CHROM. 22 637

Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis

A. S. COHEN, D. R. NAJARIAN and B. L. KARGER* Barnett Institute, Northeastern University, Boston, MA 02115 (U.S.A.)

ABSTRACT

This paper demonstrates the potential of capillary gel electrophoresis with laser induced fluorescence detection as a tool for DNA sequence determination. Both synthetic oligonucleotides and single-stranded phage DNA were utilized as templates in the standard chain termination procedure. Primer molecules were tagged at the 5' end with the fluorescent dye, JOE. First, baseline resolution of a dA extended primer from 18 to 81 bases long, a total of 64 fragments, was observed. A second synthetic template was designed to yield alternating stretches of dA and dT extensions of the primer. Thirdly, the sequence reaction products from a synthetic oligonucleotide template containing all four bases was analyzed in four independent runs, one for each of the four base-specific reactions. In all cases, the expected number and patterns of peaks were observed by capillary gel electrophoretic analysis. Finally, separation of sequence reaction products generated with single-strand M13mp18 phage DNA as template exhibited baseline resolution of fragments differing in length by a single nucleotide and from 18 to greater than 330 bases total length.

INTRODUCTION

High performance capillary electrophoresis (HPCE) has become an important separation tool as a result of its high resolving power and speed¹⁻⁴. The technique has been applied in the open tube capillary format for separation of a wide variety of compounds including large biomolecules such as proteins and nucleic acids.

This laboratory has developed polyacrylamide gel-filled capillaries for the ultrahigh resolution of nucleic $acids^{5-7}$. Efficiencies for separation of single stranded fragments up to at least 160 bases in length were found to be as high as $30 \cdot 10^6$ plates/m, with separation complete in under 30 min. In addition, the columns were shown to be stable for well in excess of 150 injections⁶. Such performance makes these columns very attractive for DNA sequence analysis which relies on the separation and identification of fragments having one common endpoint and differing in length by a single nucleotide. These fragments can be generated by base-specific chemical degradation⁸ or, more commonly, by the chain-termination method⁹. In either case, correlation of the size of each fragment with a base-specific reaction allows the order of the nucleotides to be determined.

At present, separation of the DNA sequencing fragments is conventionally achieved by electrophoresis on high-resolution denaturing polyacrylamide slab gels¹⁰. The migration rate of a single-stranded DNA molecule is primarily dependent on chain length in these systems. It is routine to obtain sequence information for fragments greater than 400 bases in length in a single slab gel run. In conventional sequencing, radioisotopes are incorporated into the analyte during chain elongation, and detection is accomplished by autoradiography. In addition, automated systems using either radioisotopic¹¹ or laser-induced fluorescence detection^{12–16} have been developed to detect the products of Sanger sequencing reactions⁹ on slab gels.

DNA sequencing methodologies are currently under intense evaluation as a result of the Human Genome Project¹⁷. The goal of this program is to sequence the haploid human genome within 15 years. This effort involves the determination of roughly $3 \cdot 10^9$ base pairs, and with present slab gel technology, 100 of the most advanced automated slab gel systems operating for 60 years would be required¹⁸. Clearly, methodologies for much higher throughput are necessary.

Relative to slab gel operation, polyacrylamide gel HPCE offers more rapid separation, higher resolving power, and with appropriate instruments, subattomole level of detection. Ultimately, with columns operating in a parallel processing mode, throughputs as high as $1 \cdot 10^5$ base pairs per hour may be envisioned [*i.e.* 100 columns \times 1000 (bases/h/column]. With such high throughputs, costs will easily be contained at levels well below the current US\$ 2–5 per base¹⁷. The first step in transforming this potential into reality is the development of gel HPCE for separation and detection of DNA sequence reaction products.

This paper demonstrates the use of gel-filled capillaries with laser-induced fluorescence for the separation and subattomole detection of DNA sequence reaction products. In addition, in approximately 1 h, baseline resolution of fragments greater than 330 bases long and differing in length by a single nucleotide can be achieved. The columns are useable for multiple injections. Other researchers are also investigating HPCE for the analysis of DNA sequence reaction products^{19,20}.

