TECHNICAL PROGRESS REPORT
Arrayed Capillary Electrophoresis:
High Throughput Separation and Identification of DNA Sequencing Fragments

Norman J. Dovichi
Department of Chemistry
University of Alberta

Funding for this renewal was received on February 1, 1994. Since that time, we have published 19 papers and book chapters plus one patent issued and two patents pending on our work in DNA sequencing (1-22). This work has focused on several areas.

First, we have developed a number of instruments for separation and detection of DNA sequencing fragments. These instrumentation include 5, 16, 32, and 96 capillary sequencers. All instruments are designed for operation with ABD's four-color sequencing protocol.

Second, we have developed a system to refill the capillaries with fresh matrix as needed. This project is in collaboration with Lawrence Berkeley National Laboratory.

Third, we have developed systems for injection of large numbers of samples, again in collaboration with LBNL.

Fourth, we have performed a number of fundamental studies on DNA separations at high electric fields by capillary electrophoresis. These studies have focused on the use of non-crosslinked polyacrylamide, a matrix with low viscosity that can be easily replaced within the capillary.

Fifth, we have evaluated the behavior of DNA separations as a function of temperature.

Sixth, we have developed a very efficient sample loading protocol that injects up to 75% of the DNA in the sample onto the capillary.

Seventh, we have modified BASS, a base calling program from the Whitehead Institute, for use with the capillary data.

Finally, we have begun to evaluate the instrumentation for realistic sequencing projects.

We have generated a huge amount of data. For brevity, we only focus on several issues that are germane to the renewal.

**MULTIPLE CAPILLARY DETECTION-LINEAR ARRAY**

We have developed a number of multiple capillary instruments, based on a detector constructed from a highly modified sheath flow cuvette (18). These detectors provide efficient optical excitation and sensitive detection.

We have constructed 5, 16, and 32 capillary DNA sequencers based on a linear array of capillaries placed within a rectangular sheath flow cuvette, figure 1 (20). Sheath fluid passes through the interstitial spaces between the capillaries, drawing the sequencing samples as thin streams through the cuvette, with one sample stream per capillary. The sample streams have identical refractive index as the surrounding sheath buffer. As a result, there is no light scattered at the sample stream-sheath stream interface. A laser beam can pass through all of the sample streams without distortion or loss of intensity.

**Figure 1 - linear arrayed CE**
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

 Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
In our instruments, two low-power laser beams are sequentially focused into the cuvette. A beam excites fluorescence simultaneously from all the sample streams, figure 2. An argon ion laser beam at 488 nm excites FAM and JOE labeled fragments; while a green laser beam (from either a HeNe at 543.5 nm or a frequency doubled neodymium-YAG laser at 532 nm) excites TAMRA and ROX labeled fragments. A single microscope objective images the fluorescence from the illuminated sample streams onto either a CCD camera or a set of fiber-optic coupled photodiodes. The high optical quality of the flow chamber minimizes light scatter, which produces both low background signals and excellent detection limits.

Our 16 and 32 capillary devices are based on micromachined cuvettes, figure 3. These cuvettes, manufactured at the Alberta Microelectronics Center at the University of Alberta, are constructed from glass microscope slides. We machine a set of grooves or fingers into the top of the cuvette to hold the capillaries on precise centers. We include other features to hold the capillaries at a fixed distance into the cuvette.

Figure 4 presents a photograph of the 16 capillary cuvette in operation. An argon ion laser beam is used to excite fluorescence in the cuvette; a yellow filter is used to block scattered laser light. The alignment fingers are not visible in this photograph. The system was deliberately misaligned, generating a large amount of light scatter that excites fluorescence from the polyimide coating on the capillaries. Raman scatter generated by the laser beam is visible as the horizontal line below the capillaries. The sheath stream draws a dilute solution of fluorescein from each capillary, generating one bright spot beneath each capillary. These spots are imaged onto the photodetectors.

