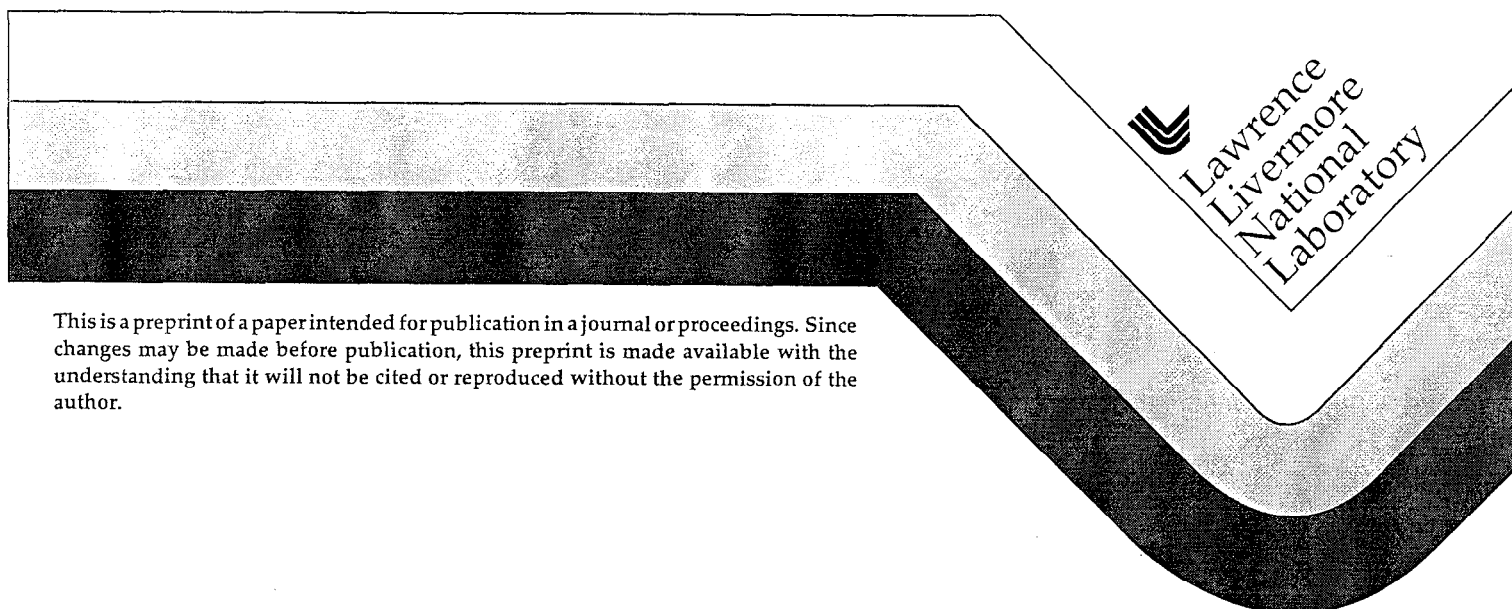


A Microchannel Electrophoresis DNA Sequencing System

J.P. Fitch, J.W. Balch
M.S. Bass, L.R. Brewer
A.C. Copeland, J.C. Davidson
L.M. Kegelmeyer, J.R. Kimbrough
R.S. Madabhushi, P.M. McCready
D.O. Nelson, R.L. Pastrone
P.M. Richardson, S.P. Swierkowski
L.A. Tarte, M. Vainer
T.E. Warth

This paper was prepared for and presented at the
33rd Annual Conference on Information Sciences and Systems
Baltimore, MD
March 17-19, 1999

February 1999



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes 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 the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

A Microchannel Electrophoresis DNA Sequencing System

J. P. Fitch, J. W. Balch, M. S. Bass^{SF}, L. R. Brewer, A. C. Copeland, J. C. Davidson, L. M. Kegelmeyer, J. R. Kimbrough, R. S. Madabhushi, P. M. McCready, D. O. Nelson, R. L. Pastrone, P. M. Richardson, S. P. Swierkowski, L. A. Tarte, M. Vainer, T. E. Warth^{SB}

