Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions

ARTICLE *in* ANALYTICAL CHEMISTRY · NOVEMBER 1998

Impact Factor: 5.64 · DOI: 10.1021/ac980457f · Source: PubMed

Article

Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions

Oscar Salas-Solano, Emanuel Carrilho, Lev Kotler, Arthur W. Miller, Wolfgang Goetzinger, Zoran Sosic, and Barry L. Karger

Anal. Chem., **1998**, 70 (19), 3996-4003 • DOI: 10.1021/ac980457f

Anal. Chem. **1998,** *70,* 3996−4003

Articles

Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions

Oscar Salas-Solano, Emanuel Carrilho,† Lev Kotler, Arthur W. Miller, Wolfgang Goetzinger,‡ Zoran Sosic, and Barry L. Karger*

Barnett Institute and Department of Chemistry, Northeastern University, Boston, Massachusetts 02115

Long, accurate reads are an important factor for high-throughput de novo DNA sequencing. In previous work from this laboratory, a separation matrix of high-weight-average molecular mass (HMM) linear polyacrylamide (LPA) at a concentration of 2% (w/w) was used to separate 1000 bases of DNA sequence in 80 min with an accuracy close to 97% (Carrilho, E.; et al. *Anal. Chem.* 1996, *68,* 3305−3313). In the present work, significantly improved speed and sequencing accuracy have been achieved by further optimization of factors affecting electrophoretic separation and data processing. A replaceable matrix containing a mixture of 2.0% (w/w) HMM (9 MDa) and 0.5% (w/w) low-weight-average molecular mass (50 kDa) LPA was employed to enhance the separation of DNA sequencing fragments in CE. Experimental conditions, such as electric field strength and column temperature, as well as internal diameter of the capillary column, have been optimized for this mixed separation matrix. Under these conditions, in combination with energy-transfer (BigDye) dye-labeled primers for high signal-to-noise ratio and a newly developed expert system for base calling, the electrophoretic separation of 1000 DNA sequencing fragments of both standard (M13mp18) and cloned single-stranded templates from human chromosome 17 could be routinely achieved in less than 55 min, with a base-calling accuracy between 98 and 99%. Identical read length, accuracy, and migration time were achieved in more than 300 consecutive runs in a single column.

To achieve the goal of the Human Genome Project (HGP) to sequence de novo more than 3 billion bases of human DNA by the year 2005, cost-effective, high-throughput, and high-fidelity automated sequencing systems that improve on current slab gel electrophoresis technology are required.[1] The throughput issue is exacerbated by the fact that a significant redundancy level is necessary for the highest accuracy of the generated sequence. Therefore, one factor necessary to increase throughput of the process is to read the maximal possible number of bases in every run. Long read lengths also significantly decrease the number of templates needed to sequence DNA contigs at a given redundancy level, as well as reduce the problems of sequence assembly.[2] Beyond the HGP, other sequencing organisms will multiply the demand for rapid, long-read-length DNA sequencing.

Currently, substantial attention has been given to capillary electrophoresis (CE) using replaceable polymer solutions as an alternative to slab gel elctrophoresis.[3−11] This methodology offers high sensitivity, fast separation, simple sample introduction, and, because of the elimination of gel pouring, the possibility of full overall automation of the process from sample preparation to finished assembled DNA contigs.[6,9,12,13] Since labor is a major expense, full automation and multiplexing of the sequencing process should substantially reduce the cost per finished base.

Several groups are developing multicapillary instruments for DNA sequencing capable of running 96 or more samples simultaneously.[14−20] When fully commercialized, such instruments may significantly accelerate DNA sequencing. However, to utilize

* To whom reprint requests should be sent: (e-mail) bakarger@lynx.neu.edu.
† Current address: University of São Paulo, IQSC/DQFM 13560-970, São Carlos S.P., Brazil.
‡ Current address: Arqule Inc., Medford, MA 02155.

(1) Duboule, D. *Science* **1997,** *278,* 555.
(2) Craig, I. W. *J. Inherited Metab. Dis.* **1994,** *17*, 391−402.
(3) Carrilho, E.; Ruiz-Martinez, M. C.; Berka, J.; Smirnov, I.; Goetzinger, W.; Miller, A. W.; Brady, D.; Karger, B. L. *Anal. Chem.* **1996,** *68*, 3305−3313.
(4) Bashkin, J.; Marsh, M.; Barker, D.; Johnston, R. *Appl. Theor. Electrophor.* **1996,** *6*, 23−28.
(5) Quesada, M. *Curr. Opin. Biotechnol.* **1997,** *8*, 82−93.
(6) Swerdlow, H.; Jones, B. J.; Wittwer, C. T. *Anal. Chem.* **1997,** *69*, 848−855.
(7) Kim, Y.; Yeung, E. *J. Chromatogr., A* **1997,** *781*, 315−325.
(8) Chiari, M.; Riva, S.; Gelain, A.; Vitale, A.; Turati, E. *J. Chromatogr., A* **1997,** *781*, 347−355.
(9) Dovichi, N. *Electrophoresis* **1997,** *18*, 2393−2399.
(10) Sunada, W.; Blanch, H. *Electrophoresis* **1997,** *18*, 2243−2254.
(11) Madabhushi, R. S. *Electrophoresis* **1998,** *19*, 224−230.
(12) Ruiz-Martinez, M. C.; Salas-Solano, O.; Carrilho, E.; Kotler, L.; Karger, B. L. *Anal. Chem.* **1998,** *70*, 1516−1527.
(13) Tan, H. D.; Yeung, E. S. *Anal. Chem.* **1997,** *69*, 664−674.
(14) Ueno, K.; Yeung, E. S. *Anal. Chem.* **1994,** *66*, 1424−1431.
(15) Kheterpal, I.; Scherer, J. R.; Clark, S. M.; Radhakrishnan, A.; Ju, J. Y.; Ginther, C. L.; Sensabaugh, G. F.; Mathies, R. A. *Electrophoresis* **1996,** *17*, 1852−1859.
(16) Dovichi, N.; Zhang, J.-Z.; Yan, J.; Rong, J.; Rong, L.; Bay, S.; Roos, P.; Voss, K.; Dellinger, S., Santa Fe, NM, 1/28/96−2/1/96 1996; U.S. Department of Energy Workshop.
(17) Bashkin, J.; Roach, D.; Leong, J.; Bartosiewics, M.; Barker, D.; Johnston, R. *J. Capillary Electrophor.* **1996,** *3*, 61−68.
(18) Quesada, M. A.; Zhang, S. P. *Electrophoresis* **1996,** *17*, 1841−1851.