The 5 and 16 capillary instruments use a microscope objective to collect fluorescence from the cuvette. A rotating filter wheel isolates fluorescence in four spectral channels. Fluorescence is then imaged onto a set of GRIN lenses, which couple fluorescence into fiber optics that are pigtailed to avalanche photodiodes. The five capillary instrument uses commercial single photon counting.
modules, while the 16 capillary instrument uses bare photodiodes and a locally constructed cooling chamber and detection electronics. The signal from each capillary is digitized and sent to a Macintosh computer for analysis.

The 32 capillary instrument uses a camera lens to collect fluorescence, a filter wheel to discriminate fluorescence into four spectral channels, a second camera lens to image fluorescence onto a cooled CCD camera to record an image of the entire capillary array. Based on the relative cost of avalanche photodiodes and a high quality CCD camera, it is cheaper to use discrete photodiodes for systems with less than ~25 capillaries and a CCD camera for larger systems.

**Thermal control**

All of our multiple capillary instruments are equipped with temperature control, which is provided by a World Precision Instruments (WPI) temperature controller. This inexpensive unit blows heated air over the capillaries. Unlike a proportionally controlled heater, this instrument is equipped with a thermostat, which causes regular pulses of hot air to pass over the capillaries. The background signal in the fluorescence detector is proportional to temperature, and a periodic fluctuation in the background signal can be observed with use of the WPI heater. Enclosing the capillaries in a Teflon tube provided sufficient heat capacity to eliminate the thermally-induced background fluctuations and generate a flat baseline.

**Multicapillary fluorescence detection limits**

We evaluated detection limits for the five-capillary detector with fluorescein in free-zone electrophoresis. Detection limits are 130 ± 30 molecules (2 × 10^{-13} M) injected onto the capillary. These detection limits are as good as the best single capillary system in our laboratory. These excellent detection limits reflect the good collection efficiency of our optical system, the low background signal generated in the sheath flow cuvette, and the high quantum efficiency and low noise of the avalanche photodiodes.

**Throughput**

The five capillary instrument has been in operation since the summer of 1994. The 16 capillary instrument has been in routine use since January of this year. They have been used for a number of applications, including evaluation of gel chemistry, development of improved gel chemistry, and sequencing of target DNA. In the last two years, we have generated over 4,000 electropherograms with the separation of nearly two million bases of raw sequence. The instruments are rugged and robust. We routinely run two separations per day on the 5 and 16 capillary instruments, and we have done three runs per day when the sample throughput warrants.

**Multiple Capillary Detection-Two-dimensional array**

**First generation system**

We have developed two generations of a two-dimensional capillary array detector. In the first system, figure 5, the capillaries terminate in a two-dimensional array in a square sheath flow cuvette. The capillaries are arranged so that the back row is located furthest downstream in the sheath flow cuvette. The sheath stream draws the sample from each capillary, generating a two-dimensional array of sample streams in the cuvette. Fluorescence is generated from all the capillaries simultaneously with an elliptically shaped laser beam, which passes immediately downstream from the

![Diagram of first two-dimensional array](image-url)
ends of the capillaries. By tilting the array at 45°, fluorescence from adjacent rows does not overlap when imaged onto the camera. The sheath fluid is removed from the sheath flow cuvette and sent to a waste container. The flow chamber is made of quartz and held with stainless steel plumbing. The steel plumbing is held at ground potential, ensuring that the sheath stream is also at ground potential.

We successfully demonstrated this detector with a 5 × 5 capillary array. However, it became obvious that there is a flaw in this design. There are no fixtures within the flow chamber that hold the capillaries on uniform centers; gaps form randomly between capillaries as the array relaxes to fit the dimensions of the flow chamber. These gaps produce localized regions of high sheath flow rate within the chamber, which tends to distort the sample-stream flow profile. As a result, the sample streams are distorted and have nonuniform spacing. The poor hydrodynamic properties of this first generation system lead to poor detection performance.