EXPERIMENTAL

Apparatus

The HPCE system used in this work was configured as described previously⁵ except that a 60 kV high voltage d.c. power supply (Model PS/MK30, Glassman, Whitehouse Station, NJ, U.S.A.) was used to generate the potential across the gelfilled capillary. The laser detection system employed is presented schematically in Fig. 1 and is similar to that described previously²¹. An argon ion laser (Model 532 AT, Omnichrom, Chino, CA, U.S.A.) mounted on a 4×6 ft. optical table (Model 10531/ 13825, Oriel, Stamford, CT, U.S.A.), was operated in the light-regulated mode at 0.03–0.05 W. The laser light was passed through a narrow band filter (Model D1-488, Corion, Holliston, MA, U.S.A.), directed by reflection using a beam steerer (Model M670, Newport) and focused into the capillary with a 25-mm focal length lens (Model KBX043, Newport). Fluorescence from the sample was collected with a 40 × microscope objective (Model M-Set, Newport) and passed through an interference filter

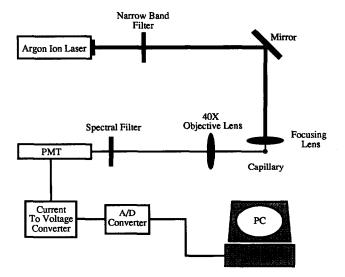


Fig. 1. Laser-induced fluorescence detection system for HPCE. The instrument consists of three major components: a detection system (laser, PMT = photomultiplier tube and accessories), a separation system (gel-filled capillary and power supply), and a data acquisition/analysis system (A/D interface and computer). See Experimental for complete details.

(Model S10-520-R, Corion) and a colored glass filter (Model OG520, Schott Glass Technol., Duryea, PA, U.S.A.). A photomultiplier tube (Model R928, Hamamatzu, San Jose, CA, U.S.A.) operated at 700 V and a photomultiplier readout (Model 7070, Oriel) were used to detect fluorescence. The resulting voltage output was displayed on a strip chart recorder and was simultaneously transmitted to an analog-to-digital (A/D) interface (Model 760 SB, Nelson Analytical, Cupertino, CA, U.S.A.) for transfer to a PC (Model ZBF-2526-EK, Zenith Data Systems).

Capillary columns

Polyacrylamide gel HPCE was performed in fused-silica tubing (Polymicro Technologies, Phoenix, AZ, U.S.A.), 75 μ m I.D., 375 μ m O.D., effective length (l) = 500–750 mm, total length (L) = 750–940 mm. Capillaries were prepared as described^{5,6}. Columns were first treated with methacryloxypropyltrimethoxysilane (Petrarch, Bristol, PA, U.S.A.). A solution of acrylamide (3% T/5% C)²² in 0.1 *M* Tris-borate, pH 7.6, 2.5 m*M* EDTA and 7 *M* urea was prepared, degassed, and introduced into the treated capillary column after adding ammonium persulfate and N,N,N',N'-tetramethylethylenediamine (TEMED).

Samples were injected into the column by dipping the cathodic end of the capillary into the sample solution and applying a voltage of 10 kV for 10–30 s. Separation was achieved at a typical applied field of 300 V/cm. Each column was used for multiple injections. Periodically, a short section of the capillary at the injection end was trimmed.

Sample preparation

Sequences for all oligonucleotides are given in the figure legends. All synthetic oligonucleotides except TEM80.1 and TEM80.2 were synthesized on a Cyclone DNA synthesizer (Milligen/Biosearch, Burlington, MA, U.S.A.). The two oligonucleotides, TEM80.1 and TEM80.2, were purchased from the Protein Chemistry Facility at Tufts Medical School, Boston, MA, U.S.A. Oligonucleotides to be fluorescently tagged were synthesized with a modified 5' end using N-TFA-C₆ Aminomodifier (CLONTECH, Labs, Palo Alto, CA, U.S.A.). Dye attachment reactions were performed with the fluorescein-based "JOE" dye purchased as the N-succinimide ester (Applied Biosystems, Foster City, CA, U.S.A.)²³. JOE-labelled DNA was purified by Sepharose G-25 chromatography followed by reversed-phase high-performance liquid chromatography²⁴. Single-stranded M13mp18 DNA was purchased from New England Biolabs, Beverly, MA, U.S.A. Deoxy- and dideoxynucleoside triphosphates were purchased from Pharmacia, Piscataway, NJ, U.S.A.; the Klenow fragment of DNA polymerase I and Sequenase 2.0 were obtained from U.S. Biochemical Corp., Cleveland, OH, U.S.A.