The Department of Energy Joint Genome Institute

Lawrence Livermore National Laboratory, MS L-452, Livermore, CA, 94550

^{SF}University of California, San Francisco and ^{SB}University of California, Santa Barbara

Abstract

Last year, the Joint Genome Institute was the third leading sequencing group in the world with over 20 million bases of human DNA finished. The goal of the human genome project is to sequence the 3 billion bases of human DNA sequence by 2003. In order to increase the DNA sequencing throughput of the Joint Genome Institute, we have developed a microchannel electrophoresis system. One critical new and unique element of this system is a process for the production of 96 and 384 microchannels on bonded glass substrates up to 14 x 58 cm. In order to utilize these instruments for DNA production sequencing, we have been evaluating and implementing software to convert raw electropherograms into called DNA bases with an associated probability of error. Our original intent was to utilize the DNA base calling software known as Plan and Phred developed by the University of Washington. In our tests and evaluations of this software applied to microchannel data, we observed that the electropherograms are of a different statistical and underlying signal structure compared to slab gels. We have modified Plan and Phred to improve base calling performance for the microchannel data. In this paper, we will present 1) the microchannel DNA sequencing system and show the advantages compared to current slab gel and capillary systems. 2) The signal processing modules needed for DNA base calling including correction of multiple wavelength channels, signal averaging, non-uniform sampling, variable DNA mobility, and peak shape and spreading effects. 3) A comparison of the DNA base signatures in the raw data of microchannels vs. slab gels including some simple modeling results. This will be propagated through the base calling software to show the impact on DNA sequencing.

Keywords: DNA sequencing, electrophoresis, signal processing

1. Introduction

Electrophoresis may be defined as the differential drift velocity of electrically charged chemical ion or molecular species under the force of an electric

field in a fluidic separation medium. Pioneering work in 1937 by Tiselius on electrophoretic protein separation in large tubes resulted in a Nobel Prize. Polyacrylamide and agarose sieving media were introduced to reduce resolution losses from diffusion and convection. These gels are still used in the popular slab gel DNA sequencers. Large improvements in separation performance were introduced with very narrow (75 μm) bore fused silica capillaries in the 1980s. Early work with microchannels etched in flat glass substrates [1-3], versus drawn glass capillaries, indicated the potential of chemical separation done with the methodology of micromachining. Our channels are fabricated in two borosilicate [4] float glass substrates that are fusion bonded together into one microchannel plate. This plate may be cleaned or regenerated simply by pumping chemicals and sieving media into a single slot manifold port on the output end of the plate. The borosilicate float glass is rugged (both mechanical and thermal) and chemically resistant.

Compared to current slab gel systems, microchannel electrophoresis arrays share many of the benefits of capillary arrays such as: short separation times, no lane tracking required, sieving media pumped directly into the array, and resolution of over 500 DNA bases in less than 3 hours of run time. Furthermore, compared to capillary arrays our microchannel arrays have several additional positive attributes:

- The integrated array of microchannels in a single plate is very robust and dense.
- Only two components are fabricated and bonded to make a plate. Replicas can be made economically.
- Plate cleaning and sieving media fill and replenishment are done with a single O-ring sealed port.
- A planar optical window and precision placement of channels and fiducials enable better laser fluorescence detection.

- The channel shape is wide and shallow—optimizing signal detection and sieving media heat dissipation.
- Designs can be changed with CAD tools and the fabrication method can handle diverse structures such as single ended loads, cross-

channel loads, and fiducials or electrodes in the channels. Precise fiducials (1 μm) fabricated with the channels can also be used for alignment to the rest of the system.

Additional information on the 96-lane plates and plate fabrication are described by Swierkowski [5].