![Figure 6 - second generation two-dimensional array instrument](image)

Second generation system

A second generation two-dimensional array has been developed and applied to large-scale DNA sequencing. This system differs from the first generation instrument in four ways. First, the capillaries are held on uniform centers by use of a metal shim, which has an array of micromachined holes. The capillaries are threaded through the holes in the shim, creating a precisely aligned capillary array in the flow chamber. Second, the array terminates in a plane, with each capillary extending the same distance into the flow chamber. Third, a barrier shim is placed downstream from the capillary tips. The sheath stream surrounds the capillary tips and passes through the holes, producing uniformly spaced sample streams in the cuvette. Finally, fluorescence is detected along the capillary axis, rather than at an oblique angle as in the first generation two-dimensional array.

Figure 6 presents an overview of the instrument. Cylindrical lenses project an elliptically shaped laser beam in the cuvette. The beam skims between the shim and the capillaries, exciting fluorescence from all of the sample streams. Fluorescence is collected through the holes in the barrier shim with a camera lens, which forms a collimated beam. Fluorescence is detected in four spectral channels with a filter wheel. The transmitted light is imaged onto a cooled CCD camera.

The capillaries are held with a set of metal shims, machined with an array of holes, figure 7. The holes hold the capillaries on uniform centers. In our current design, a 24 × 24 array of 200-μm
diameter holes on 400-μm centers are electroformed in a thin stainless steel shim. We block the outer holes with tape, leaving the central 96 holes available for use. The 8 × 12 array of open holes is shown in gray in the figure. The cuvette can be used as a 576 capillary array by removing the tape and inserting more capillaries.

Four larger alignment holes are formed at the corners of the shim. These holes fit snugly over stainless steel rods, to hold the shims in precise alignment. The cuvette is assembled by placing three shims, spaced by 1-mm thick Teflon washers, over the stainless steel rods, figure 8. Capillaries are threaded through the holes in the shims, forming a precisely spaced two-dimensional array. A soft rubber insert is squeezed between the shims inside the Teflon washers; the rubber acts as a septum, forming a leak-tight seal around the capillaries.

A barrier shim is located 1-mm downstream from the capillary array. The holes in this shim match the capillary array spacing, with one hole per capillary.

Figure 9 describes the hydrodynamics of the cuvette. Sample migrates from the capillary into the fluid-filled chamber, which is held at ground potential. Sheath fluid draws the sample as a discrete stream from each capillary, creating a sample stream that is forced to flow through the matched hole in the barrier shim. A window is placed opposite the opaque metal shim.

Fluorescence, shown as black dots in figure 10, is excited by the laser beam and imaged through the holes in the shim. By aligning the laser beam so that it skims near the metal shim, we maximize the solid angle of fluorescence that is subtended by the holes in the shim. The opaque metal shim blocks scattered laser light,
preventing illumination of the chamber between the shim and the exit window. As a result, there is no background signal generated in this region. The second-generation two-dimensional arrayed capillary electrophoresis cuvette provides a very low background measurement of fluorescence.

In our system, fluorescence is collected and collimated by a 50-mm focal length camera lens. A filter wheel equipped with four spectral filters is used to spectrally resolve fluorescence from the four dyes. A second lens images the fluorescence onto a CCD camera. The image of the sample streams consist of an array of spots with 50-μm diameter and spaced on a 300-μm grid. This image is monitored as each filter swings into position.

The filter wheel turns at 0.5 Hz. The relatively low sampling frequency is limited by the camera readout speed. Our camera is not equipped with a frame transfer chip and the camera can not be illuminated during read-out of the image. A shutter blocks the image during read-out to prevent smearing of the image. We reduce the read-out time to 180 ms by only recording the central portion of the image. During read-out, the filter wheel position is incremented; after a slight delay to allow the wheel’s position to settle, a 250-ms exposure is taken. The shutter is closed, the image is digitized, and the filter is indexed to the next position. Each filter is exposed for only 250 ms out of the 2-second filter rotation period, corresponding to a 12% duty cycle.

Each image is software binned and decimated to an array of 96 points, one from each capillary, before storage in memory. The time-multiplexed spectral data are stored in an array, with one vector per capillary. The data files are surprisingly compact. A 2.5 hour run with 96 capillaries occupies only 4.4 megabytes of memory.