S0003-2700(98)00457-0 CCC: $15.00
Published on Web 09/01/1998

the full potential of multicapillary instruments, all parameters of the sequencing process, including sample preparation and separation of DNA fragments, must be optimized.[6,12,13,21] Furthermore, given 96 simultaneous runs, rugged operation is a major challenge; however, it is essential, as column or sample failures would severely limit throughput.

In 1993, our laboratory introduced replaceable polymer solutions (linear polyacrylamide, LPA) for DNA sequencing.[22] Such solutions could be conveniently pumped out of the capillary after each run and replaced with a fresh portion of the same matrix. Other polymers have also been tested for this purpose, e.g., poly(ethylene oxide) (PEO),[23] hydroxyethyl cellulose (HEC),[24] poly(dimethylacrylamide),[11] and poly(vinylpyrrolidone).[25] However, to-date, high-weight-average molecular mass (HMM) LPA is the polymer of choice for the longest read lengths.[3,26] It is a very hydrophilic polymer, combining additionally such features as high fidelity, reproducibility, and ease of preparation and handling.

In recent work, a 2% (w/w) HMM LPA was used to separate 1000 bases of DNA sequence in 80 min with an accuracy close to 97%.[3] More recent developments have included an inverse emulsion polymerization procedure for HMM LPA preparation in powder form[26] and a method for sample purification for both template removal and desalting to concentration levels below 10 μM.[12] Sample purification is an essential feature to achieve rugged operation in DNA sequencing.

In this paper, significant improvements in 1000-base sequencing have been achieved in terms of accuracy, analysis time, and routine operation. These improvements included optimization of the LPA composition and electrophoretic conditions, the use of energy-transfer dyes (BigDye labeled primers[27]) and a base-calling expert system. A mixed matrix containing 0.5% (w/w) low-weight-average molecular mass (50 kDa, LMM) LPA and 2% (w/w) HMM LPA was found optimal. Mixed polymer matrixes have also been used by other groups for DNA sequencing by CE.[7,28,29]

Under the optimized conditions, the separation of 1000 bases has been achieved in less than 55 min at an accuracy greater than 99% with M13mp18 as template. Similar results under optimized conditions have been obtained for the sequencing of templates from a fragment of human chromosome 17. Importantly, it has also been found that separation characteristics of a coated capillary column remained essentially identical at the above high-performance level for over 300 cumulative runs. Taken together, the above improvements substantially advance DNA sequencing by CE.

EXPERIMENTAL SECTION

Instrumentation. The design of a single-capillary instrument with detection based on laser-induced fluorescence (LIF) has been described previously.[30] Briefly, the fluorescence emission was collected with a microscope objective (model 13600, Oriel, Stamford, CT), and the spectra of the labeled sequencing fragments were acquired continuously in the range from 500 to 660 nm by a photodiode array (model 1461, EG&G Princeton Applied Research, Princeton, NJ). The CE columns were fused-silica capillaries of 50-, 75-, and 100-μm i.d, 365-μm o.d. (Polymicro Technologies, Inc., Phoenix, AZ), covalently coated with poly(vinyl alcohol) (PVA).[31] The specific lengths and electrophoretic conditions are listed in the figure captions.

Chemicals. Acrylamide, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED), ammonium persulfate, and urea were purchased from ICN Biomedicals, Inc. (Aurora, OH). TRIS, TAPS, and EDTA were from Sigma (St. Louis, MO). All chemicals were either electrophoresis or analytical grade, and no further purification was performed. Span 80 emulsifier and petroleum special with a boiling range from 180 to 220 °C were purchased from Fluka Chemicals (Buchs, Switzerland). Water used in all reactions and solutions was deionized to 18.2-MΩ grade with a Milli-Q purification system (Millipore, Worcester, MA).

High-weight-average molecular mass (9 MDa) LPA powder was produced in-house, as previously described.[26] Low-weight-average molecular mass LPA (50 kDa) was prepared by polymerization of acrylamide dissolved in methanol at 60 °C for 4 h using 2,2′-azobisisobutyronitrile (Aldrich, Milwaukee, WI) as initiator. After the polymerization was complete, the LPA powder was washed with acetone and vacuum-dried.

For this work, mixed LPA solutions containing 2.0% (w/w) 9-MDa LPA and 50-kDa LPA in 0.5 (w/w), 1.0 (w/w), or 2.0% (w/w) concentration were utilized to separate the DNA sequencing reaction products. To prepare a typical mixed LPA solution (20 g), appropriate amounts of dry 9-MDa LPA and 50-kDa LPA, urea (denaturant, final concentration 7 M), 10× buffer concentrate (500 mM Tris/500 mM TAPS/20 mM EDTA), and water were added in a glass jar and slowly stirred with a magnetic bar. The solutions were typically homogenized after 2 days and ready for use. The polymer solution was replaced from the capillary after each run. The electrophoresis buffer was changed every day on both the injection and detection sides of the capillary column. For a freshly prepared polymer, no preelectrophoresis was necessary; however, the voltage was applied for 5 min before injection, if the polymer solution was more than 1 month old. LPA powder is stable and has almost unlimited shelf life. The prepared working solutions of LPA could be stored in the refrigerator at 4 °C for up to 3 months.