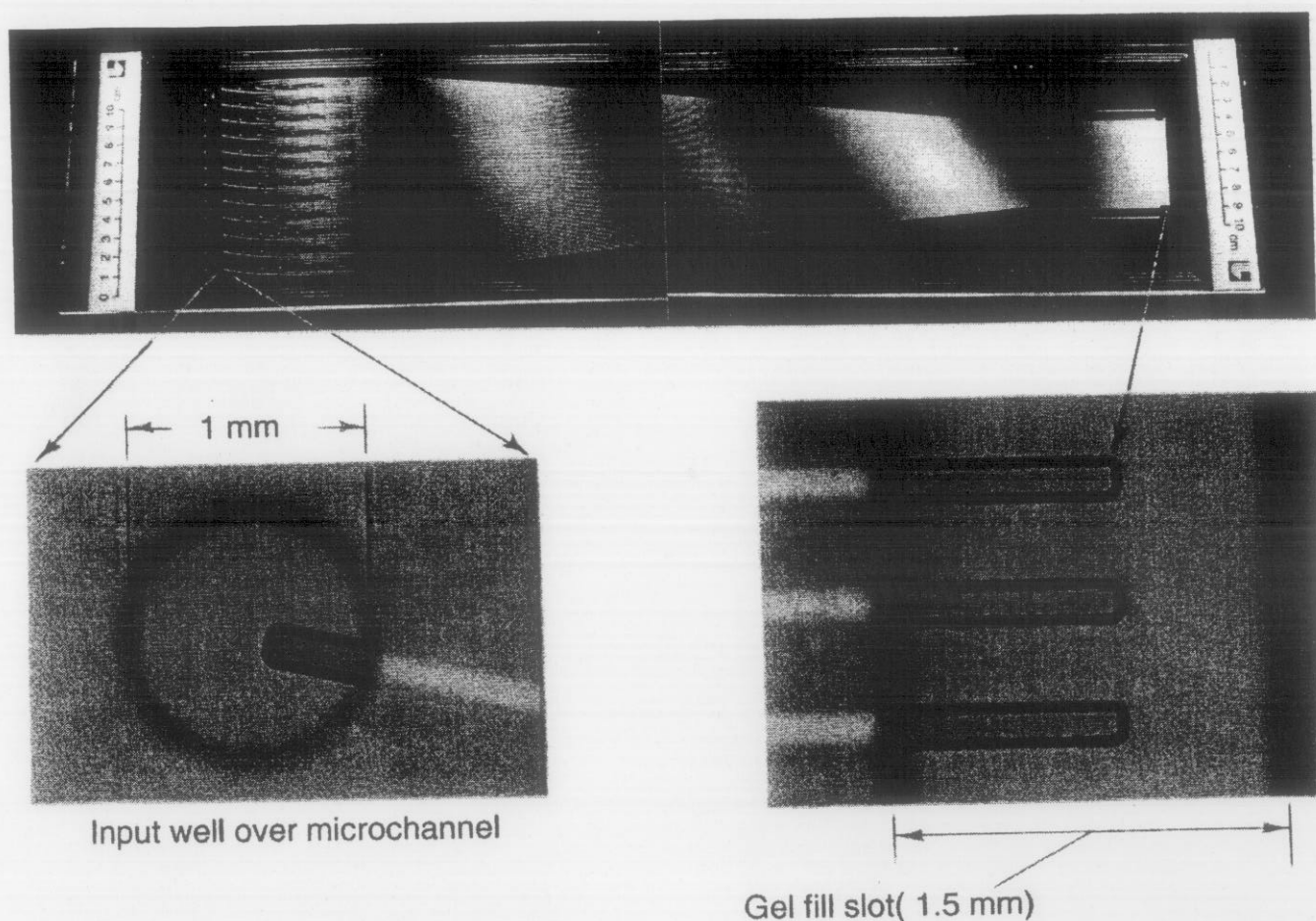


Figure 1. A 96-microchannel glass plate with microtiter input format on the left and the detection zone and fill slot on the right. The titer plate input port format makes each row of channels 9 mm different in length. The longest channels are 418 mm to the optical readout; the shortest ones are 63 mm shorter. The microchannels are about 45 x 170 μm in cross section.

