13mp19 corresponding to the data shown. The underlined region at the beginning (bases 1–18) corresponds to the primer sequence employed. Thus the first incorporated base is the G at position 19. The Ts identified in Table I are also underlined.

In each case, the measurements were made for the center T of a group of three T bases (see Figure 2).

DISCUSSION

Detection sensitivity was a key factor in designing a DNA sequencing instrument based on capillary electrophoresis. This is due to the limited amount of sample that may be applied to the gel without adversely affecting resolution. For example, on the sequencing instrument manufactured by Applied Biosystems (ABI 370A), an optimal sample size is about 0.4 picomole (13) of template, giving about a femtomole (13) of DNA per band (11). Extrapolating from such a standard slab gel to the gel-filled capillaries suggested that only attomoles (13) of DNA would be present per band. In order for the multiple fluorophore strategy to be successful in capillaries, a system capable of detecting these low levels of each of the four fluorophores with adequate sensitivity was necessary.

The system described in Materials and Methods met this criterion. The basic optical design is a simple and widely used configuration for single wavelength laser-based fluorescence

detection in which the use of a spatial filter to eliminate background laser radiation reflected from the capillary walls is critical (7,14). The design was modified to permit simultaneous detection at multiple wavelengths by using beamsplitters to divide the emitted fluorescence into parts. This simple approach to multispectral detection employs no moving parts and has excellent sensitivity (approximately 0.1 attomole or 60,000 molecules of fluorescein at a signal to noise of 2:1).

The data presented in Figure 2 and Table 1 are representative of the performance of this system for DNA sequencing by capillary electrophoresis. The most significant effect is clearly the high speed of the CE-based sequence analysis. In the most rapid analysis presented (70/40 cm capillary at 428 V/cm), speed was increased more than 14 fold relative to slab gel electrophoresis (31 minutes versus 452 minutes for DNA fragments 317 bases in length). This high speed was obtained at no cost in separation efficiency, as measured either by resolution or by plates obtained per meter (Table I).

Several factors can affect these separation speeds, including gel composition, column length, and the applied field. It is well known that the mobility of DNA fragments is higher in low

TABLE I. SUMMARY OF ELECTROPHORESIS RESULTS¹

Length ² (cm)		Capillary Electrophoresis ⁶			Slab ⁷
		100/70	70/40	70/40	41.5/26
Field		300 V/cm	300 V/cm	428 V/cm	30-40 V/cm
Base ⁵ 99	Retention time (minutes), t_R	39.2 ± 2.3	23.0 ± 2.9	17.7 ± 1.1	175.7 ± 3.5
	Resolution, R^3	2.47 ± 0.53	1.92 ± 0.24	1.99 ± 0.10	1.13 ± 0.26
	Theoretical plates (per meter), N^4	9.1 ± 5.3 × 10 ⁶	7.2 ± 3.0 × 10 ⁶	13.2 ± 1.5 × 10 ⁶	1.7 ± 1.0 × 10 ⁶
Base 227	t_R	59.4 ± 1.8	33.8 ± 2.9	25.7 ± 1.7	336.2 ± 5.2
	R	1.93 ± 0.52	1.34 ± 0.45	1.23 ± 0.30	1.02 ± 0.26
	N	8.9 ± 4.6 × 10 ⁶	7.8 ± 2.0 × 10 ⁶	7.2 ± 3.5 × 10 ⁶	3.9 ± 1.9 × 10 ⁶
Base 317	t_R	73.8 ± 2.0	41.4 ± 2.7	31.3 ± 3.4	452.2 ± 8.7
	R	1.27 ± 0.22	0.94 ± 0.33	0.70 ± 0.16	0.73 ± 0.11
	N	6.7 ± 1.4 × 10 ⁶	7.0 ± 2.5 × 10 ⁶	3.8 ± 1.0 × 10 ⁶	3.8 ± 0.9 × 10 ⁶

¹ Data are reported as the mean and standard deviation of three separate runs under identical conditions (employing different gel columns) for the CE results, and of five lanes on two gels for the slab results.

² x/y, where x is the overall column length and y is the distance from the beginning of the column to the detector, i.e. the effective column length.

³ Resolution R computed according to the formula $R = 1.18 \frac{[(t_R)_B - (t_R)_A]}{[(W_{1/2})_A + (W_{1/2})_B]}$.

⁴ Theoretical plates N computed according to $N = 5.54 (t_R/W_{1/2})^2$.

⁵ Bases 99, 227, and 317 are each the middle T of three consecutive T_s and are underlined in Figure 2B.

⁶ Performed on a 3.2% total acrylamide/2.7 % bis acrylamide crosslinker gel containing 7 M urea (7).

⁷ Performed on 0.4 mm thick, 6 % total acrylamide/5 % bis acrylamide crosslinker gels containing 7 M urea (7) on an ABI DNA sequencer.

percentage gels (15). The gel matrix employed in the capillary columns was 3.2% total acrylamide with 2.7% bisacrylamide crosslinker (3.2%T/2.7%C), a composition found empirically to provide a good balance between speed and resolution. In the slab gel experiments a 6%T/5%C gel is employed, in accord with manufacturers recommendations. The lower gel percentage that may be employed in CE with good resolution is one factor in the observed high separation speeds.

Shorter columns also have an obvious benefit for more rapid separations. In order to use a short column successfully, however, its use must not substantially degrade resolution. This issue was examined cursorily by comparing the results obtained on 40 and 70 cm gel columns (Table I). Although no significant effect of length upon resolution is evident in these results (the only CE resolutions which differ with a confidence of greater than 95% are those in columns 1 and 3 of row 3) the run to run variability present in these experiments could easily mask important trends in the data. For example, if resolution depends upon the square root of length as occurs in standard models of chromatographic separations (16), the effect would not be apparent. Further work will be required to clarify this issue.