Sequence reactions performed with the Klenow enzyme contained 5 pmoles template, 20 pmoles primer and 60 units Klenow in 187 μ l of 10 mM Tris · HCl pH 7.5, 10 mM MgCl₂, 50 mM NaCl, 3.3 μ M dATP, 60 μ M ddATP and 33 μ M each of dCTP, dTTP and dGTP. After incubation at 37°C for 30 min, the reaction was stopped by heating to 65°C for 10 min. DNA was recovered by ethanol precipitation. For sequencing reactions performed with Sequenase 2.0, 4 pmoles each of template and primer in 25 μ l 10 mM Tris · HCl pH 7.5, 10 mM MgCl₂ and 50 mM NaCl were annealed by heating to 55°C for 2 min, followed by cooling to 37°C for 30 min. The mixture was adjusted to 37.5 μ l and made 200 μ M in dATP, dGTP, dCTP and dTTP, 7 mM in dithiothreitol and 5 μ M in the appropriate dideoxynucleotide. After addition of 7 units of Sequenase 2.0, the mixture was incubated at 37°C for 5 min, then stopped by heating to 65°C for 10 min. DNA was recovered by ethanol precipitation. All samples were stored in the dark at -20° C until use. Immediately prior to analysis the sample was resuspended in 2–5 μ l of 80% (v/v) formamide, 8 mM EDTA, heated to 90°C for 2 min and placed on ice.

RESULTS AND DISCUSSION

Our goal in pursuing this research was to demonstrate the feasibility of capillary gel electrophoresis as a tool in DNA sequencing. Our initial strategy was to examine products of a standard enzymatic DNA sequence reaction performed using synthetic oligonucleotides as both primer and template. This approach allowed us to predict exactly each product, while limiting the number and length of the resulting fragments.

TEM80.1 was sequenced using JOE-PRM18.1 as primer (see caption of Fig. 2 for oligonucleotide sequences). When annealed, these two oligonucleotides formed a double-stranded region of 17 base pairs. The Klenow fragment of DNA polymerase I was used to extend the primer in the presence of ddATP, and the electropherogram of the reaction products is presented in Fig. 2. Sixty-four baseline resolved peaks were observed in a time interval of less than 8 min, out of a total of 30 min run time. Based on the migration time separately determined for the primer, JOE-PRM18.1, the earlier

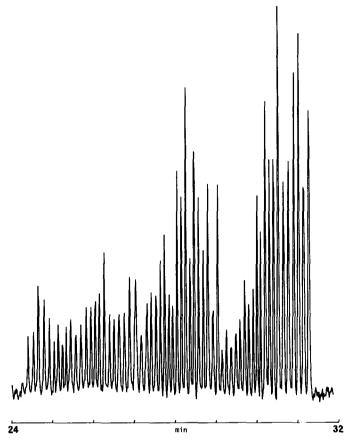


Fig. 2. Electropherogram of chain-termination sequencing reaction products. Template, TEM80.1 = 5'- $T_{63}ACAACGTCGTGACTGGGG-3'$. Primer JOE-PRM18.1 = 5'-JOE-TCCCAGTCACGACGTTGT-3'. Primer was extended by the Klenow fragment of DNA polymerase I in the presence of ddATP. Sequence reactions and electrophoretic conditions are described in the Experimental section. Running conditions: l = 540 mm, L = 680 mm, E = 300 V/cm, injection = 10 kV, 30 s, buffer = 0.1 *M* Tris-borate, 2.5 m*M* EDTA, 7 *M* urea, pH 7.6.

est peak in the electropherogram was the unextended primer. The other peaks represented primer extended by 1, 2, 3 ... 63 residues and terminated by ddA. Upon examination, the migration time for each fragment was found to be a linear function of length with a correlation coefficient of better than 0.999. We interpreted this to mean that the contribution of the JOE dye moiety to each fragment added a constant increment to the mobility of fragments in this size range.