2. Results

Microchannel plates containing 96 channels (see Figure 1) have been used extensively over the past six months. One plate has been used more than 130 times; generating nearly 500,000 high quality base calls from various test samples. Depending on the purity of the samples we load into the system, we can use the plate 15 to 30 times before performance begins to degrade due to fouling of the channel walls. When this happens, the channels can be easily regenerated *in situ* by flushing with solvent followed by an acid etch and water rinse.

Samples are introduced into the channels through ultrasonic milled holes in the top plate [6]. Top

plates are typically 1.1mm in thickness. A small volume syringe is used to transfer the 100-500 nL required for analysis. We typically use commercially available gel loaders that contain 8 or 12 syringes for loading the plate. The sample input holes are arranged in an 8 x 12 array on 9 mm centers corresponding to the industry standard microtiter plate. A 96-syringe device to load samples directly from a microtiter plate is a logical extension of our current technique. After loading the wells, DNA is then electrokinetically loaded onto the microchannels. Finally, the DNA drifts and separates under the influence of an applied electric field of about 100 - 150 V/cm. Typical run times are

180 minutes, which allow us to collect data from over 500 peaks per lane.

A lane map for a 96-microchannel plate is shown in Figure 2. The arrival time of the primer peaks shows a sawtooth pattern because of the different drift lengths, for the different input rows. The built in fiducial pattern is shown on the sides of the

figure as two constant lanes on the left and three constant lanes on the right; they are etched with the same pitch as the microchannels. A sample of a raw electropherogram from a lane in the middle of the plate (48 lanes in from the edge) is shown in Figures 3 and 4. One can clearly see the peak width and separation and good signal to noise ratio even before smoothing and filtering.