DNA sequencing reactions were conducted using standard cycle sequencing chemistry with AmpliTaq-FS and both BigDye and conventional FAM/JOE/TAMRA/ROX labeled (−21) M13 universal primers (Applied Biosystems/Perkin-Elmer Corp., Foster City, CA) on an M13mp18 single-stranded template (New

(19) Anazawa, T.; Takahashi, S.; Kambara, H. *Anal. Chem.* **1996,** *68*, 2699−2704.
(20) Carrilho, E.; Miller, A.; Ruiz-Martinez, M. C.; Kotler, L.; Kesilman, J.; Karger, B. L. *SPIE Proc. Ultrasensitive Biochem. Diag. II* **1997,** *2985A*, 4−18.
(21) Salas-Solano, O.; Ruiz-Martinez, M. C.; Carrilho, E.; Kotler, L.; Karger, B. L. *Anal. Chem.* **1998,** *70*, 1528−1535.
(22) Ruiz-Martinez, M. C.; Berka, J.; Belenkii, A.; Foret, F.; Miller, A. W.; Karger, B. L. *Anal. Chem.* **1993,** *65*, 2851−2858.
(23) Fung, E. N.; Yeung, E. S. *Anal. Chem.* **1995,** *67*, 1913−1919.
(24) Bashkin, J.; Marsh, M.; Barker, D.; Johnston, R. *Appl. Theor. Electrophor.* **1996,** *4*, 39−41.
(25) Gao, Q.; Yeung, E. S. *Anal. Chem.* **1998,** *70*, 1382−1388.
(26) Goetzinger, W.; Kotler, L.; Carrilho, E.; Ruiz-Martinez, M. C.; Salas-Solano, O.; Karger, B. L. *Electrophoresis* **1998,** *19*, 242−248.
(27) Lee, L.; Spurgeon, S.; Heiner, C.; Benson, S.; Rosenblum, B.; Menchen, S.; Graham, R.; Constantinescu, K.; Upadhya, K.; Cassel, J. *Nucleic Acids Res.* **1997,** *25*, 2816−2822.
(28) Barron, A. E.; Sunada, W. M.; Blanch, H. W. *Electrophoresis* **1996,** *17*, 744−757.
(29) Wu, C. H.; Quesada, M. A.; Schneider, D. K.; Farinato, R.; Studier, F. W.; Chu, B. *Electrophoresis* **1996,** *17*, 1103−1109.
(30) Ruiz-Martinez, M. C.; Carrilho, E.; Berka, J.; Kieleczawa, J.; Miller, A. W.; Forest, F.; Carson, S.; Karger, B. L. *Biotechniques* **1996,** *20*, 1058.
(31) Goetzinger, W.; Karger, B. L. Int. Pat. Appl. WO 96/23220, 1996.

England Biolabs, Beverly, MA). For sequencing reactions with single-stranded templates from human chromosome 17, BigDye(−21) M13 universal primers (ABI/PE) were only used. The temperature-cycling protocol for this sequencing chemistry was performed on a PTC100 thermocycler (MJ Research, Inc., Watertown, MA), consisting of 15 cycles of 10 s at 95 °C, 5 s at 50 °C, and 1 min at 70 °C, followed by 15 cycles of 10 s at 95 °C and 1 min of 70 °C. After completion of a reaction, the samples were heated for 5 min at 100 °C in order to inactivate the enzymes prior to cleanup.

Purification of DNA Sequencing Reaction Products. Sequencing reaction products were purified using a method previously described.[12] Template DNA (M13mp18 or from chromosome 17) was removed using spin columns with a polyether sulfone ultrafiltration membrane, molecular weight cutoff of 300 000 (MWCO 300K, Pall Filtron, Northborough, MA) pretreated with 0.005% (w/w) solution of linear polyacrylamide with molecular mass 700−1000 kDa. The filtrate was dried under vacuum and dissolved in 50 μL of deionized water. The reconstituted template-free sequencing samples were then desalted using two prewashed Centri-Sep (gel filtration) spin columns (Princeton Separations, Adelphia, NJ) per sample. All results with conventional dye-labeled primers were achieved using this cleanup procedure.

An alternative, more rapid, manual desalting procedure was used for sequencing with BigDye-labeled primers, based on ultrafiltration in spin columns with polyether sulfone membranes with MWCO 30K (Pall Filtron). To prevent nonspecific binding of single-stranded DNA fragments, these columns were pretreated using 500 μL of 0.0001% (w/w) 700−1000-kDa LPA and spun at 3000 rpm for 15 min. A sample solution, after the template removal (120 μL of 30% (w/w) 1-propanol), was placed on the column, diluted to 500 μL with water and spun at 8000 rpm for 3 min. After spinning, the concentrated sample solution remaining in the column was again diluted to 500 μL and spun. Rinsing was performed with five column volumes of water, to lower the salt content below 10 μM (i.e., equivalent to the salt level using the gel filtration spin columns). The final volume of the concentrated DNA sequencing sample was ∼50 μL. A 5-μL aliquot of the purified sample was diluted with 20 μL of deionized water prior to injection. The specific injection conditions are described in the figure captions. The purified sequencing samples were stored at −20 °C in deionized water.

Software Design. An expert system recently developed in this laboratory was used for base calling. Data processing began by determining the primer dye spectra from the relatively intense peaks in the data, performing color separation by a least-squares fit to these spectra, and applying a median filter to remove high-frequency noise. The electropherogram was divided into sections containing ∼20 bases, and the fifth percentile value among the amplitudes of all data points in each section was computed. This calculation established the background at the center of each section; elsewhere, the background was derived by linear interpolation. After background subtraction, a search was conducted for long stretches of peaks with similar height, with the first peak in the stretch having highest median height being selected as the starting point for sequence determination. Peak width, separation, and dye mobility shifts were determined at this point. A set of empirical rules was then employed to find peak boundaries and to estimate the number of bases in each peak. Full details of this basecaller will be presented elsewhere.[32]

RESULTS AND DISCUSSION

The goal of the present work was to increase the speed, accuracy, and ruggedness of long reads in DNA sequencing by CE. In earlier work, our laboratory found that, using low-concentration 2% (w/w) HMM LPA polymer solution, the read length for DNA sequencing of M13mp18 by CE reached 1000 bases with close to 97% accuracy and with an analysis time of 80 min.[3] In that study, HMM LPA was prepared by solution polymerization at a monomer concentration of 6% (w/w), followed by dilution in the running buffer to 2% (w/w). Preparation of the separation matrix from solution-polymerized LPA was, however, slow, due to the high viscosity of the concentrated polymer solution. To overcome this problem, a new method for preparation of HMM LPA in powder form using inverse emulsion polymerization has recently been developed.[26] HMM LPA prepared by this method has a molecular mass close to 9 MDa, as determined by size exclusion chromatography. LPA powder has practically unlimited shelf life, and working solutions can be prepared fresh as needed by gentle stirring of the required amount of polymer and buffer components in water.