Figure 11 presents the four-color sequencing data generated by the 96 capillary instrument. Each column represents the fluorescence signal for each capillary. Red, yellow, green, and blue are used to denote the spectral channel with the highest intensity at each point in time. Figure 12 is a conventional presentation of the electropherogram from one of the capillaries.

**Figure 11 - four color sequencing data from 96 capillary instrument. Each column presents the time-resolved fluorescence signal from one capillary from 60 to 120 minutes into the run. Four colors are plotted, corresponding to the most intense spectral channel at that time.**

**Figure 12 - conventional presentation of an electropherogram from one capillary.**

### Multiple capillary refilling

We are collaborating with Joe Jaklevec’s group at Lawrence Berkeley National Laboratory to develop capillary refilling instrumentation. Two systems have been designed and constructed at Berkeley and evaluated by us. The first instrument is a high pressure chamber used for refilling 96 capillaries, figures 13 and 14. The massive bomb is about 6.5 inches in diameter and accommodates a microtiter plate. The bottom of the bomb is a hemisphere while the top is a flat plate with 96 holes that align with the wells in a microtiter plate. The capillaries are epoxied into the holes in the plate. They feed through the plate and mate with a microtiter plate, which holds polymer matrix for refilling. After the capillaries have been loaded with polymer, the bottom of the bomb is removed and the polymer reservoir can be replaced with a microtiter plate that contains 96 samples.

The first bomb was engineered for operation at 500 PSI but with a much higher safety margin, figure 15. This design results in a massive chamber that is cumbersome to use. Robust construction is required to support the total force across the 6.5-inch diameter plate at the top of the bomb; this force is about 7.5 tons at 500 PSI.
Electrophoresis Results Obtained from a 2-D Array of 96 Capillaries
M13mp18 Sequencing Performed on the 96-Capillary Instrument
operation, which represents a significant over-
design. We found that ~100 PSI of nitrogen is
sufficient to force 4% T 0% C polyacrylamide
through an array of 35-cm long, 50-μm inner
diameter fused silica capillaries in 15 minutes at
room temperature.

Based on our feedback, the Berkeley
group constructed a much more compact second
bomb, which is 3.5 inches in inner diameter.
This system is made from thin wall, high
strength stainless steel. The hemispherical
bottom weights one pound while the top is only
slightly heavier. This bomb was designed for the
16 capillary instrument. The lid is equipped
with 16 Valco compression fittings that are used
to hold capillaries. The Valco fittings were used
for easy of replacement of capillaries as we
evaluate different capillary coatings.

SAMPLE INJECTION

In most of our instruments, a dedicated electrode is associated with each capillary. The sample
reservoir is brought into contact with both the electrode and capillary, electric field is applied for a
few seconds, and the sample is replaced with running buffer before the electrophoretic potential is
applied. It is awkward to have a dedicated electrode for each sample in a large-scale instrument. The
electrodes tend to bend with time, and it is difficult to hold a large number of capillaries and elec-
trodes in close alignment for loading 96 small volume samples.

We have developed a very simple sample loading method for large numbers of samples.
Microtiter plates are coated with a 300 nm layer of gold at the Alberta Microelectronics Centre; this
very thin layer consumed roughly 30 mg of gold. Samples are placed in the wells and the plate is
placed in contact with the capillary array. An electrode is dipped into a buffer filled well on the
plate. We have reused the gold-plated microtiter plate with no sign of degradation or contamination.

Microtiter plates are awkward to use with the 16 capillary refilling system. The group at
LBNL has constructed several 16-well injectors for us. The gold-plated metal plate is machined with
16 funnel-shaped holes, which are 4-mm wide at the top, taper to 1.75 mm diameter, are 4-mm deep,
and have a volume of 10 μL. These plates can be reused without degradation or contamination.

FUNDAMENTAL STUDIES OF ELECTROPHORESIS SEPARATION OF SEQUENCING FRAGMENTS

Early reports of the use of non-crosslinked polyacrylamide for DNA sequencing described
separation of fragments 350 bases in length in 30 minutes. These reports were quite disappointing—
the short read length was not practical for most sequencing applications. We investigated the cause
of this short read length.