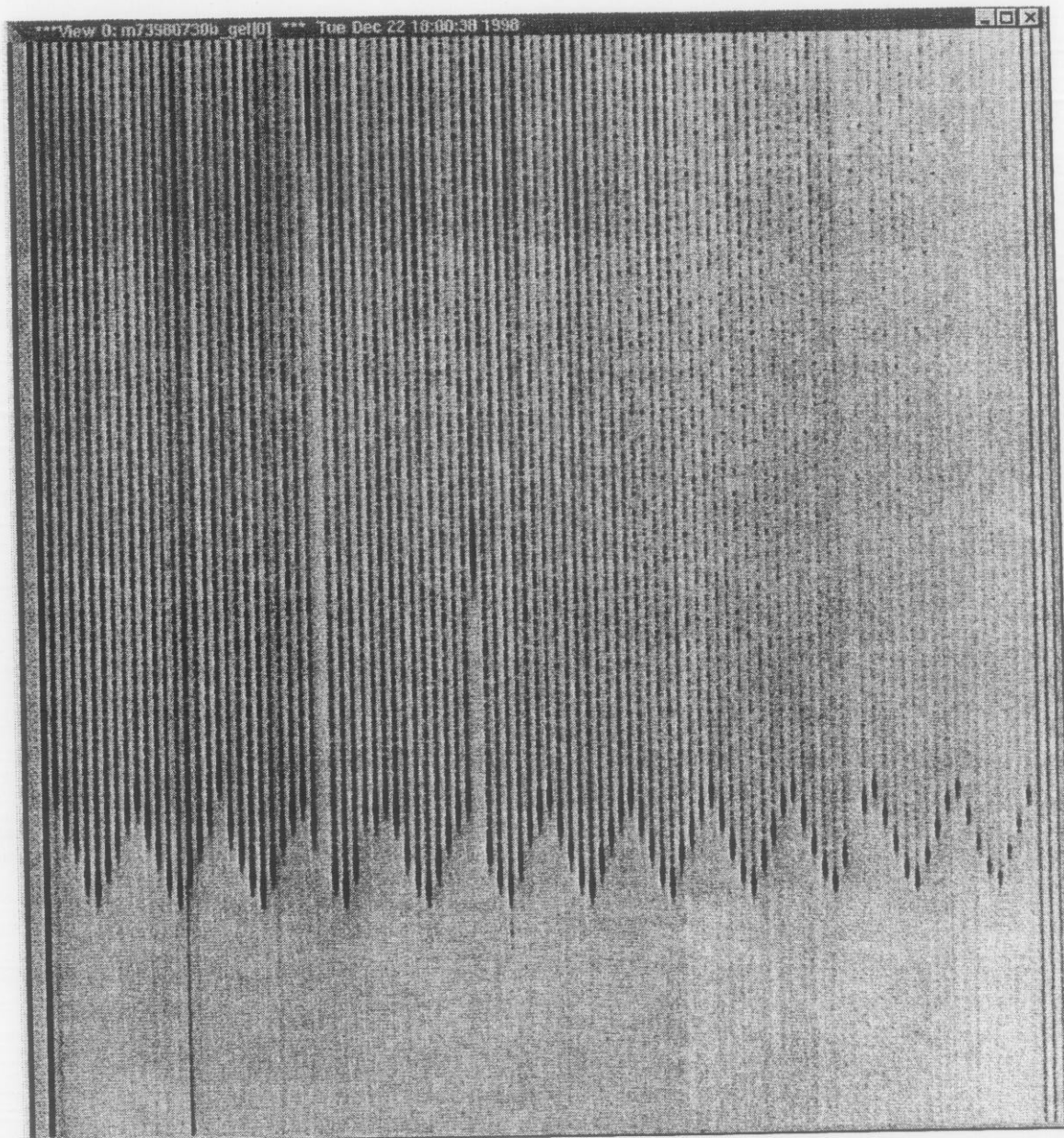


Figure 2. Lane map showing all 96 microchannel lanes for the blue dye that labels the DNA base C. Scanning across the plate every second, the displayed data is collected in 2.5 hours (time origin at bottom), with an applied voltage of 4.8 kV. The sample is an energy transfer dye labeled M13 sequencing standard. Note the uniformity of separation across the microchannels. Three lanes in this example show anomalous behavior due to accidental load variation but produced good data on subsequent runs.