To increase ruggedness of the DNA sequencing process by CE, a sample purification procedure has been developed to remove the template and decrease the concentration of low-molecular-weight components of the reaction, e.g., chlorides and nucleotides, to less than 10 μM.[12] Template was removed using an ultrafiltration membrane, and desalting was performed by utilization of either an ultrafiltration membrane or a gel filtration column. Both versions of desalting maintain the same principle of reducing the salt and nucleotide content to or below 10 μM; however, ultrafiltration is more rapid for manual operation, because the drying procedure between template removal and desalting steps is eliminated. For automated purification, either gel filtration or ultrafiltration columns could be employed. As noted, in this work all studies on polymer composition, electric field, temperature, and internal diameter were conducted using the gel filtration approach, whereas sequencing with BigDye-labeled primers was achieved with the ultrafiltration method.

Optimization of Separation Conditions. *Composition of the Polymer Solution.* Previously, using a 2% (w/w) HMM LPA solution matrix and M13mp18 as template, several errors typically occurred in the early part of the electropherogram; then no further errors were found until base numbers higher than 700.[3] The base miscalls at the beginning of the electropherogram resulted in part from the relatively low separation efficiency of the HMM LPA matrix. For example, the efficiency of separation for the fragment 49 bases long was measured at 3 million theoretical plates/m (M N/m), whereas for fragment 750 bases it was 14 M N/m. Despite the fact that selectivity was the highest at the beginning of the electropherogram, it was not sufficiently high to separate effectively all short (i.e., below 100 bases long) DNA fragments. It is known that the selectivity of small DNA fragments strongly depends on the overall concentration of LPA in the separation matrix.[3,28] Therefore, it was decided to increase the LPA concen-

(32) Miller, A.; Karger, B. L., in preparation.

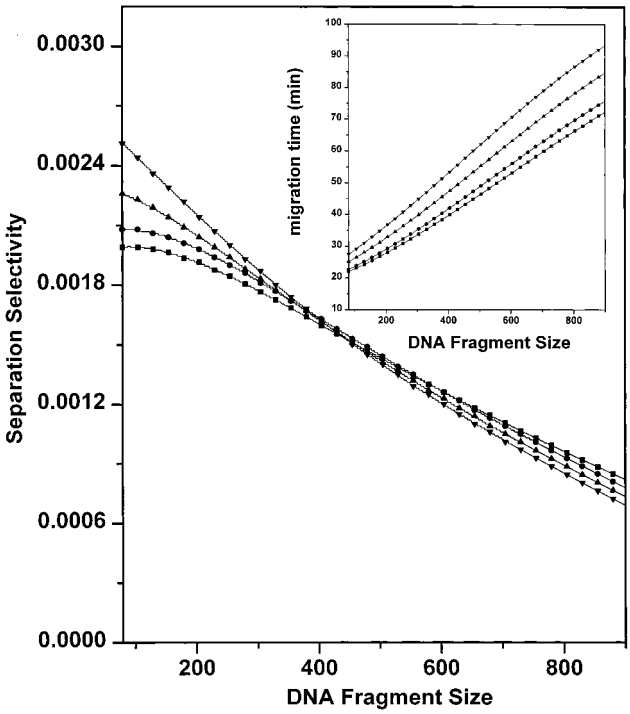

Figure 1. Separation selectivity ($\Delta\mu/\mu_{av}$) as a function of DNA fragment length for LPA solutions of different composition: 2.0% (w/w) 9 MDa (■); 2.0% (w/w) 9 MDa, and 0.5% (w/w) 50 kDa (●); 2.0% (w/w) 9 MDa and 1.0% (w/w) 50 kDa (▲); and 2.0% (w/w) 9 MDa and 2.0% (w/w) 50 kDa (▼). The plots of migration time vs base number for the different LPA matrixes are shown in the inset. Samples: dye-labeled primer cycle sequencing (AmpliTaq-FS) using (−21) primers on ssM13mp18. Electrophoretic conditions were as follows: effective length 30 cm, total length 45 cm, 75-μm-i.d., 365-μm-o.d. coated capillary (poly(vinyl alcohol)); both cathode and anode running buffers 50 mM Tris/50 mM TAPS/2 mM EDTA. Cathode running buffer also contained 7 M urea, the same as in the separation matrix. The samples were injected at a constant electric field of 25 V/ cm (0.7 μA) for 50 s and electrophoresed at 150 V/cm (6.7 μA) at 50 °C.

tration in the separation matrix by addition of low molecular mass LPA, because small DNA molecules interact preferentially with these polymer fibers.[7,33] In addition, low-weight-average molecular mass LPA present in the separation matrix decreases the average pore size. Moreover, the matrix viscosity increase over the 2% (w/w) HMM LPA solution was small.

On the basis of the above considerations, LMM LPA with a molecular mass of 50 kDa was prepared in powder form. To find the optimal composition of the mixed LPA solution, several LPA mixtures were made by combining a constant weight of HMM LPA powder (resulting in a concentration of 2% (w/w)) with 0.5, 1.0, and 2.0% (w/w) LMM LPA, respectively. Using these mixed matrixes, four-color sequencing reactions with dye-labeled primers were electrophoresed at 150 V/cm and 50 °C, conditions identical to those employed previously.[3] The separation selectivity of sequencing fragments was calculated as $\Delta\mu/\mu_{av}$, where $\Delta\mu$ is the difference in mobility of adjacent base number fragments and μ_{av} is the average of the two mobilities.[3]

Figure 1 presents a plot of selectivity as a function of fragment size for the different polymer compositions, and the inset presents migration time vs fragment size. As seen in this figure, upon increasing the concentration of LMM LPA in the separation matrix, the selectivity of the DNA fragments shorter than 400 bases was, as expected, enhanced. The shorter the DNA fragment, the greater the difference in selectivity. On the other hand, it is important to note that the selectivity of DNA fragments longer than 400 bases decreased with increase of the concentration of LMM LPA in the separation matrix. This latter behavior has also been observed by others.[28,34] Similar to selectivity, the efficiency of separation also improved at the beginning of the sequencing run upon increasing the LPA concentration in the matrix. In the late region of the electropherogram, however, efficiency was found not to be significantly dependent on the LPA concentration in the matrix (data not shown). Of course, as can be seen from the inset of Figure 1, all DNA fragments migrated proportionally slower with the increase of total concentration of the separation matrix.