The resolution of DNA sequencing fragments degrades for longer fragments, which is associ-
ated with a decrease in peak spacing. A plot of migration time versus fragment length is sigmoidal,
reaching an asymptotic value for longer fragments. There are a number of models for this phenom-
emon. Biased reptation appears to be the most popular model, but it is only valid for longer frag-
ments. The model predicts that mobility, \( \mu \), scales inversely with fragment length

\[
\mu = \chi \left[ \frac{1}{N} + \frac{1}{N^{*}} \right] = \chi \left[ \frac{1}{N} + \left( \frac{\alpha E}{T} \right)^{\beta} \right]
\]
where $\chi$ is the free solution mobility of DNA, $N$ is fragment length, $N^*$ is the fragment length for the onset of biased reptation with orientation, $E$ is electric field, $T$ is absolute temperature, $\alpha$ is a collection of constants, and $\beta$ is an exponent that describes the electric field dependence of $N^*$. Fragments shorter than $N^*$ do not suffer appreciable crowding whereas fragments much longer than $N^*$ are very difficult to resolve. $N^*$ serves as an estimate of the longest possible read-length.

**The dependence of the onset of biased reptation with orientation on electric field**

Classic models of biased reptation state that $\beta = -2$; that is, sequence read length is inversely proportional to the square of electric field (23). This model is distressing, because it predicts that read length rapidly shrinks at high fields. A recent model predicts that $\beta = -1$, which implies that relatively long reads are possible at higher fields (24).

To test these two models, we separated DNA sequencing fragments in non-crosslinked polyacrylamide at electric fields ranging from 100 to 300 V cm$^{-1}$. The data were analyzed to determine $N^*$ at each electric field. A log-log plot of $N^*$ versus $E$ was linear (reduced $R^2 = 0.03$ for five degrees of freedom) with a slope of $-1.27 \pm 0.02$. We obtained similar results for crosslinked polymers. The slope of the log-log plot is equal to $\beta$, the exponent in the biased reptation equation. A value of $-1$ for $\beta$ implies that the onset of biased reptation is nearly inversely proportional to electric field.

The inverse relationship between read length and electric field allows the generation of very rapid sequence at high fields, albeit with short read length. As an extreme example, we have successfully sequenced DNA in crosslinked gels at an electric field of 1,200 V cm$^{-1}$, where fragments 250 bases in length are separated in four minutes.

More importantly, we published the first separation of long sequencing fragments by use of non-crosslinked polyacrylamide. We demonstrated that fragments over 550 bases in length could be separated in two hours at room temperature. These results were subsequently verified by Terabe (25). As we show below, much longer fragments can be separated with lower concentration, non-crosslinked polyacrylamide at elevated temperatures.

**Activation Energy of Electrophoresis Separation of Sequencing Fragments**

Most research groups in the capillary electrophoresis community have performed their separations at temperatures below 35 °C. We were presented with some clones from the *Pseudomonas* genome; this genome has roughly 70% GC content and generates particularly nasty compressions and abysmal quality sequencing data at low temperature. This result motivated us to investigate the effect of temperature on DNA separations in capillary electrophoresis. We have presented the first activation energy data for DNA separations with either crosslinked or non-crosslinked polymers (3, 16); this progress report only considers the results from the non-crosslinked polymer.

As expected, an increase in temperature results in relaxation of compressions. Also, as expected, an increase in temperature results in a decrease in separation speed. The mobility of short fragments increases roughly 2% per degree temperature rise. Most importantly, there is a dramatic improvement in resolution for longer sequencing fragments at higher temperatures with non-crosslinked polymers. This improvement is associated with an increase in $N^*$ with temperature.