It can be observed that peak heights in Fig. 1 varied by up to six-fold. This variation was a consequence of the enzymatic nature of the Klenow fragment in which the relative rates of incorporation of dideoxy- and deoxynucleotides are sequence dependent⁹. Sequenase 2.0 has been shown to be superior to the Klenow fragment²⁵ and was used in all other experiments described in this report.

We next increased the complexity of the sequencing sample by using TEM80.2

as template (see caption of Fig. 3 for sequence). Two independent reactions were performed, one in the presence of ddATP and the second in the presence of ddTTP. Fig. 3 presents the electropherograms for the two reactions. The reaction containing ddATP was expected to generate a set of fragments all of which begin with Joe-PRM18.1 primer and which end at any position corresponding to dT in the template. The predicted pattern for the 34 expected products of this reaction was: sextet (unextended primer plus 5 ddA-terminated fragments), space (no dT residues at these positions), quintet (5 ddA-terminated fragments), space, quintet, space ... triplet (final 3 ddA-terminated fragments). In an analogous manner, 31 peaks having a pattern of

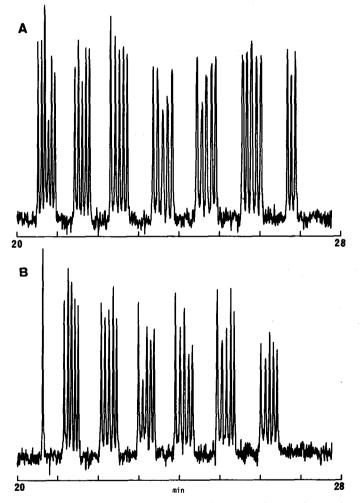


Fig. 3. Electropherogram of chain-termination sequencing reaction products. Template, TEM80.2 = 5'- $T_3A_5T_5A_5T$

singlet (unextended primer), space, quintet, space ... quintet were predicted for extensions in the presence of ddTTP. The electropherograms in Fig. 3A and B depict exactly the expected patterns with baseline resolution of all fragments and with a run time of less than 30 min. It can also be noted that peak height variation in Fig. 3 was significantly less than that in Fig. 2, a direct consequence of replacing the Klenow fragment with Sequenase 2.0.

With the successful correlation between predicted and observed peak patterns for the synthetic templates in Figs. 2 and 3, we next investigated a more realistic template, TEM80.3. This DNA represented the sequence of the polylinker cloning site of the phagemid pBluescript SK(+) from positions 585 to 664 and contained all four nucleotides (see caption of Fig. 4 for sequence). Four separate base-specific reactions were performed, and the products of each reaction were independently analyzed three times.