For the mixed solutions containing 1 or 2% LMM LPA, the decrease in selectivity for the longer DNA fragments was significant, and even with an improvement of separation at the beginning of the electropherogram, these polymer mixtures resulted in substantially reduced read lengths. For example, at 2% LMM LPA in the separation matrix, the read length was decreased by almost 100 bases (data not shown). On the other hand, close examination of Figure 1 reveals that the mixed separation matrix containing 2% (w/w) HMM LPA and 0.5% (w/w) LMM LPA did not significantly decrease the selectivity of longer DNA fragments. At the same time, substantial improvement in selectivity at the beginning of electropherogram over 2% (w/w) HMM LPA was achieved. As a result, this mixed matrix was selected for use as a compromise to balance the requirements for separation of short and long DNA fragments. Since each matrix requires optimization of electrophoretic parameters of separation, these parameters were next studied.

Column Temperature and Electric Field Strength. We first investigated the selectivity and efficiency of the 2.5% (w/w) mixed matrix as a function of column temperature, in the range from 50 to 70 °C, in 5 °C increments. The selectivity for fragments shorter than 400 bases gradually decreased upon increasing the temperature to 70 °C (data not shown), probably due to fluctuations of the entangled polymer fibers at higher temperatures.[35,36] On the other hand, for DNA fragments longer than 400 bases, the selectivity increased from 50 to 60 °C (data not shown). These results were expected since the onset of biased reptation, which produces size-independent DNA migration, should be extended to larger base numbers as the temperature is raised.[37,38] However, from 60 to 70 °C, there was a gradual decrease in selectivity for the long fragments, likely again due to the loss in stability of LPA at the higher temperature.[36]

The efficiency of separation of DNA sequencing fragments from four-color dye primer reactions at various temperatures is

(33) Barron, A. E.; Blanch, H. W.; Soane, D. S. *Electrophoresis* **1994,** *15*, 597−615.
(34) Mitnik, L.; Salome, L.; Viovy, J. L.; Heller, C. *J. Chromatogr., A* **1995,** *710*, 309−321.
(35) Fang, Y.; Zhang, J. Z.; Hou, J. Y.; Lu, H.; Dovichi, N. J. *Electrophoresis* **1996,** *17*, 1436−1442.
(36) Kleparnik, K.; Foret, F.; Berka, J.; Goetzinger, W.; Miller, A. W.; Karger, B. L. *Electrophoresis* **1996,** *17*, 1860−1866.
(37) Slater, G.; Rousseau, J.; Noolandi, J.; Turmel, C.; Lalande, M. *Biopolymers* **1988,** *27*, 509−524.
(38) Duke, T.; Viovy, J. L. *Phys. Rev. E* **1994,** *49*, 2408−2416.

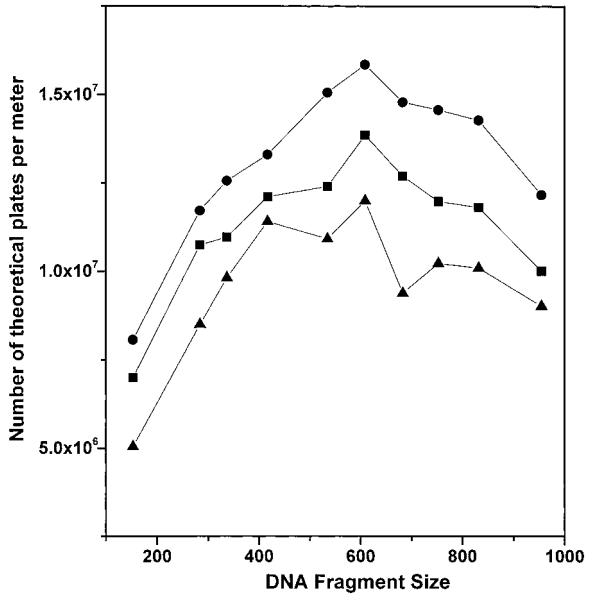

Figure 2. Number of theoretical plates per meter as a function of DNA fragment length for different temperatures of the capillary column: 50 (■), 60 (●), and 70 °C (▲).[5] The LPA was 2.0 (w/w) and 0.5% (w/w) 50 kDa. Sample and other electrophoretic conditions as in Figure 1.

shown in Figure 2. At any temperature, the efficiency of the bands was lowest at the beginning of the electropherogram, increasing with the DNA fragment size up to the maximum at 600 bases long. For example, at 60 °C, the maximum efficiency was ~15 M *N*/m. For fragments longer than 600 bases, the efficiency was gradually decreased. From Figure 2, efficiency for all DNA fragments was substantially higher at 60 °C than at 50 °C. This is may due to the fact that, at the higher temperature, the elasticity of entanglement−disentanglement interactions between DNA and LPA fibers increased, thus reducing the contribution of these interactions to band broadening.[39] However, at temperatures higher than 60 °C (at 65 °C and especially at 70 °C), a significant decrease in efficiency of the DNA fragments was observed. As noted above, at these temperatures, the positive effects of disruption of DNA elongation by thermal energy were likely outweighed by the loss in separation due to alteration of the stability of LPA. We concluded from these results that, for mixed 2.5% (w/w) LPA, the optimum temperature was 60 °C, 10 °C higher than previously used.[3] Note that the higher temperature will further help to overcome compressions as well as reduce analysis time.