Figure 15 is a plot of $N^*$ as a function of absolute temperature. The smooth curve is the least squares fit of a quadratic to the data. $N^*$ increases dramatically with temperature, which implies that longer sequencing fragments can be separated at higher temperature. This result is different from that observed for crosslinked gels, where there was a decrease in $N^*$ at elevated temperature (4). Note that the data of figure 16 are not consistent with equation 1; $N^*$ is proportional to $(T-290)^2$, not to $T^2$. Undoubtedly, this variation in $N^*$ with temperature reflects subtle changes in the polymer and DNA structure, rather than simply a variation in DNA conformation.
The separation of DNA fragments in electrophoresis is an activated process (25). DNA fragments must distort to pass through narrow obstructions in the polymer, and the polymer network must distort to allow passage of the DNA fragments. The mobility of a DNA fragment moving through the polymer can be described by an Arrhenius equation and the activation energy may be extracted from a plot of log mobility versus 1/T. The activation energy for DNA fragments in the non-crosslinked polymer drops with increasing fragment length, figure 16. This result is different from the behavior observed in crosslinked gels, where the activation energy increases with fragment length (4). This dramatic difference in activation energy for DNA separations in the crosslinked gel and non-crosslinked polymer must reflect fundamental differences in the behavior of both the DNA and the polymer during electrophoresis. The non-crosslinked polymer should be able to distort to allow passage of large DNA fragments. The distortion of the polymer is driven by the electrostatic force of the DNA fragment. Larger DNA fragments, which carry high charge, are driven more strongly by the electric field, which results in greater distortion of the non-crosslinked polymer. In contrast, a crosslinked gel will not be able to distort a significant amount without breaking the chemical bonds that crosslink the gel fibers. Presumably, a very high electric field would be required to provide enough energy to break the polymer bonds. DNA fragments must instead distort to pass through the pores, which requires more energy as the fragment length increases.

**Separation of very long sequencing fragments at elevated temperature**

The results from our thermodynamic study suggest that sequencing read length increases with temperature. We reported the first high temperature sequencing data with non-crosslinked polymers in capillary electrophoresis; we obtained sequencing read lengths of 640 bases with 5% T non-crosslinked polyacrylamide at 60 °C (12). These experimental conditions were quite exciting because they represent a significant improvement in speed and read-length compared with conventional capillary protocols.

We have improved our separation by use of lower concentration polyacrylamide and higher temperature separation. Figure 17 presents an electropherogram obtained at 70 °C. The M13mp18 sample was separated at 150 V/cm. The data have been corrected for baseline drift and for spectral overlap; no other processing has been done to the data. Close-up views are provided of the region for fragments 750 and 1,000 bases in length.

**Figure 17**—Separation of M13mp18 at 70 °C and 125 V/cm in 4% T 0% C polyacrylamide. First two pages are the color corrected data. Last two pages are close-ups of selected regions.
These results are quite satisfying. Until we began our investigations, there had been no reports of the separation of fragments longer than 350 bases in length with the use of non-crosslinked polyacrylamide. Careful attention to the fundamentals of the separation reveal several relationships between read length and experimental condition. To obtain long reads, it is necessary to use low concentration polyacrylamide, modest electric fields, and high temperatures. When the experimental conditions are properly tuned, exceptional resolution of kilobase fragments is possible in a few hour separation. Barry Karger has also recently reported read lengths of over 1,000 bases in less than two hours with 0% C polyacrylamide at elevated temperatures (26).

Efficient injection of sequencing fragments—pseudocoulometric loading

Most sequencing reactions generate several microliters of sequencing product. In conventional slab-gel electrophoresis, all of the sample is loaded in the well at the top of the gel. However, in capillary electrophoresis, the capillary is dipped into the sample. Electric field is applied briefly, drawing a small fraction of the DNA sample onto the capillary. After injection, the capillary is placed in running buffer, potential is reapplied, and the separation proceeds.

Only a small amount of DNA is loaded onto the capillary. However, a large amount of DNA is present in the sequencing sample. We estimate that roughly 1% of the DNA in the sequencing sample is actually loaded onto the capillary; the remainder is discarded. Clearly, there are important opportunities to reduce reagent consumption if sample loading were more efficient.