The third factor which has a major effect upon speed is the applied voltage. In all experiments performed to date we have seen linear increases in migration velocity with applied voltage, and no significant deleterious effect of the higher field upon resolution (compare the 40 cm column experiments employing

300 V/cm and 428 V/cm fields in Table I). In principle, one could therefore increase the field arbitrarily and continue to increase the separation speed. In practice, this approach is limited by the stability of the gel-filled capillaries. As fields are increased the lifetimes of the capillaries decrease, and they exhibit failure modes in which the capillary current drops sharply. Subsequent microscopic examination shows that bubbles have formed within the gel matrix. We have had reasonable success with fields of 300 V/cm, but often encounter problems when higher fields are applied. The development of more stable gel matrices will thus play an important role in the continued development of these methods.

Several new directions are suggested by these improved methods for sequence analysis. First, in order for the speed of these separations to substantially increase the throughput of automated sequencers it is essential to develop the capability for the simultaneous analysis of multiple samples. The most obvious approach to this problem is to employ multiple capillaries in parallel. However, inasmuch as the critical feature of the capillaries is their small dimensions, one could consider alternatively using ultra-thin slab gels operated under high-voltage conditions. Such gels have been described in the literature (17), and offer a potentially much simpler and less expensive approach to high speed separations. The design of an appropriate optical detection system for such an instrument will face stringent requirements for high spatial and temporal resolution as well as

extreme sensitivity. However, such a system, capable of simultaneously acquiring 500 bases of sequence data from 50 samples every hour, would have a theoretical capacity of 600,000 bases per day if operated continuously.

Secondly, the potential for yet higher resolution which may be attainable by CE could permit significantly more sequence to be obtained in a single analysis. Existing automated instruments are able to routinely obtain 300 to 400 bases of sequence information with reasonable accuracy (1). If it were possible to routinely obtain 1000 or more bases of sequence information in a single analysis, it would greatly facilitate the task of assembling sequence for a large region.

Finally, an area of considerable potential is the analysis of minute samples. In conventional sequence analysis, the template DNA to be sequenced is isolated from a small culture of bacteria containing the recombinant phage or plasmid to be analyzed. The labor and expense of these steps add considerably to the overall complexity of the sequencing process. However, the amount of material actually analyzed in capillary electrophoresis is much less than in conventional slab gel electrophoresis. We estimate that about 10^9 molecules of template DNA are loaded onto the capillary in a given sequence analysis (7). Since up to 10^{10} viral particles or plasmid molecules are present in a single bacteriophage M13 plaque or bacterial colony, respectively, this suggests that it may be possible to skip the entire step of template amplification, whether in vivo by bacterial growth in culture, or in vitro via the Polymerase Chain Reaction (18, 19), and instead to determine the sequence directly from DNA molecules isolated from individual plaques or colonies. To achieve this it will be necessary to develop new methods permitting most or all of a given sample to be applied to the capillary column, in contrast to current methods in which only a minute fraction of a sample may be analyzed.

ACKNOWLEDGEMENTS

Support for this work has been generously provided by Applied Biosystems, the Whitaker Foundation, and NIH (GM 42366).

REFERENCES

- Smith, L. M. (1989) *Amer. Biotech. Lab.* **7**(5), 10–25.
- Smith, L. M. and Hood, L. E. (1987) *Biotechnol.* **5**, 933–939.
- Jorgenson, J. W. and Lukacs, K. D. (1981) *Anal. Chem.* **53**, 1298–1302.
- Hjerten, S. (1983) *J. Chrom.* **270**, 1–6.
- Ewing, A. G., Wallingford, R. A., and Olefirowicz, T. M. (1989) *Anal. Chem.* **61**(4), 292A–303A.
- Guttman, A., Cohen, A. S., Heiger, D. N., and Karger, B. L. (1990) *Anal. Chem.* **62**, 137–141.
- Drossman, H., Luckey, J. A., Kostichka, A., D'Cunha, J., and Smith, L. M. (1990) *Anal. Chem.* **62**, 900–903.
- Connell, C., Fung, S., Heiner, C., Bridgham, J., Chakerian, V., Heron, E., Jones, B., Menchen, S., Mordan, W., Raff, M., Recknow, M., Smith, L., Springer, J., Woo, S., and Hunkapiller, M. (1987) *Biotechniques* **5**(4), 342–348.
- Smith, L. M., Kaiser, R. J., Sanders, J. Z., and Hood, L. E. (1987) *Meth. in Enz.* **155**, 260–301.
- Smith, L. M., Fung, S., Hunkapiller, M. W., Hunkapiller, T. J., and Hood, L. E. (1985) *Nucl. Acids Res.* **13**(7), 2399–2412.
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) *Nature* **321**, 674–679.
- Fung, S., Woo, S. L., Haugland, R. P., Menchen, S. M., and Connell, C. R. U.S. Patent #4,855,225 (August 8, 1989)
- picomole– 10^{-12} moles; femtomole– 10^{-15} moles; attomole– 10^{-18} moles.
- Cheng, Y. F. and Dovichi, N. J. (1988) *Science* **242**, 562–564.
- Kambara, H., Nishikawa, T., Katayama, Y., and Yamaguchi, T. (1988) *Biotechnol.* **6**, 816–821.
- Skoog, D. A., West, D. M., and Holler, F. J. (1988) in *Fundamentals of Analytical Chemistry*, Fifth Ed., p. 621. Saunders College Publishing.
- Radola, B. J. (1980) *Electrophoresis* **1**, 43–56.
- Wilson, R. K., Chen, C., and Hood, L. (1990) *Biotechniques* **8**(2), 184–189.
- Gyllenstein, U. B. (1989) *Biotechniques* **7**(7), 700–704.