Using 2.5% (w/w) mixed LPA matrix at the optimized temperature of 60 °C and electric field of 150 V/cm, the number of errors at the beginning of the electropherogram was reduced, compared to earlier work,[3] from 6 to 3. Moreover, for the latter part of the electropherogram, the number of errors did not increase. We then studied the effect of electric field strength from 150 to 300 V/cm on the separation of the DNA fragments. The results indicated that, with the given LPA matrix, the optimum resolution was achieved at 200 V/cm, where the gain in efficiency was substantial, as expected.[39−41] For example, for the fragment 750 bases long, the efficiency measured in five samples increased from 15 M *N*/m at 150 V/cm to 20 M *N*/m at 200 V/cm (RSD = 5%). At 250 V/cm, the resolution of adjacent fragments at the end of electropherogram decreased (data not shown), which may be the consequence of more pronounced stretching of the DNA fragments at high electric fields.[3,37,38] The overall resolution of DNA fragments with the mixed LPA matrix at 60 °C and 200 V/cm was better than with 2% HMM LPA alone. An additional advantage of separation of DNA fragments at higher field was again a reduction of analysis time.

Table 1. Effect of Capillary Internal Diameter on the Separation Efficiency of DNA Sequencing Fragments[a]

capillary column i.d. (μm)	current (μA)	efficiency[b] base 153	base 417	base 753
100	20.0	4	5	5
75	10.2	8	15	20
50	5.5	8	16	21

[a] Samples: dye-labeled primer cycle sequencing (AmpliTaq-FS) using ABI Prism BigDye (−21) primers on ssM13mp18. Electrophoretic conditions were as follows: capillary length 30 cm, total length 45 cm. Polymer solution was a mixture of 2.0% (w/w) HMM LPA and 0.5% (w/w) LMM LPA, both cathode and anode running buffers, 50 mM Tris/50 mM TAPS/2 mM EDTA. Cathode running buffer also contained 7 M urea, the same as in the separation matrix. The samples were injected at a constant electric field of 25 V/cm (0.7 μA) for 15 s and electrophoresed at 200 V/cm at 60 °C. [b] Efficiency was calculated in millions of theoretical plates per meter. Analysis was performed with three different capillary columns for each internal diameter. The resulting RSD for the number of theoretical plates was 6% (n = 3) under each experimental condition.

Internal Diameter of the Capillary Column. It is known that one of the factors limiting the separation speed in electrophoresis is heat production from high electric fields.[39,42] A main cause for the separation efficiency losses is the thermal gradient across the capillary;[39] therefore, columns with smaller internal diameter may be expected to separate the DNA fragments better than wider capillaries. However, the smaller the internal diameter of the capillary, the higher the pressure that must be applied to replace the separation matrix. With column temperature and electric field strength optimized for the mixed LPA matrix, the effect of the capillary internal diameter on the separation efficiency of DNA sequencing reaction products was next determined.

The efficiency of separation of DNA fragments for capillary columns of 50-, 75-, and 100-μm i.d. is shown in Table 1. The efficiency of DNA fragments in the 75 μm columns was significantly higher (~70%) than that in the columns with 100 μm. The number of theoretical plates increased by an average of ~70%. However, further reduction of the internal diameter from 75 to 50 μm did not appreciably change the separation efficiency. A probable reason for this result could be that, under the already optimized electric field strength and column temperature conditions, heat production in the 75-μm column during the run was balanced with the heat dissipation provided by these columns. Columns with 50 μm therefore did not provide additional advantages in the heat dissipation. Taking in account the heat balance of 75-μm columns under the optimized separation conditions and

(39) Gibson, T.; Sepaniak, M. *J. Chromatogr., B* **1997,** *695*, 103−111.
(40) Grossman, P. D. *J. Chromatogr., A* **1994,** *663*, 219−227.
(41) Karger, A. E. *Electrophoresis* **1996,** *17*, 144−151.
(42) Slater, G. W.; Mayer, P.; Grossman, P. D. *Electrophoresis* **1995,** *16*, 75−83.

the ease of the separation matrix replacement, these columns were selected for all experiments except those in Table 1.

DNA Sequencing under the Optimized Conditions. *Base-Calling Software.* In previous work,[3] base calling by a modification of a graph-theoretic method[43] was used, alone and in combination with a signal-processing procedure known as adaptive equalization. While work on equalization has continued,[44] it was found possible to attain even greater accuracy with manual editing where peak overlap was high, suggesting the use of an expert system. Such a system has been developed for the present work, capitalizing on features specific to dye primer chemistry; extension to dye terminators is in progress. Compared to other base callers,[43,45−48] the expert system relies heavily on the relatively uniform peak heights provided by the dye primers. The main advantage of this new base caller is for the case where peaks are poorly resolved, such that precise peak positions are difficult to determine. On data obtained previously with the 2% (w/w) HMM LPA matrix and the standard primers,[12] the expert system extended the read length at 99% accuracy by 50 bases. We have utilized this expert system in the sequencing reads throughout this study. A full description of the new software will be reported elsewhere.[32]

BigDye Sequencing Chemistry and CE. Recently, a new set of energy-transfer dye labels for DNA sequencing (BigDye) has been commercialized.[27] Under identical preparation and injection conditions, the signal-to-noise ratio was found to be 10-fold higher for BigDye-labeled reactions than for reactions made with conventionally labeled primers (data not shown). This result was a consequence of both the greater quantum efficiency of the BigDye primers and their smaller spectral overlap. To prevent saturating the detector with the higher signal, the electrokinetic injection time was decreased from 50 to 10 s at a field of 25 V/cm. The BigDye primers are also known to exhibit smaller changes in mobility shift early in the sequence than conventional dye primers,[27] which increased accuracy there.