We have carefully studied the injection process. Passage of the electric field through the sample transfers a number of Coulombs of charge onto the capillary. The number of Coulombs transferred depends on the current and time associated with the injection. Essentially, injection is a Coulometric process. In conventional formamide-EDTA injection buffers, the majority of the charge loaded onto the capillary are EDTA. Only a small fraction of ions consists of DNA molecules, and sample loading is inefficient.

If DNA constituted a larger fraction of ions, then it should be possible to inject more DNA onto the capillary. Three steps are important in efficient loading of sequencing samples. First, it is necessary to eliminate the use of EDTA. Second, DNA must be carefully desalted. Third, a deionized loading solvent must be used. We have employed both deionized formamide and water. Replicate injections from the same sample demonstrated an exponential decrease in the DNA signal with injection number. This decrease in signal is due to the removal of DNA from the sample to the capillary. That is, the use of a low ionic strength sequencing buffer dramatically increases the fraction of DNA loaded onto the capillary.

We have successfully loaded over 75% of the DNA in a sample onto the capillary (14). This high efficiency loading is quite exciting, because it suggests that only a minute amount of DNA sequencing fragments need be prepared, which will result in a significant reduction in reagent cost for sample preparation.

However, any cost saving will be illusionary if expensive consumables are used to desalt the DNA sample. We have investigated three methods to desalt samples. In the simplest case, we perform a conventional ethanol precipitation with 95% ethanol, followed by three washes with 70% ethanol. In the second case, we pass the sequencing sample through a Microcon ultrafiltration column. In the third case, we use a homemade Sephadex G50 column. All three methods generate low ionic strength DNA samples that can be injected efficiently. While Microcon columns are quite expensive, the G50 columns are inexpensive to prepare.

Base calling software

Rather than creating a new base calling algorithm, we have modified a public domain program called BASS (pronounced ‘base’), available from the Whitehead Institute (27). Five major modifica-
tions are being made to the program. First, we removed the lane-finding routine from the program, which is not needed when dealing with capillary data. Second, we are modifying the program to reflect the high sampling rate employed with the capillary instrument. A number of filter parameters are written into the routine based on ABD data; we are modifying the code to adjust these parameters for the capillary data. Third, we are modifying the peak-spacing parameter in the routine. BASS assumes that the electrophoresis peaks are uniformly spaced during the run. However, the effects of biased reptation with orientation generate nonuniform peak spacing, with decreased spacing at the start and end of the run. The data are divided into eight windows, and the routine is applied to each window independently with an appropriate peak spacing. Fourth, we are modifying the routine to better account for mobility shifts. Fifth, we have modified the user interface.

We have not begun to optimize the speed of the program. It currently takes 1 minute on a SPARCstation 5 computer to process the data from a single capillary.

Figure 18 presents the BASS processed version of an M13 sequence, which was generated with the multiple capillary instrument. The first sheet presents the raw data, the second sheet presents the spectrally corrected data, and the last page shows the filtered sequence. The sequence is determined with reasonable accuracy—97% across 500 bases. However, it is clear that more work is required. The peak spacing and the color conversion algorithm need to be improved. Finally, a more efficient algorithm is needed to correct for mobility shifts between dyes.

**Sequencing Real DNA Inserts**

To date, virtually all capillary electrophoresis sequencing data has been generated with standard sequencing vectors, such as M13mp18 or pGEM. We have published two reports on the use of capillary electrophoresis to sequence real DNA inserts (1, 19) In the first paper, six clones taken from _Plasmodium falciparum_ (malaria) were sequenced, which generated a sequencing accuracy of at least 97.5%. Capillary gel electrophoresis, at 200 V/cm, showed a three-fold increase in speed compared to an automated slab-gel sequencer. In the second paper, another malaria clone was sequenced; over 570 bases of sequence were determined in the two hour separation. These clones are AT rich and provide beautiful sequence at 25 °C.

Buoyed by success with these templates, we obtained several clones from the _Pseudomonas_ genome from Roger Leveque at Laval University. These clones are GC rich and generate significant compressions when run at room temperature. We had pioneered the use of formamide as an additional denaturant in capillary gel electrophoresis (27-28); however, even the addition of 30% formamide to the separation matrix did not relax the compressions.