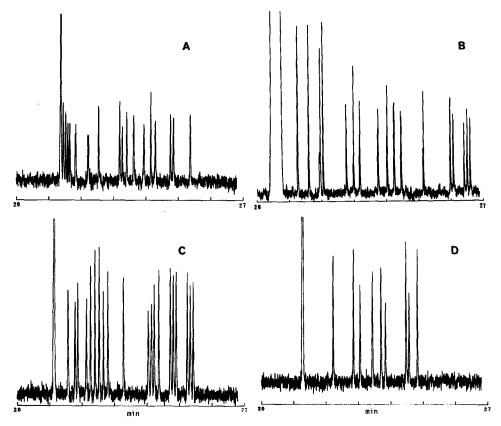
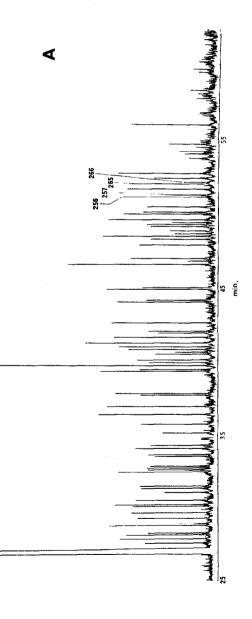
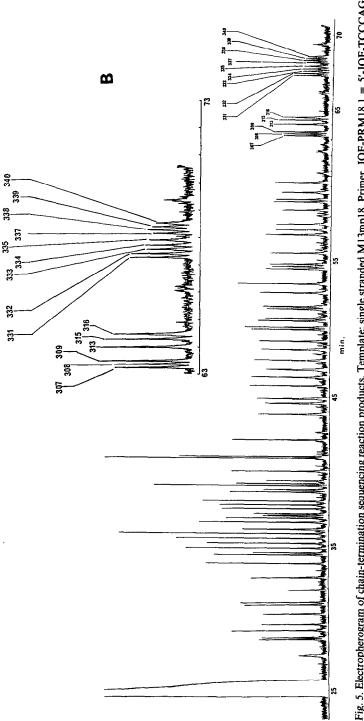


Fig. 4. Electropherogram of chain-termination sequencing reaction products. Template, TEM80.3 = 5'-GGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGT-CGTTTTACAACGTCGTGACTGGGG-3'. Primer, JOE-PRM18.1 = 5'-JOE-TCCCAGTCACGAC-GTTGT-3'. (A) dA reaction: primer was extended in the presence of ddATP. (B) dC reaction: primer was extended in the presence of ddGTP. (D) dT reaction: primer was extended in the presence of ddTTP. Sequencing reactions were performed using Sequenase 2.0. Sequence reaction and electrophoretic conditions are described in the Experimental section. Running conditions: see Fig. 2.







CACGACGTTGT-3'. (A) dC reaction: primer was extended by Sequenase 2.0 in the presence of ddCTP. (B) dT reaction: primer was extended by Sequenase 2.0 in the presence of ddTTP. Sequence reactions and electrophoretic conditions are described in the Experimental section. Running conditions: (A) l = 500 mm, L = 650Fig. 5. Electropherogram of chain-termination sequencing reaction products. Template: single stranded M13mp18. Primer, JOE-PRM18.1 = 5'-JOE-TCCCAGTmm, E = 350 V/cm, buffer = 0.1 M Tris-borate, 2.5 mM EDTA, 7 M urea, pH = 8.0; (B) l = 750 mm, L = 920 mm, E = 310 V/cm, pH = 8.0, injection = 10 kV, 15 s.

The data from representative runs of each sample are shown in Fig. 4A–D. The number and pattern of peaks in each electropherogram were as predicted. For example, the reaction in the presence of ddATP, Fig. 4A, was expected to produce five peaks at the earliest migration time: the unextended primer plus primer extended by 1, 2, 3 and 4 residues and terminated with ddA. Second, for the reaction containing ddCTP (complementary to dG, Fig. 4B), it was predicted that a doublet, space, triplet would occur at the longest migration time. Third, in Fig. 4C, the expected pattern –triplet, singlet, triplet, triplet– was observed at the end of the run for the reaction containing ddGTP. Finally, in Fig. 4D, the expected total of 10 peaks was found, 9 from the reaction with ddTTP and one from the unextended primer. In all four electropherograms, baseline resolution was observed even when fragments differed in length by a single nucleotide.

A preliminary investigation of proper peak alignment in Fig. 4A–D was undertaken. In this work, relative migration, using the primer as internal standard, was determined for all 63 bands. The relative migration of the bands occurred in the proper sequence order; however, in some cases the experimental error of 0.2% R.S.D. was greater than the increment from one base to the next. With proper care, nevertheless, this result suggests that four separate runs might be utilized for sequence determination. However, a better approach is to employ one column run with four different fluorescently tagged primers, as currently used in slab gels¹⁸. This strategy for capillary gel electrophoresis is under study in our laboratory.

