ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/14949557

Ruiz-Martinez, M.C. et al. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. Anal. Chem. 65, 2851-2858

ARTICLE *in* ANALYTICAL CHEMISTRY · NOVEMBER 1993

Impact Factor: 5.64 · DOI: 10.1021/ac00068a023 · Source: PubMed

CITATIONS
226

READS
132

6 AUTHORS, INCLUDING:

Jan Berka

21 PUBLICATIONS **6,733** CITATIONS

SEE PROFILE

Frantisek Foret

Institute of Analytical Chemistry, CAS, Brno, …

177 PUBLICATIONS **5,288** CITATIONS

SEE PROFILE

Barry Karger

Northeastern University

363 PUBLICATIONS **16,808** CITATIONS

SEE PROFILE

Available from: Frantisek Foret
Retrieved on: 29 January 2016

DNA Sequencing by Capillary Electrophoresis with Replaceable Linear Polyacrylamide and Laser-Induced Fluorescence Detection

Marie C. Ruiz-Martinez, Jan Berka,† Alexei Belenkii, Frantisek Foret,‡ Arthur W. Miller,§ and Barry L. Karger*

Barnett Institute, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Replaceable linear polyacrylamide (LPA) has been utilized as a sieving matrix for DNA sequencing by capillary electrophoresis (CE). Difficulties associated with cross-linked polyacrylamide gel stability have been overcome for the routine application of CE to DNA sequencing. A simple laser-induced fluorescence (LIF) detection system based on a single laser and two photomultipliers (PMT) has been adopted for this work. Sequencing information for four bases has been obtained from two fluorescent dyes and two peak height ratios, detected in two optical channels. FAM- and JOE-labeled M13 (-21) primers have been chosen because both dyes are efficiently excited with a low-power argon ion laser, can be optically separated, and exhibit minimal dye-based shifts in DNA fragment mobilities. Addition of denaturants to the electrophoresis running buffer (1 × TBE, 3.5 M urea, 30% formamide) and column operation at 32 °C permitted the resolution of difficult compressed sites in the sequence of phage M13mp18. Careful examination of the polymerization reaction of LPA has led to methodology that has proven to be reproducible for obtaining DNA sequencing information of M13mp18 phage for 350 nucleotides in close to 30 min.

INTRODUCTION

High-speed DNA sequencing methods are under rapid development, both for the needs of the Human Genome Project and for general sequencing applications. At present, high-resolution electrophoretic techniques are the standard procedures for the separation of DNA sequencing reaction products. The limited throughput of standard gel electrophoretic techniques, however, remains a barrier for the sequencing of large genomes. Furthermore, rapid, automatable methods that are cost effective for clinical, forensic, and general molecular biological applications are required.

Since the first reports on ultrahigh resolution of DNA mixtures using capillary gel electrophoresis,[1] CE has been successfully applied to different kinds of DNA analysis, including the following: DNA sequencing,[2-6] separation of restriction fragments,[7-14] synthetic oligonucleotides,[6,15] and PCR products.[6,16-18] The relatively low currents in fused silica capillary columns enable high electric fields to be applied without damage to the separation matrix from overheating. Large electric field strengths result in an order of magnitude increase in separation speed over conventional methods. The overall throughput of CE can be further increased by analyzing many samples in parallel, e.g., through the use of capillary arrays.[19-21] A related modification of standard slab gel electrophoresis has been described utilizing horizontal ultrathin slab gels.[22,23]

In the case of CE, adoption for DNA sequencing has been limited. Problems related to the stability of gel-filled capillaries with sequencing reaction products hinder the full automation of CE-based systems. The instability of cross-linked gels in capillary columns is manifested by either formation of bubbles or a steady decline of current, caused by a resistivity increase normally at the injection end of the capillary. These problems make the gel-filled capillaries useable for only a few injections, even when the capillary end is trimmed after each run. Although the effects of formation of anomalous conductivity zones, resulting in ion depletion at the gel/liquid interface,[24] can be reduced by adding denaturant and acrylamide monomers to the buffer reser-

* Author to whom reprint requests and correspondence should be sent.
† On leave from The Institute of Pathological Physiology, School of Medicine, Masaryk University, Brno, Czech Republic.
‡ On leave from The Institute of Analytical Chemistry, Veveri 97, Brno, Czech Republic.
§ College of Computer Science, Northeastern University.

(1) Cohen, A. S.; Najarian, D. R.; Paulus, A.; Guttman, A.; Smith, J. A.; Karger, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9660–9663.
(2) Drossman, H.; Luckey, J. A.; Kostichka, A. J.; D'Cunha, J.; Smith, L. M. *Anal. Chem.* **1990**, *62*, 900–903.
(3) Cohen, A. S.; Najarian, D. R.; Karger, B. L. *J. Chromatogr.* **1990**, *516*, 49–60.
(4) Luckey, J. A.; Drossman, H.; Kostichka, A. J.; Mead, D. A.; D'Cunha, J.; Norris, T. B.; Smith, L. M. *Nucleic Acids Res.* **1990**, *18*, 4417–4421.
(5) Karger, A. E.; Harris, J. M.; Gesteland, R. F. *Nucleic Acids Res.* **1991**, *19*, 4955–4963.
(6) Swerdlow, H.; Zhang, J. Z.; Chen, D. Y.; Harke, H. R.; Grey, R.; Wu, S.; Dovichi, N. J. *Anal. Chem.* **1991**, *63*, 2835–2841.
(7) Heiger, D. N.; Cohen, A. S.; Karger, B. L. *J. Chromatogr.* **1990**, *516*, 33–48.
(8) Grossman, P. D.; Soane, D. S. *J. Chromatogr.* **1991**, *559*, 257–266.
(9) Motsch, S. R.; Kleemis, M. H.; Schomburg, G. *J. High Resolut. Chromatogr.* **1991**, *14*, 629–632.
(10) Bocek, P.; Chrambach, A. *Electrophoresis* **1992**, *13*, 31–34.
(11) Guttman, A.; Wanders, B.; Cooke, N. *Anal. Chem.* **1992**, *64*, 2348–2351.
(12) Stege, M.; Lagu, A. *Anal. Chem.* **1991**, *63*, 1233–1236.
(13) MacCrehan, W. A.; Rasmussen, H. T.; Northrop, D. M. *J. Liq. Chromatogr.* **1992**, *15*, 1063–1080.
(14) Baba, Y.; Matsuura, T.; Wakamoto, K.; Tsukako, M. *J. Chromatogr.* **1991**, *558*, 273–284.
(15) Paulus, A.; Ohms, J. I. *J. Chromatogr.* **1990**, *507*, 1113–1123.
(16) Schwartz, H. E.; Ulfelder, K.; Sunzeri, F. J.; Busch, M. P.; Brownlee, R. G. *J. Chromatogr.* **1991**, *559*, 267–283.
(17) McCord, B. R.; McClure, D. L.; Jung, J. M. *J. Liq. Chromatogr.*, in press.
(18) Nathakarnkitkool, S.; Oefner, P. J.; Bartsch, G.; Chin, M. A.; Bonn, G. K. *Electrophoresis* **1992**, *13*, 18–30.
(19) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 2149–54.
(20) Kambara, H.; Takahashi, S. *Nature* **1993**, *361*, 565–566.
(21) Taylor, J. A.; Yeung, E. A. *Anal. Chem.* **1993**, *65*, 956–960.
(22) Kostichka, A. J.; Marchbanks, M. L.; Brumley, R. L., Jr.; Drossman, H.; Smith, L. M. *Bio/Technology* **1992**, *10*, 78–81.
(23) Ansorge, W.; Voss, H.; Wiemann, S.; Schwager, C. H.; Sproat, B.; Zimmermann, J.; Stegemann, J.; Erfle, H.; Hewitt, N.; Rupp, T. *Electrophoresis* **1992**, *13*, 616–619.
(24) Spencer, M. *Electrophoresis* **1983**, *4*, 36–41.