DNA Sequencing under the Optimized Conditions. We combined all the parameters discussed above, along with the new base caller, to examine sequencing on a M13mp18 single stranded template. With the conventionally labeled primers, the read length was found to be only between 850 and 900 bases at 99% accuracy under the optimized separation conditions (data not shown). As seen in Table 2, the average read length (see footnote *a*) with BigDye primers was found to be 1010 bases at 99% accuracy, while at 98% accuracy, the read length was 1073 bases. The improvements led to sequencing 1000 bases in less than 1 h with higher accuracy than our previous work.[3] The RSD of migration time for the DNA fragment of the same length for the runs shown in Table 2 was roughly 4% for a single capillary column. From column to column ($n = 5$), the RSDs of averages for migration time for the same DNA fragment, read length per run, and accuracy of base calling were within 5%. Besides the standard M13mp18 DNA template, eight inserts cloned from humanchromosome 17 (MIT Center for Genome Research) on M13mp18 were also sequenced, and as Table 3 demonstrates, on average 940 bases with 99% accuracy per run were generated with BigDye-labeled primers. Figure 3 shows an example of typical separation of M13mp18 sequencing fragments using BigDye primers under the optimum conditions developed in this work. The only errors up to roughly 900 bases are in the very early part of the sequence and are due to severe compressions. The results in Tables 2 and 3 and Figure 3 demonstrate that, under optimized conditions, the DNA sequencing fragments separation by capillary electrophoresis is a rugged, fast, and high-throughput procedure.

Longevity of Coated Capillary Columns. For a rugged high-throughput multicapillary analytical system, the longevity of the columns is an important consideration, since frequent column replacement could create an issue of maintenance expenses. LPA polymer solutions, due to the significant hydrophilicity of the polymer, require the walls of the capillary column be coated to suppress EOF.[22] Less hydrophilic polymers such as PEO, PVP, or poly(dimethylacrylamide) have been reported to operate in uncoated capillaries,[11,23,25] but none of the polymer solutions used for DNA sequencing is so well suited for long reads in short analysis time as LPA. In addition, in the case of PEO, the bare

Table 2. Read Length[a] at the Optimized Conditions Using a Single Coated Capillary Column[b]

run no.	migration time base 1019 (min)	read length[c] for 99% accuracy	read length[c] for 98% accuracy
1	53.1	1001	1030
30	52.4	972	1015
60	54.2	974	1064
95	53.8	1103	1134
129	57.1	1025	1067
154	53.2	1015	1119
181	54.1	972	1070
267	56.4	989	1101
310	54.0	1042	1057
av[d]	54.3[e]	1010[e]	1073[e]

[a] In this study, read length was defined as the largest number of consecutive base calls starting with the fragment 49 bases in length. [b] For DNA standard reactions on M13mp18 template. Other experimental conditions as in Figure 3. [c] The reference sequence was the published M13mp18 sequence,[49] with three corrected discrepancies found at points of the sequence where the data consistently were at variance over hundreds of runs. Discrepancies found between published and real sequence in another region of M13mp18 (gene III) have been reported.[50] [d] A total of 325 runs were conducted on this column with no loss in performance. [e] The resulting RSDs for migration time and read lengths were 3, 4, and 4%, respectively.

Table 3. Sequencing Data of the Clone hCIT.91_J_4 Located on Human Chromosome 17

parameter	average[a]	reproducibility[a] (% RSD)
migration time base 1019 (min)	56.1	2
read length for 99% accuracy[b]	940	3
read length for 98% accuracy[b]	1003	3

[a] $n = 10$ samples. [b] The consensus sequence was obtained from Whitehead Institute/MIT Center for Genome Research. Read length is defined in footnote *a* in Table 2.

(43) Berno, A. J. *Genome Res.* **1996,** *6*, 80−91.
(44) Brady, D.; Kosic, M.; Miller, A. W.; Karger, B. L., in preparation.
(45) Daly, M. J. "Trout" signal processing and base calling software, Whitehead Institute/MIT Center For Genome Research, 1996.
(46) Ewing B.; Hillier, L.; Wendl, M. C.; Green, P. *Genome Res.* **1998**, *8*, 175−185.
(47) Giddings, M. C.; Brumley, R. L.; Haker, M.; Smith, L. M. *Nucleic Acids Res.* **1993,** *21*, 4530−4540.
(48) Ives, J. T.; Gesteland, R. F.; Stockham, T. G. *IEEE Trans. Biomed. Eng.* **1994,** *41*, 509−519.
(49) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985,** *33*, 103−119.
(50) Ebright, R.; Dong, Q.; Messing, J. *Gene* **1992,** *114*, 81−83.