This experience motivated us to perform capillary electrophoresis at elevated temperatures. Before our publications, DNA sequencing had been done at 35 °C or cooler in capillary electrophoresis. There had been a fear of enhanced polymer degradation at higher temperatures. The rigid capillary walls were thought to resist expansion of crosslinked polymer during heating. Also, the decreased viscosity of the non-crosslinked material was expected to lead to loss of the material as residual electroosmosis pumps the polymer from the capillary. Fortunately, as demonstrated above, these fears were unfounded; we routinely obtain high quality sequence at 60 °C and we have obtained occasional good runs at 70 °C.

It is clear that further evaluation of the capillary system is warranted. We have begun a quick experiment to shot-gun sequence wild type M13. This circular bacteriophage is 6,407 bases in length and provides a convenient test for our sequencing protocol.

**Figure 18 - M13mp18 sequencing data. The first page is the raw data. The second page is the data, treated with BASS to correct for spectral overlap. Note that windowing results in discontinuous peak heights. The third page is the filtered and called data.**
Fragments were first ultrasonically sheered, then size-fractionated with agarose gel electrophoresis into ~1000 base fragments, and finally cloned into pUC19. Three different strategies were used to prepare sequencing template from these clones. The first strategy simply involves lysis of the pUC19 colonies. Primers are synthesized from both ends of the sequencing vector, and PCR is used to amplify across the insert. The PCR products are then cycle sequenced using Thermosequenase. While very simple, this strategy results in a fairly reproducible artifact. A large baseline disturbance is observed roughly halfway through the run. At normal operating conditions, the large disturbance swamps the sequencing signal. High temperature separations, ~65°C, relaxes this feature and allowed sequence to be read. However, high temperature operation of the capillary is not robust.

It appears that the sequencing artifact is associated with double stranded sequencing vector undergoing biased reptation in the polymer. The double stranded DNA has roughly twice the mobility of single stranded DNA, and it migrates with roughly twice the velocity of the single-stranded biased reptation peak. The high ionic strength associated with the double stranded DNA perturbs the conductivity in its vicinity, generating an electric field perturbation that distorts the migration of the sequencing fragments. Operation at elevated temperature denatures the template into single stranded DNA which migrate at the end of the run.

A biotin-labeled PCR primer can be used to generate single stranded template from the PCR products. One of the PCR primers is biotinilated. After PCR, the amplified insert is captured on streptavidin coated beads. The complementary strand is denatured and washed, leaving the captured single strand for use as a template in cycle sequencing. However, the cost and additional manipulation of the magnetic bead technology does not seem useful in large scale sequencing protocols.

In our preferred method, we pick clones, transfer them to a growing medium, incubate, and then isolate DNA using a modified miniprep protocol. Cycle sequencing is performed using Thermosequenase and dye labeled primers. As noted in the loading section, we use a size-exclusion column to desalt the samples before resuspension in formamide or deionized water for injection.

We have generated 59 electropherograms from 42 different clones of the wild-type M13 genome. Several runs were repeats from the same clone. Sequence read length of up to 626 bases at 97% accuracy have been obtained. We have obtained 75% coverage of the genome, with the largest gap associated with the origin of replication, which apparently does not produce viable clones in pUC19.

This exercise has proven to be quite enlightening. Sequence read length from double stranded template tends to be much shorter than with fragments obtained from single stranded M13mp18 template using a universal primer. Repeated sequence of DNA from the same clone generates similar electropherograms and similar read-length. Template quality is of highest importance in obtaining useful sequence data. However, the process has been invaluable in improving the sample preparation protocol for routine use of the instrument.

Finally, we have obtained several cosmids from the human genome from a colleague at the University of Calgary. We will begin a shot-gun sequencing project for these clones this month.
REFERENCES


18. "Capillary gel electrophoresis for large scale DNA sequencing: separation and detection", in *Automated Technologies for Genome Characterizations* Edited by Tony Beugelsdijk, in press.