0003-2700/93/0365-2851$04.00/0 © 1993 American Chemical Society

voirs,[25] cross-linked or high % T nonreplaceable linear polyacrylamide-filled capillaries still cannot generally be used for more than 5–10 sequencing runs without a significant loss in signal and resolution.[26,27] In addition, the cross-linked gel columns cannot be used above room temperature. To permit multiple consecutive DNA sequencing runs on the same capillary, additional methods for sample purification, including DNA template removal, have been described.[25,28] Despite these developments, there is still a critical need for improved column design for routine use.

In this work, low- to moderate-viscosity linear polyacrylamide (LPA) matrices for DNA sequencing analysis have been developed in order to overcome the problems of gel stability. The advantage of such matrices is that, along with providing powerful separation, the LPA can be replaced after each run. In effect, this means that each sequencing run is conducted on a fresh column. Furthermore, the denaturing agent of the running buffer has been modified from the typical 7 M urea to 3.5 M urea plus 30% (v/v) formamide in order to reduce the viscosity of the medium while enhancing the denaturing ability of the buffer with the addition of formamide.[29] With replaceable LPA, the most laborious and time-consuming steps in CE, i.e., the preparation and alignment of the separation column, are reduced to a simple reloading of the sieving medium.[30]

The sequencing strategy utilizes modification of Sanger–Coulson dideoxy sequencing chemistry to achieve approximately equal product amounts,[31,32] along with a single-laser, two-fluorescent-dye primer approach, related to that recently reported;[33] however, on-column detection has been used instead of the flow sheath cuvette. Such a simple and relatively inexpensive system can be accessible to many laboratories. Moreover, this system is, in principle, compatible with a capillary array arrangement. Of course, other sequencing strategies, e.g., four-dye-labeled primers or terminators, could be incorporated into the replaceable matrix format. In principle, CE using replaceable LPA matrices should also be compatible with membrane collection/multiplexing schemes.[34]

EXPERIMENTAL SECTION

Instrumentation. The instrumental design using one laser and two photomultiplier tubes (PMTs) is illustrated in Figure 1. The power supply (point 16 in Figure 1) was 60 kV dc (Model PS/MK30, Glassman, Whitehouse Station, NJ). Laser light from the 488-nm emission of a low-power argon ion laser (7 mW) (point 1) (Model 532, Omnichrom, Chino, CA) was first passed through a narrow band-pass filter (488 nm) (point 2) (Melles Griot, Irvine, CA), and then the intensity of the light was decreased by a neutral density filter (Melles Griot) (point 3) to 1.2 mW. The laser light was next reflected by a 45° mirror (point 4) (Newport, Fountain Valley, CA) and focused by an achromatic lens (focal length 20 mm) (point 5) (Melles Griot) onto a detection window of a horizontally positioned fused silica capillary of dimensions 75-μm i.d., 360-μm o.d. (point 6) (Polymicro Technologies, Inc., Phoenix, AZ). The detection window (3 mm in length) was formed

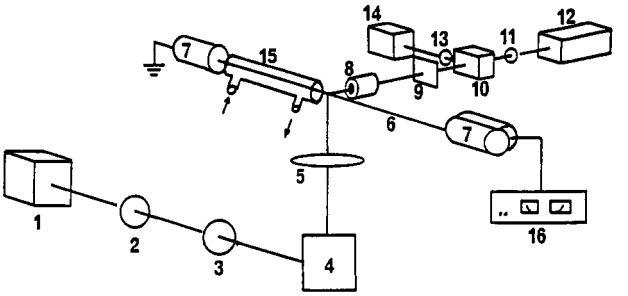

Figure 1. Diagram of the DNA sequencing instrument. Details of the component parts can be found in the Experimental Section.

by stripping the polyimide coating off the capillary with a razor blade. The buffer reservoirs (point 7) were positioned at the same height and refilled daily with fresh, filtered buffer. At the anodic end, a double buffer reservoir filled with 1 × TBE, which separated the capillary from the electrode chamber, was utilized to protect the capillary from water electrolysis products which can react with gel or buffer components, causing an increase of background fluorescence noise during prolonged operation under high electric fields.[35]

At the cathodic end, a single buffer reservoir filled with 1 × TBE, 7 M urea buffer was used. The capillary was placed inside a stainless steel tube of 380-μm i.d., which was then inserted into a water jacket 10 cm in length (point 15). For capillaries with an effective length of 18 cm, about 4 cm of the column protruded from the water jacket at the cathodic end for insertion into the buffer reservoir and another 4 cm of the capillary protruded from the end of the water jacket to the detection window. The water jacket was controlled by a circulating bath at 32 °C.