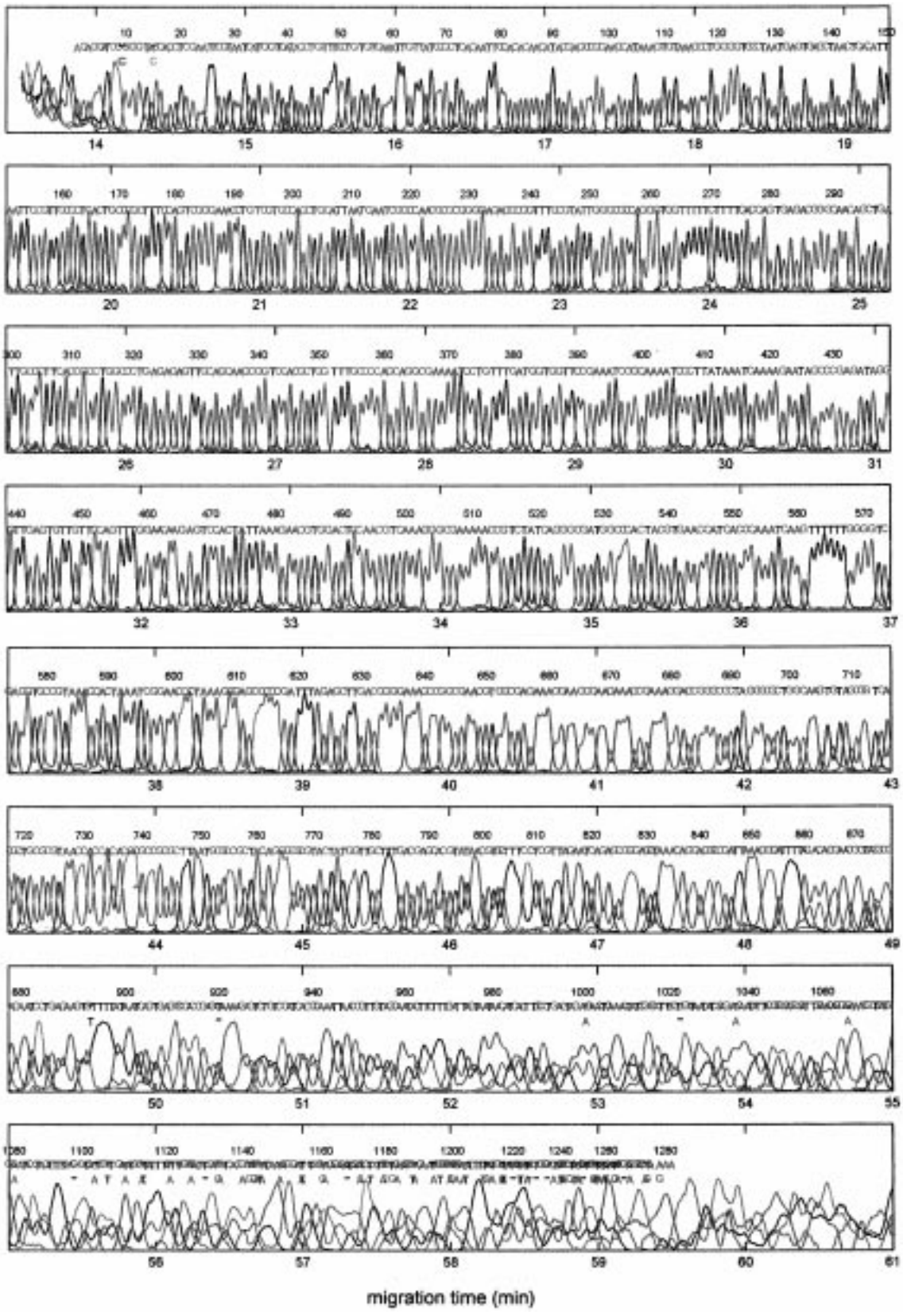

Figure 3. Electrophoretic separation of DNA sequencing fragments generated on ssM13mp18 with BigDye-labeled universal (−21) primer and AmpliTaq FS at the optimum experimental conditions: 2.0% (w/w) LPA 9 MDa and 0.5% (w/w) 50 kDa LPA, 200 V/cm and 60 °C. Electrophoretic conditions were as follows: effective length 30 cm, total length 45 cm, 75-μm-i.d., 365-μm-o.d. coated capillary (poly(vinyl alcohol)); both cathode and anode running buffers 50 mM Tris/50 mM TAPS/2 mM EDTA. Cathode running buffer also contained 7 M urea, the same as in the separation matrix. The samples were injected at a constant electric field of 25 V/ cm (0.7 μA) for 10 s and electrophoresed at 200 V/cm (10.2 μA) at 60 °C. Base-calling errors are marked under the following sequence: dash = overcall; base letter = substitution; base letter with dash above it = undercall.

columns require reconditioning for a long time between runs, and in many cases, they last only a limited number of runs.[23] With PVP, the reconditioning procedure is simpler, but the lifetime of the column was determined to be ~30 runs.[25] With poly(dimethylacrylamide) separation matrixes, columns may operate for 100 runs; however, the lifetime of failed columns even after reconditioning did not exceed 20 runs.[11]

Poly(vinyl alcohol)-coated capillaries with LPA polymer solutions are routinely used in this laboratory for DNA separation.[31] This coating is very effective, and the columns provided excellent DNA separations. Initially, when working with such coated columns, it was found that, after about 7−10 consecutive runs with direct replacement of the separation matrix, the efficiency of separation gradually decreased. However, it was determined that, washing with 20 or more column volumes of deionized water after 5 consecutive runs, the efficiency of the column was maintained. Furthermore, for overnight storage, the capillary was also washed and filled with deionized water. The gradual decrease in the column performance without the water wash step may be the result of partial trapping of DNA fragments by the coating polymer network. Such trapping at the wall would lead to significant negative charge on the coating and, thus, electroosmotic flow, which could reduce separation performance of the LPA polymer solution.

To test column longevity, we performed more than 300 consecutive sequencing runs on a single column (see Table 2). We analyzed different samples, e.g., sequencing standard samples made with M13mp18 DNA, clones of human chromosome 17, DNA ladders, etc. The results, shown in Table 2, indicate that the separation characteristics of the column throughout its use remained constant, even under demanding separation conditions. We arbitrarily stopped testing the column after 325 runs; therefore, the column could conceivably have a much longer lifetime. The polymer-coated columns were found to provide excellent column-to-column and sample-to-sample reproducibility of DNA sequencing. These are very important issues for the ruggedness in an automated multiple capillary DNA sequencer. In the high-throughput sequencing system, use of PVA-coated columns would more than compensate the additional cost of coating by the savings from less frequent replacement of the capillary columns.

CONCLUSIONS

Substantial improvements in DNA sequencing technology are needed to achieve routinely long reads in rapid time. This paper describes progress in high throughput and rugged sequencing technology by CE using mixed replaceable LPA solutions that can be implemented on multicapillary instrumentation. The electrophoretic separation of more than 1000 DNA sequencing fragments of M13mp18 in less than 55 min, with a base-calling accuracy of 99% for 1000 called bases, is now routinely performed in our laboratory. The separation characteristics of polymer-coated capillary columns have remained at the same high performance level for over 300 cumulative runs. Implementation of the optimized sequencing procedure into a high-speed capillary array instrument is currently underway in this laboratory.

ACKNOWLEDGMENT

The authors gratefully acknowledge the DOE under the Human Genome Project Grant DE-FGO2-90ER 60985 for support of this work. Support by DOE does not constitute an endorsement of the views expressed in this article. The authors also thank the Whitehead Institute/MIT Center for Genome Research for providing the cloned sequencing templates and helpful discussions. This is contribution No. 738 from the Barnett Institute.

Received for review April 28, 1998. Accepted July 10, 1998.

AC980457F