M13 DNA sequence reactions

To evaluate the performance of gel-filled capillaries for separation of longer sequencing fragments, single-stranded M13mp18 DNA was utilized as template. JOE-PRM18.1 was again used as primer, and reaction conditions were identical to those used when synthetic templates were employed. The electropherogram of the reaction products obtained in the presence of ddCTP is displayed in Fig. 5A. In 1 h, a large number of peaks were resolved which were identified based on the known sequence of M13mp18. It can be observed that baseline resolution of fragments 256–257 and 265–266 was readily achieved. Clearly, gel-filled capillary columns provide sufficient resolution to separate fragments as large as 300 nucleotides and differing in length only by one residue. Signal intensity decreased rapidly for fragments longer than 300 bases. This effect was probably a consequence of reaction conditions and sample matrix. The exact causes are currently under investigation.

The electropherogram of reaction products obtained in the presence of ddTTP is displayed in Fig. 5B. Separation of fragments up to 340 bases long was achieved in approximately 1 h with high resolution. Identification of peaks was again made from the known sequence of M13mp18. The clusters of fragments from 331–355 and 337–340 bases long, shown in the inset of Fig. 5B, demonstrate clearly the resolution of large fragments which differ in length by a single nucleotide. In addition, the temporal distance between peaks 335 and 337 was sufficient to accommodate a fragment 336 bases long which was not a predicted product of the ddTTP reaction. Also, a peak corresponding to a fragment 328 bases long was predicted but not observed. This was most likely due to a loss in signal intensity specific for this fragment coupled to the decreased signal generally seen for larger fragments in both electropherograms of Fig. 5. When migration time is examined with respect to fragment length, a linear relation-

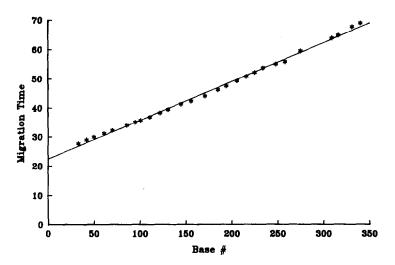


Fig. 6. Plot of migration time vs. fragment length. Values for migration time were taken from the electropherogram presented in Fig. 5B (dT reaction for M13mp18). Migration time of peaks representing ddTterminated fragments approximately 10n bases long (where n is an integer from 2 to 36) were selected. Peaks were assigned fragment lengths based on the known sequence of M13mp18.

ship (correlation coefficient = 0.997) is observed (Fig. 6). This linear behavior is an important feature for DNA sequencing. It is important to recognize that no peak compression⁹ has been observed in the samples we have examined. Nevertheless, further studies of this point are necessary.

Detection limits

A preliminary assessment of detection limits was made in this work. We first determined the detection level in the open tube by examining the signal-to-noise ratio of various concentrations of fluorescein flowing by the detector. A 10^{-11} M solution yielded a signal-to-noise ratio greater than 5:1, using the apparatus described in Fig. 1. If we assume a detection cell volume of low pl, this signal represents approximately 0.001 attomole of mass determined at the detection zone. Next we examined the gel column. An aqueous solution of $1.5 \cdot 10^{-10}$ M JOE-PRM18.1 (10 kV, 2 s) yielded a signal-to-noise ratio of greater than 5:1. It is difficult to quantify the actual amount injected through the gel capillary, given such factors as sample focusing upon injection; however, a reasonable estimate would be 0.1 attomole or less, based on a comparison of the open tube to gel results.

CONCLUSIONS

The results of this work demonstrate the potential of polyacrylamide gel capillary electrophoresis and laser-induced fluorescence detection for DNA sequencing. The features of this approach include (1) very high resolution, (2) rapid separation, (3) high sensitivity, (4) multiple injections on a given column and (5) automation. Baseline resolution of fragments one base different in length is presented for fragments up to 340 bases in length, and detection limits are estimated to be subattomole. Current studies include extending resolution to well in excess of 300 bases, exploring reaction and sample handling conditions and developing a single column instrument for detection of four different fluorescent dyes in one separation.