Light emitted from the capillary was collected by a 40× microscope objective (Oriel, Stamford, CT) (point 8) and passed through a high-pass, 488-nm reflecting filter (Melles Griot) (point 9), followed by a visible spectrum 50/50 beam splitter (Type BS-550-S, Corion, Holliston, MA) (point 10). The fluorescence signal transmitted through the beam splitter was filtered by an interference filter (515 ± 10 nm) (point 11) (Corion), which isolated the emission from the DNA fragments that were fluorescently labeled with the primer FAM (Applied Biosystems, Foster City, CA), for detection by a PMT (Hamamatsu Type 647, Hamamatsu Photonics K.K., Japan) (point 12). The light reflected from the beam splitter was filtered by a 550 ± 10-nm interference filter (Corion) (point 13), which isolated the emission of DNA fragments that were fluorescently labeled with the primer JOE (Applied Biosytems), for detection by a second PMT (Hamamatsu Type 928) (point 14). The signals were collected at a 10-Hz sampling rate by a Perkin-Elmer Nelson data acquisition system using Turbochrom software (Version 3.2) running on an IBM compatible 486 DX-2 personal computer. The residual FAM emission in the JOE channel was subtracted by computer data processing. Finally, the sequence was coded and analyzed by the software described in the Appendix section.

DNA Sequencing Reaction and Sample Preparation. Two commercially available labeled primers FAM and JOE (−21) M13 were utilized (Applied Biosystems). One picomole of fluorescent dye-labeled primer, FAM or JOE, was annealed to 0.5 pmol of M13mp18 single-stranded DNA template (New England Biolabs, Inc., Beverly, MA) in 1 × sequencing buffer (USB, Cleveland, OH) by heating the solution to 65 °C and cooling it to 20 °C for 30 min. Nucleotide termination mixtures containing 720 μM portions of each of the deoxynucleotides (dNTP) (USB) and a 7.2 μM portion of a single dideoxynucleotide (ddNTP) (USB) were prepared. The proper product ratio, 3:1, was obtained by mixing 16.5 μL of the termination mixture ddC and 3.5 μL of the termination mixture ddT for the JOE-primed sequencing reaction. For the FAM-primed sequencing reaction, 12 μL of the termination mixture ddA plus 8 μL of the termination mixture ddG were mixed to obtain the product ratio of 3:1, respectively.

(25) Swerdlow, H.; Dew-Jager, K.; Brady, K.; Grey, R.; Dovichi, N. J.; Gesteland, R. *Electrophoresis* **1992**, *13*, 475–83.
(26) Smith, L. M. *Nature* **1991**, *349*, 812–813.
(27) Pentoney, S. L.; Konrad, K. D.; Kaye, W. *Electrophoresis* **1992**, *13*, 467–74.
(28) Tong, X.; Smith, L. M. *Anal. Chem.* **1992**, *64*, 2672–2677.
(29) Rocheleau, M. J.; Grey, R. J.; Chen, D. Y.; Harke, H. R.; Dovichi, N. J. *Electrophoresis* **1992**, *13*, 484–486.
(30) Sudor, J.; Foret, F.; Bocek, P. *Electrophoresis* **1991**, *12*, 1056–1058.
(31) Tabor, S.; Richardson, C. S. *J. Biol. Chem.* **1990**, *265*, 8322–6.
(32) Ansorge, W.; Zimmermann, J.; Schwager, C.; Stegeman, T.; Erfle, H.; Voss, H. *Nucleic Acids Res.* **1990**, *18*(11), 3419–3420.
(33) Chen, D. Y.; Harke, H. R.; Dovichi, N. J. *Nucleic Acids Res.* **1992**, *20*, 4873–80.
(34) Church, G. M.; Kieffer-Higgins, S. *Science* **1988**, *240*, 185–188.
(35) Carson, S.; Cohen, A. S.; Belenkii, A.; Ruiz-Martinez, M. C.; Berka, J.; Karger, B. L. *Anal. Chem.*, in press.

Chain extension and termination reactions were initiated in the presence of 5.5 mM DTT (USB) and 1 × Mn buffer (USB) by the addition of 1 μL (13 units) of Sequenase (Version 2.0 USB) and 1 μL (0.005 units) of inorganic pyrophosphatase (USB). After 10 min of incubation at 37 °C, the reactions were terminated by ammonium acetate/ethanol precipitation at −20 °C for 2 h. DNA was then washed twice with 70% ethanol, dried under vacuum, and resuspended in 5 μL of 95% formamide:5% (50 mM) EDTA. The FAM- and JOE-labeled fragments were mixed in equal amounts, and the sample was heated to 95 °C for 5 min, followed by chilling on ice before electrokinetic injection into the capillary.

Capillaries and Linear Polyacrylamide Preparation. The capillaries were coated with LPA as previously described by Hjerten et al.[36] The replaceable LPA matrices were prepared as follows: 0.60 g of acrylamide (ultrapure, ICN, Costamesa, CA) was dissolved in 10 mL of 1 × TBE, 30% formamide, 3.5 M urea buffer, filtered (0.2-μm pore size), and degassed. The monomer solutions were typically polymerized by addition at room temperature of 2 μL of 10% (w/v) ammonium persulfate (ICN) and 1 μL of 100% TEMED (Bio-Rad, Richmond, CA) per milliliter of monomer solution (i.e., final concentration of APS = 0.02% (w/v), TEMED = 0.1% (v/v)). (Recently, the polymerization procedure has been optimized for milliliter volume sizes.) The solutions were filled into 100-μL gas-tight syringes (Rainin, Woburn, MA) and then placed in either a cooling water bath thermostated at 10 °C or a refrigerator set at 0 °C. Polyacrylamide-filled syringes were stored in the refrigerator (0 °C) and could be used for at least 10 days.

The LPA matrix was replaced in the capillary before each run by connecting the 100-μL syringe needle to the capillary with a short piece of Teflon tubing. After fresh polymer matrix was loaded into the column, the capillary was electrodialyzed for 10 min at a constant electric field of 250 V/cm in order to decrease the background fluorescence. Electrodialysis can be conducted on the polymer matrix directly; however, in this work, it was more practical to electrodialyze the matrix in the capillary prior to DNA sequencing analysis.

The electrophoretic conditions for separation of the reaction products of DNA sequencing were as follows: electric field = 250 V/cm, current = 7 μA, effective capillary length = 18 cm, total length = 33 cm, column temperature = 32 °C (10 cm of the capillary effective length heated). The whole analysis time, including the replacement of the matrix and electrodialysis, was about 45 min, with separation achieved in close to 30 min.

Viscosity and Oxygen Content Measurements. The viscosity of the polymer networks were measured by means of a falling ball, size 3, tantalum ball viscometer (Gilmont, Barrington, IL) in a water bath at a constant temperature of 18 °C. The oxygen concentration of the acrylamide solutions was measured at 20 °C by a dissolved oxygen meter with RTD/polarographic oxygen sensor (Type G-05505-00, Cole Palmer, Niles, IL).

RESULTS AND DISCUSSION

As noted in the Introduction, several reports have discussed the lack of the stability of gel-filled capillaries as an impediment to the routine application of CE to DNA sequencing. If sufficient resolution were achievable, low to moderate-viscosity LPA networks would permit the use of replaceable sieving matrices for the separation of DNA sequencing reaction products. Previously, this laboratory showed that LPA (9% T) provided single-base resolution of polyadenylic acids,[7] but this matrix was too viscous to allow easy replacement from the capillary. In the present work, we therefore studied lower percentages of LPA for the single-base separation of DNA fragments. The sequencing strategy utilized a simple base coding and a single-laser, two-PMT instrument that could be easily incorporated into a practical setting.

Instrument Configuration. The instrumentation design was related to that recently described using one laser and two PMTs.[33] The difference between that system and the present one was that the detection of fluorescence was on-column and not with a sheath flow cuvette. Previous studies from this laboratory showed similar detection levels of dye-labeled DNA fragments with on-column detection, as reported for the sheath flow cuvette.[35] For the detection of small fragments (e.g., up to 300–400 bases in length), the four-dye-labeling approach (each dye specific for one base), which provides the best spectral resolution, was not required because of suffiicent electrophoretic separation for the low-base-number fragments. The sequencing information was further encoded using different relative peak heights for two bases, each detected in one of two channels, FAM or JOE. The selection of these two-dye-labeled primers was based in part on their negligible electrophoretic migration difference for the same base length. Indeed, the software used to analyze the raw data determined that the difference in mobility between the two primers was about 0.7 s, in agreement with that recently reported.[19] Furthermore, FAM and JOE primers permitted the use of a low-power argon ion laser, since both dyes were efficiently excited at 488 nm.

Table I. Detection System Specification

	FAM channel	JOE channel
λ excitation	488 nm	488 nm
λ emission max	535 nm	557 nm
λ detection	515 ± 10 nm	550 ± 10 nm
FAM emission	100%	30 ± 3%
JOE emission	<3%	100%

Table I presents the system configuration and the percent cross-talk between the emissions in the FAM and JOE spectral channels. The emission wavelength maxima of FAM and JOE are 535 and 557 nm, respectively, but in this work the FAM detection band was shifted to 515 ± 10 nm in order to eliminate the overlap of the JOE signal in the FAM channel. The JOE emission in the FAM channel was then below 3%, which was negligible for our purposes. The JOE channel (detected at 550 ± 10-nm band, determined by a filter), contained 30 ± 3% interference of the FAM emission, and this interference was removed by software. In addition, due to the high migration velocity of the DNA fragments in this separation protocol, the sampling rate for data acquisition was selected to be 10 Hz.

DNA Sequencing Reaction. Various DNA sequencing protocols have been developed using T7 DNA polymerase, manganese, and inorganic pyrophosphatase for sequencing reactions where of (a) single-dye-based coding of bases with four different peak heights,[31,32] or (b) single-dye-based coding of three bases by peak height ratios plus one base coded by a gap,[27,32] (c) two-dye binary coding of three bases with one base coded by a gap and with two optical channels,[19] or (d) two-dye and two peak height ratio coding of four bases with two optical channels[33] have been used. For this work we selected approach d for simplicity while maintaining good discriminating power.

We found that the use of a 100:1 concentration ratio of dNTPs:ddNTPs resulted in good signal-to-noise ratio, even for small peak-coded bases, in the range up to at least 350 nucleotides. For peak height ratio coding, the adjustment of volumes of the dNTP/ddNTP termination mixtures was straightforward for a given batch of reagents. Ratios from 2.5:1 to 3:1 produced a low base identification error. Moreover, a second sequencing run using inverted peak height ratios could be easily accomplished for sequence confirmation. The volumes of the particular dNTPs/ddNTP termination mixes necessary for the inverted ratio were simply inverted in the appropriate manner.

Linear Polyacrylamide Matrix. The LPA composition was optimized in order to provide reproducible and rapid separation of DNA fragments for a minimum of 300 bases.

(36) Hjerten, S. *J. Chromatogr.* 1985, *347*, 191–8.

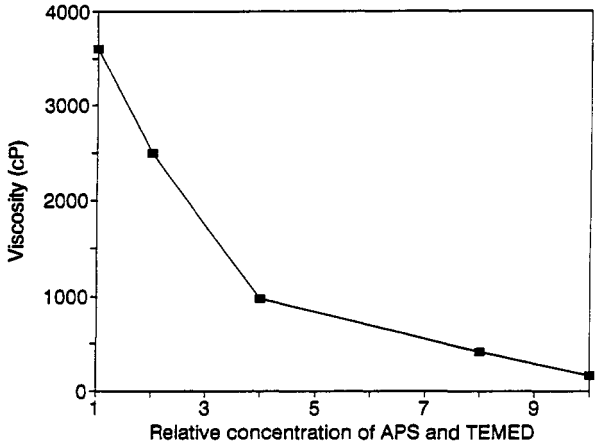

Figure 2. Viscosity of 6% T matrix as a function of concentration of APS and TEMED. Buffer: 1 × TBE, 30% formamide, 3.5 M urea. See Table II for concentrations of APS and TEMED.

Low- to moderate-viscosity LPA allowed faster separations than those reported[6,37] using high % T LPA filled capillaries. The following polymerization parameters were optimized: buffer composition, oxygen content in the monomer solution, radical initiator and catalyst concentration, and polymerization temperature.

For the facile replacement of the polyacrylamide matrix, the composition of the denaturant in the buffer was first altered to include formamide in conjunction with a smaller concentration of urea than previously used.[35] As a denaturing agent, formamide added several advantages, including decreased viscosity of the polymer network relative to 7 M urea and increased decompression of sequences with the secondary structures.[29] On the other hand, formamide or urea decomposition during prolonged electrophoretic runs, especially those performed above room temperature, could potentially lead to an increase in current and thus affect separation reproducibility. This decomposition did not play a role during 30–60-min analyses, and the replacement of fresh polymer after each run did not allow denaturant decomposition to affect the separation.