Ultimately, single column operation could be quite satisfactory for general laboratory sequencing needs. However, in the case of the Human Genome Project, a bank of columns would be required, so that multiple sequences can be analyzed simultaneously. In such an approach, each gel capillary column would represent a single lane in the current automated approaches to slab gel electrophoresis.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Beckman Instruments for support of this work. The authors further acknowledge Dr. Robert Rush for assistance in purification procedures. This is Paper No. 416 from the Barnett Institute.

REFERENCES

- 1 J. W. Jorgenson and K. D. Lukacs, Science (Washington, D.C.), 222 (1983) 266.
- 2 A. G. Ewing, R. A. Wallingford and T. '1. Olefirowicz, Anal. Chem., 61 (1989) 292A.
- 3 M. J. Gordon, X. Huang, S. L. Pentoney and R. N. Zare, Science (Washington, D.C.), 242 (1988) 224.
- 4 B. L. Karger, A. S. Cohen and A. Guttman, J. Chromatogr., 492 (1989) 585.
- 5 A. S. Cohen, D. R. Najarian, A. Paulus, A. Guttman, J. A. Smith and B. L. Karger, Proc. Natl. Acad. Sci. U.S.A., 85 (1988) 9660.
- A. Guttman, A. S. Cohen, D. N. Heiger and B. L. Karger, Anal. Chem., 62 (1990) 137.
- 7 A. Guttman, A. S. Cohen, A. Paulus, B. L. Karger, H. Rodriguez and W. S. Hancock, in C. Shaefer-Nielsen (Editor), *Electrophoresis* '88, VCH, New York, 1988, p. 151.
- 8 A. Maxam and W. Gilbert, Proc. Natl. Acad. Sci. U.S.A., 74 (1977) 560.
- 9 F. Sanger, S. Nicklen and A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A., 74 (1977) 5463.
- 10 Current Protocols in Molecular Biology, Green Publishing and Wiley-Interscience, New York, 1988.
- 11 Acugen 402 Sequencer, EG & G Biomolecular, Natick, MA.
- 12 L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent and L. E. Hood, *Nature (London)*, 321 (1986) 674.
- 13 J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jensen and K. Baumeister, Science (Washington, D.C.), 238 (1987) 336.
- 14 W. Ansorge, B. Sproat, J. Stegemann, C. Schwager and M. Zenke, Nucl. Acids Res., 15 (1987) 4593.
- 15 H. Kambara, T. Nishikawa, Y. Katayama and T. Yamaguchi, *Bio/Technology*, 6 (1988) 816.
- 16 J. A. Brumbaugh, L. R. Middendorf, D. L. Grone and J. L. Ruth, Proc. Natl. Acad. Sci. U.S.A., 85 (1988) 5610.
- 17 Understanding Our Genetic Inheritance: The U.S. Human Genome Project. The First Five Years, 1991– 1995, Publication DOE/ER-0452P, U.S. Department of Health and Human Services and U.S. Department of Energy.
- 18 G. L. Trainor, Anal. Chem., 62 (1990) 418.
- 19 H. Drossman, J. A. Luckey, A. J. Kostichka, J. D'Cunha and L. M. Smith, Anal. Chem., 62 (1990) 900.
- 20 H. Swerdlow and R. Gesteland, Nucl. Acids Res., 18 (1990) 1415.
- 21 W. G. Kuhr and E. S. Yeung, Anal. Chem., 60 (1988) 2642.
- 22 B. G. Johansson and S. Hjerten, Anal. Biochem., 59 (1974) 200.
- 23 DNA Modification Reagents for Use in Automated DNA Synthesis: A User's Manual, CLONTECH Laboratories, Palo Alto, CA.
- 24 Synthesis of Fluorescent Dye-labelled Oligonucleotides for Use as Primers in Fluorescent-based DNA Sequencing, User Bulletin 11, Applied Biosystems, Foster City, CA.
- 25 S. Tabor and C. C. Richardson, Proc. Natl. Acad. Sci. U.S.A., 84 (1987) 4767.
- 26 D. Heiger, A. S. Cohen and B. L. Karger, unpublished results.