Oxygen is known to be a radical quencher which inhibits acrylamide polymerization.[38] Thus, minimizing the oxygen concentration in the monomer solution is critical for the efficient and reproducible synthesis of polyacrylamide chains. Although the common laboratory protocols for polyacrylamide slab gel preparation recommend only vacuum degassing of the monomer solution, bubbling with helium is a more efficient technique for the oxygen removal. Bubbling a 10-mL solution with helium for 2 h gave a reproducible oxygen content of less than 2 ppm.

Experiments were next undertaken to optimize the amount of radical initiator, ammonium persulfate (APS), and catalyst, N,N,N′,N′-tetramethylethylenediamine (TEMED). The relative proportions of both compounds were kept constant, but the total concentration was varied for the synthesis of polyacrylamide using the same monomer concentration (6% T). In this study, the reaction was allowed to take place at room temperature for a period of 48 h to assure that no more reaction was occurring.

In order to characterize the LPA obtained, the viscosities of the matrices were measured at 18 °C, as described in the Experimental Section, and the results are shown in Figure 2. It can be seen that the viscosity increases by 22.5-fold with a 10-fold decrease in APS/TEMED concentration. It is known that the polymer molecular weight is related to the viscosity of matrix. As the concentrations of TEMED and APS are reduced, fewer acrylamide primary chains are initiated for the same concentration of monomer;[39] therefore, longer fibers or higher molecular weight polymers are produced. The lowest concentration of TEMED/APS, 0.05% (v/v)/0.1% (w/v), while yielding the highest viscosity, was not routinely used because of difficulty in reproducing the polymerization; however, it was a useful matrix to explore trends.

Table II. Viscosity of Linear Polyacrylamide vs Concentration of Radical Initiator and Its Effect on Separation[a]

η (cp)	TEMED (v/v) (%)	APS (w/v) (%)	N/m	α (226–227)	Rs (226–227)	t (227) (min)
3600	0.05	0.01	3.7×10^6	1.0036	0.54	64.8
2500	0.10	0.02	2.7×10^6	1.0034	0.49	50.8
980	0.20	0.04	2.5×10^6	1.0031	0.44	45.5
160	0.50	0.10	1.2×10^6	1.0030	0.40	43.6

[a] η, viscosity; N/m, number of theoretical plates per meter; α, selectivity; Rs, resolution; t, migration time. Electrophoretic conditions: capillary effective length, 24 cm (total length 40 cm); constant electric field, 200 V/cm; running temperature, 25 °C; 1 × TBE; pH = 8.4; 30% formamide; 3.5 M urea.

The pressure drop required to replace the LPA matrix of 2500 cp in the separation capillary was calculated with the assumption that the network behaved as a Newtonian fluid,

$$\Delta P = 8LQ\eta/\pi R^4 \quad (1)$$

where ΔP = the required pressure to replace the matrix (Pa), L = the length of the capillary (m), Q = the flow rate of the network (m^3/s), η = the viscosity of the polymer (Pa s), and R = the radius of the capillary (m).[40] The pressure required for replacement of the matrix at room temperature in a 33-cm-long capillary in 3 min was calculated to be 84 atm (1.25×10^3 psi). This pressure can be obtained using a 100-μL gas-tight syringe connected to the capillary with a short piece of Teflon tubing. Syringe pumps could also be employed in an automated system.

We next examined the LPA matrices in Figure 2, all at 6% T monomer concentrations, and the results are shown in Table II. In this case, ddTTP was used as terminator for the sequence reaction with M13mp18. It can be seen that both efficiency and the ratio of mobilities for T226/T227 increase as the viscosity and thus the polymer molecular weight of the matrix increases. The price to pay for the resultant improved resolution is slower migration time; however, an increase in electric field and a decrease in column length leads to a more rapid separation (see Figure 5).

Similar trends of improved resolution with increased matrix viscosity for a given monomer concentration can be seen in Figure 3. In this case, the TEMED/APS were maintained constant at 0.10% (v/v)/0.02% (w/v), and the temperature of polymerization was varied from 18 to 10 to 0 °C. Sufficient time for the reaction to go to completion was allowed. Spectrophotometric determination of residual free acrylamide[41] after 48 h of polymerization confirmed that the degree of polymerization was more than 94%. It can be seen that the resolution increased as the temperature of polymerization was decreased. The lowest temperature (0 °C) also yielded the greatest viscosity and presumably the longest fibers. For convenience, we used 10 °C in our sequencing studies, since the matrix could still be replaced in the capillary in 2–3 min. The molecular weight of this LPA was determined by light scattering to be approximately 1×10^6.

(37) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Anal. Chem.* **1992**, *64*, 967–970.
(38) Chrambach, A.; Rodbard, D. *Science* **1971**, *172*, 440–50.
(39) Nossal, R. *Macromolecules* **1985**, *18*, 49–54.
(40) *Unified Separation Science*; Giddings, J. C., Ed.; John Wiley & Sons, Inc.: New York, 1991; p 60.
(41) MacWilliams, D. C.; Kaufman, D. C.; Waling, B. F. *Anal. Chem.* **1965**, *37*, 1546–1552.

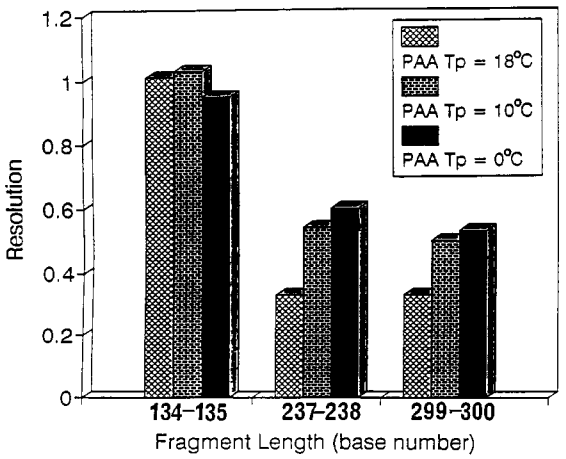

Figure 3. Resolution vs base number for 6% T linear polyacrylamide networks polymerized at three different temperatures. Conditions: capillary effective length = 18 cm (total length = 33 cm), constant electric field = 250 V/cm, running temperature = 32 °C, 6% T, 1 × TBE, 30% formamide, 3.5 M urea.

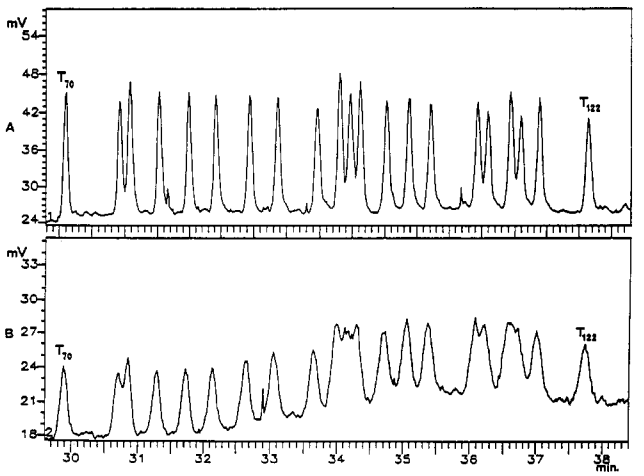

Figure 4. The effect of capillary column orientation on the separation of the FAM–ddT DNA sequencing fragments: (A) horizontal capillary and (B) vertical capillary. Conditions: capillary effective length = 20 cm (total length = 33 cm), constant electric field = 200 V/cm, room temperature. The network composition was 6% T in 1 × TBE, 30% formamide, 3.5 M urea.

Work is continuing on determining the role of molecular weight of various polymerized matrices and linear polyacrylamide standards on separation; however, the results of Table II and Figure 3 show a clear trend to improved resolution with increasing polymer size (within the polymer range studied), in agreement with the earlier conclusion of Bode on slabs.[42] It is known that when the polymer concentration is well above the entanglement threshold, the mesh size will be dependent on polymer concentration but not polymer molecular weight.[43] However, the flexibility of the fibers should become less as the molecular weight increases. As recently presented by Bae and Soane,[44] the relaxation time in a mesh of the network for polymers of moderate molecular weight ($<10^6$ Da) was found to be comparable to the residence time of 100 bp ds fragments. They predicted that above 100 bp, resolution would decrease. Moreover, the relaxation time for fiber movement was 1 order of magnitude lower for cross-linked gels. Given the increased viscosity with higher molecular weight polymers for a fixed monomer concentration, it may well be that such matrices approach more closely cross-linked gels than lower molecular weight non-cross-linked polymer networks. These ideas will be discussed in more detail in a separate paper.[45]

Separation of DNA Sequencing Reaction Products. Critical to the separation of DNA sequencing reaction products by CE is the production of columns with the highest efficiency possible. Recent reports have identified several causes of dispersion in capillary gel electrophoresis. Wicar et al.[46] determined that coiling (or uncoiling) the capillary after filling it with LPA can distort the structure of the polymer networks, especially for low-viscosity matrices. At the beginning of our studies with 6% T LPA, the instrumental system was designed with the capillary in a vertical position, and poor efficiencies were obtained. It was theorized that the vertical position of the capillary could create a pressure drop that could influence flow of the polymer network, thus distorting the mesh structure. Horizontal positioning of the capillary should then remove this pressure drop and should improve the separation. Figure 4 shows the effect of the capillary position on the separation of DNA fragments. The clear advantage of horizontal columns can be seen. This result also suggests that no electroosmotic flow at the capillary walls should occur, or the polymer network structure may be distorted.

We next optimized separation conditions for the M13mp18 DNA sequencing reaction products, using the 6% T LPA matrix, with polymerization at 10 °C and the TEMED/APS concentrations of 0.10% (v/v)/0.02% (w/v), respectively. As shown in Figure 5, where a full sequence is run, these conditions included E = 250 V/cm, effective length = 18 cm, and column temperature = 32 °C. The red lines represent the JOE channel with C > T (3:1) and the blue lines the FAM channel with A > G (3:1). As can be seen in this figure, the sequence is read to 350 bases in 30.5 min. The sequence can be read by eye up to 300-310 bases, with the software adding approximately 50 bases.

For sequence accuracy, a second run is typically made, in which the ratios of each channel are inverted. As noted in the Appendix, if there is an ambiguity, the software calls the base from the run with the higher ratio. With this approach, no errors are found up to 350 bases for M13mp18, a known sequence.

The importance of the two runs is best seen in Figure 6, where the electropherograms are shown for G57-A74 and A230-G249. The partial compressions at G59-G60, G66-A67, and especially G234-G236 are clarified when G > A (3:1). Another approach to minimize ambiguity for severe compressions is to elevate the column temperature to >50 °C to alleviate strand curvature. LPA can be successfully operated at elevated temperatures, whereas cross-linked gels would collapse under such heat. Furthermore, replacing LPA after the run minimizes column fouling due to formamide or urea decomposition at the elevated temperatures. The rapid separation and ease of refilling allows different column conditions to be used, when necessary, to increase sequence accuracy. For general purposes, 32 °C is preferred, since at higher column temperatures some increase in peak widths was observed.

We next explored the reproducibility of sequencing runs. Precise control of the capillary column temperature is important to obtain good reproducibility. Without temperature control, due to laboratory temperature fluctuations, the run-to-run reproducibility was 4% RSD in migration time (n = 9). With temperature control, the migration time reproducibility was 1.2% RSD (n = 4), see Table IIIA. Note also that the maximum number of bases read was always in

(42) Bode, H. J. *Anal. Biochem.* **1977**, *83*, 364–371.
(43) deGennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(44) Bae, Y. C.; Soane, D. *J. Chromatogr.*, in press.
(45) Smirnov, I.; Foret, F.; Ruiz-Martinez, M.; Berka, J.; Karger, B. L., manuscript in preparation.
(46) Wicar, S.; Vilenchik, M.; Belenkii, A.; Cohen, A. S.; Karger, B. L. *J. Microcol. Sep.* **1992**, *4*, 339–348.

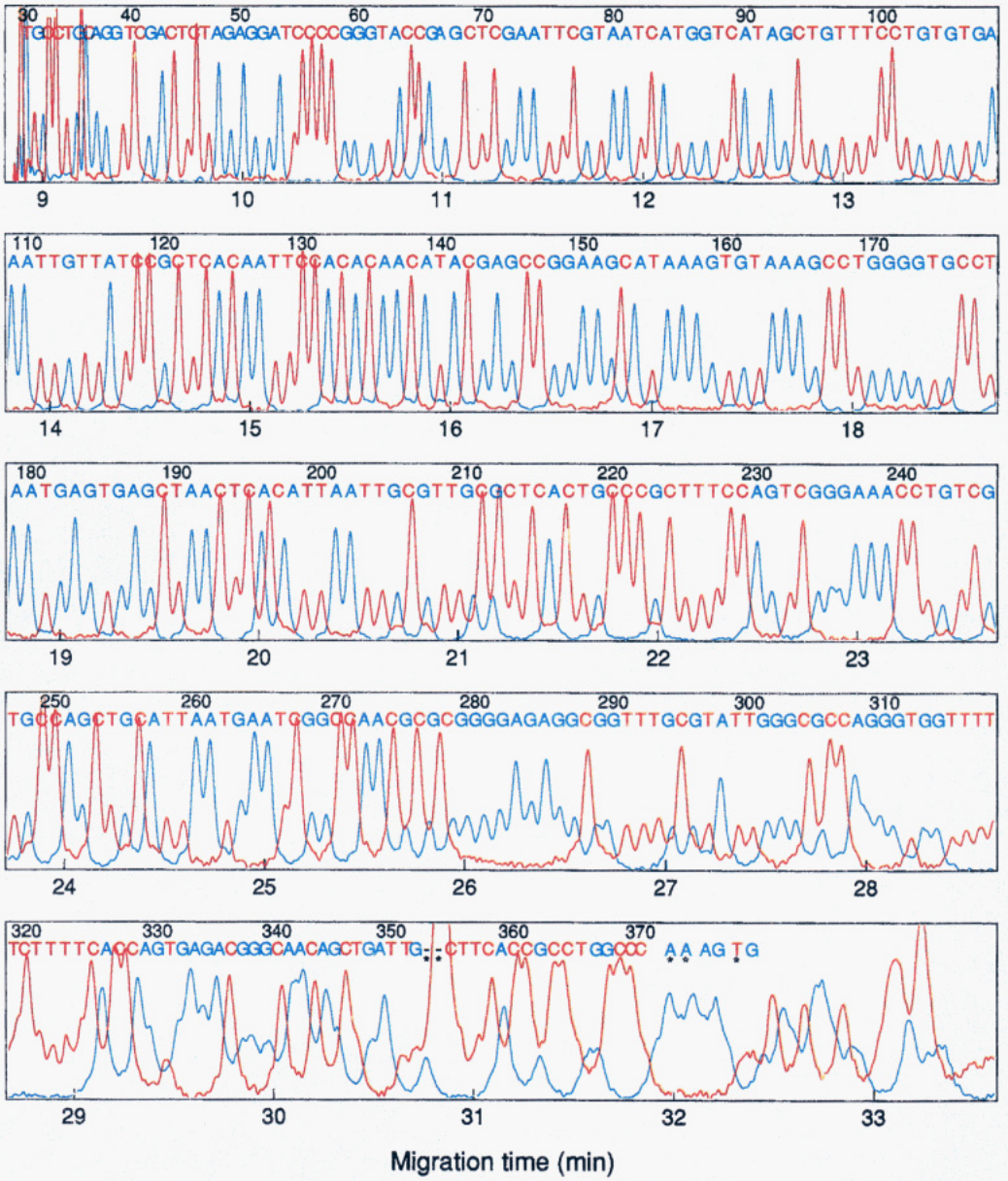

Figure 5. Sequencing of M13mp18. Red, JOE C > T; blue, FAM A > G. Electrophoretic conditions similar to those in Figure 3, except column temperature = 32 °C and the 6% T matrix was polymerized at 10 °C with 0.1% TEMED, 0.02% APS. Asterisks represent software miscalls.

excess of 350. After each run, the matrix was replaced in the column. We next polymerized eight different batches of 6% T LPA, and the reproducibility in migration time from batch to batch with the same capillary was 2.4% RSD Table IIIB. Again, all batches permitted sequencing out to at least 350 bases.

It should be noted that in Table III, the reproducibilities were run on a single-coated capillary. Whenever the capillary was changed due to coating degradation, the variation in migration time from capillary to capillary was approximately 5% RSD. However, the separation was maintained, and the sequence could always be read out to 350 bases or more.

CONCLUSIONS

A simple approach for DNA sequencing by high-performance capillary electrophoresis using a replaceable linear polyacrylamide matrix with a one-laser, two-photomultiplier system and peak height base encoding has been demonstrated. The linear polyacrylamide composition and the polymerization procedure were optimized to provide highly reproducible separation of the single-stranded fragments. DNA sequencing information was accurately obtained for 350 bases in roughly 30 min. The facile replacement of linear polyacrylamide from the capillary column overcame problems associated with the stability of capillary cross-linked gels.

The protocol of this work is applicable to recently described DNA sequencing techniques, for example, sequencing methods based on subcloning of DNA fragments to phage or plasmid vectors for which dye-labeled primers are available. Single-stranded DNA templates produced by asymmetric PCR amplification can be used as well, permitting, for instance, the sequence analysis of eucaryotic exons with average length of about 150 nucleotides for mutation mapping. A recently reported novel technique called partial-digest sequencing[47] that permits DNA of 4–6 kb in length to be sequenced by 200–300-bp-long sequencing steps without subcloning or the preparation of oligonucleotide primers is another promising technique in which the developed methodology can be directly applied. For all these methods, the instrument and matrix preparation are sufficiently straightforward that the procedure could be easily adopted in a laboratory setting.

(47) Wright, W. E. *BioTechniques* **1992**, *13*, 772–779.

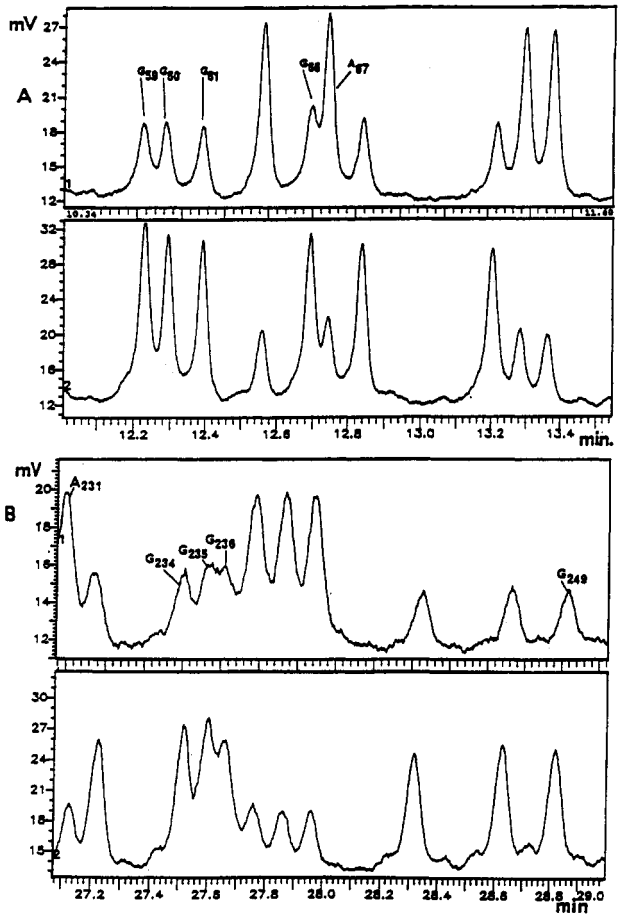

Figure 6. Two electrophoretic regions of the M13mp18, FAM G > A and A > G complementary sequencing reactions with inverted peak ratios. The separation of bases G59–A74 and A230–G248 are presented. Note the partial compressions of GG 59–60 of the GGG 59–61 triplet, of G66/A67, and G234/G235 of the GGG 234–236 triplet. Conditions: same as in Figure 3 except LPA polymerized at 0 °C.

Table III. Reproducibility for the Same Batch of Linear Polyacrylamide and from Batch to Batch

	A. For the Same Batch[a]			
run no.	RS (136/137)	t (136) (min)	t (299) (min)	maximum no. of bases read
1	1.10	16.6	27.6	369
2	0.96	16.2	26.9	364
3	0.98	15.9	27.0	359
4	0.94	16.0	27.0	355
	B. From Batch to Batch[b]			
batch no.	Rs (136/137)	t (136) (min)	t (299) (min)	maximum no. of bases read
1	1.26	16.1	27.6	370
2	1.25	16.2	27.4	369
3	1.24	16.1	27.5	360
4	0.97	17.4	28.9	357
5	1.10	16.6	27.6	369
6	0.96	16.2	26.9	364
7	0.98	15.9	27.0	359
8	0.99	16.8	27.0	367

[a] RSD of the migration time of base 299 = 1.2%. Rs, resolution; t, migration time. [b] RSD of the migration time of base 299 = 2.4%.

Current work involves extending the base number for sequence reading in excess of 500 by further optimization of the replaceable polymer network. For these longer base sequencing runs, we are utilizing fluorescently labeled terminators and an intensified diode array detector for high spectral resolution.[35] Such an approach should be fully compatible with recent primer walking strategies.[48] Results from these studies will be presented at a later date.

ACKNOWLEDGMENT

The authors thank NIH under the Human Genome Project Grant HG 00023 for support of this work. The authors wish also to acknowledge Dr. Dieter Schmalzing for discussions concerning polymerization optimization and Steve Carson for assistance in construction of the LIF detection instrument. This work is contribution no. 557 from the Barnett Institute.

APPENDIX

Sequencing Program. The program was written with the commercial programming package MATLAB (The MathWorks, Inc., Natick, MA). The software first converted the data to MATLAB binary format. The next step was spike removal. Spikes were detected as regions where the first derivative of the signal exceeded 10 times the standard deviation of that derivative for the channel. In each such region, a 1-s section at the highest signal value was excised and replaced by a straight line across the resulting gap. After spikes were removed, noise was reduced by a low-pass Fourier filter.

An arbitrary point high on the trailing edge of the primer peaks was selected as the position to begin sequencing, and a base line was estimated for each channel. For this estimate, the lowest signal point in each 2-min section of the data for a channel was found. The base line was then set as the shortest curve connecting the starting point and the final time point that passed through the 2-min minima and underlay all the intervening data.[49]

After base line subtraction, the data consisted of a $2 \times n$ matrix $\mathbf{Y}$, each row containing all the n time points for one channel. The 30% cross-talk of JOE in the FAM channel was removed by solving the matrix equation $\mathbf{Y} = \mathbf{RC}$, where $\mathbf{R}$ was the 2×2 matrix containing the response of each dye in each channel and $\mathbf{C}$ was the $2 \times n$ solution in which each row held the isolated fluorescence of a single dye. Thresholds for peak slope and peak width were derived from the derivative of each row of $\mathbf{C}$. The slope threshold decreased in rough proportion to peak height as base number increased, while the width threshold increased. These thresholds were used to locate approximate peak positions, whose neighborhoods were then searched for precise positions of peak maxima and peak boundaries.[50] Peak area was calculated as the area bounded by the dye signal, the base line, and perpendiculars dropped to the base line at the peak boundaries.

Peaks for a given dye were classified on the basis of peak area. Let the base known to produce the taller peaks be A and the other base be B. Peaks corresponding to A were chosen by first dividing the electrophoretic time profile for the dye into 1-min sections and selecting the peak with greatest area in each. An estimate of maximum peak versus time was constructed by interpolating linearly between values calculated for the selected peaks. Each such value was computed as the median of the areas of a selected peak and its two nearest selected neighbors on each side. Once maximum peak area at all times had been estimated, the ratio of peak area to that value was determined for all peaks. Peaks with ratios of 0.6 or greater were assigned to base A. These peaks were removed from consideration, and the process was repeated to assign the remaining peaks to B or to error. Peaks assigned to error were then deleted or merged with neighboring peaks.

(48) Kieleczawa, J.; Dunn, J. J.; Studier, F. W. *Science* **1992**, *258*, 1787–1791.
(49) Dyson, N. *Chromatographic Integration Methods*; The Bath Press: Bath, U.K., 1990.
(50) Pearson, W. R.; Miller, W. *Methods Enzymol.* **1992**, *210*, 575–601.

The entire classification process was repeated until no peaks were assigned to error.

The resulting sequence was aligned with the one from the inverse-labeling experiment, using the sequence comparison program FASTA.[51] Where the two sequences differed, the consensus base was taken as that from the run with the higher ratio labeling for the specific peak. An ASCII text report and a Postscript summary plot were then generated (Figure 5). Figure 5 further illustrates an option whereby the consensus sequence was compared with the known sequence, and miscalls were marked by asterisks.

Sequence ambiguities occurred at the approximate rate of 1% up to 350 bases; however, from the known sequence, these ambiguities were correctly resolved by the consensus procedure (i.e., the higher ratio assignment). Work is continuing on software improvements.

The present program is available on request or by anonymous ftp from ftp.ccs.neu.edu in directory /pub/miller. It requires MATLAB release 4.1 and the MATLAB toolboxes for signal processing and for optimization.

RECEIVED for review May 12, 1993. Accepted July 13, 1993.[⊗]

(51) Pearson, W. R. *Methods Enzymol.* **1990**, *183*, 63–98.

[⊗] Abstract published in *Advance ACS Abstracts*, September 1, 1993.