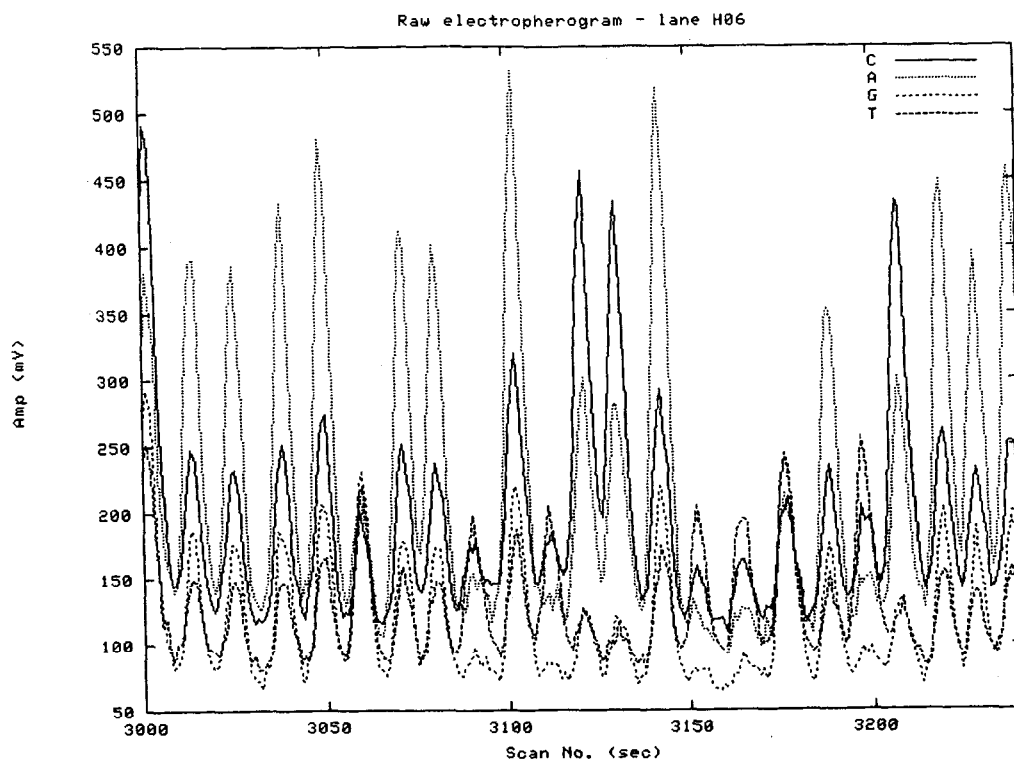


Figure 3. Raw data from lane H06 corresponding approximately to base numbers 50 through 75.

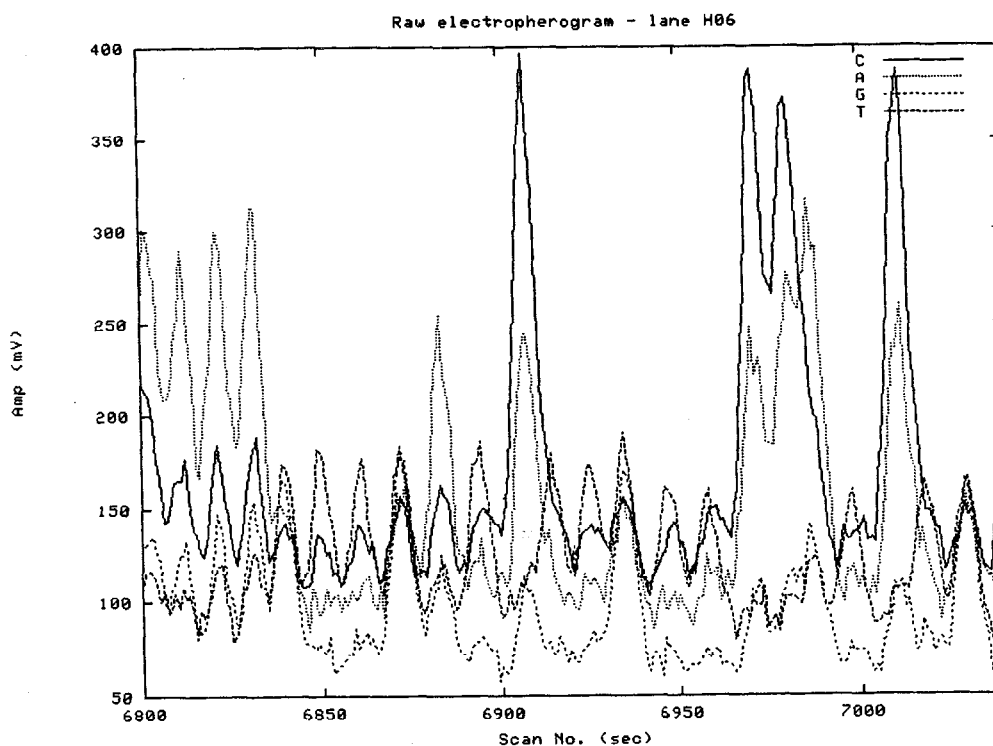


Figure 4. Raw data from lane H06 corresponding approximately to base numbers 400 to 425. This particular M13 standard was resolved to greater than 550 bases.

3. Signal Processing

Software modifications of Plan and Phred [7] focused on the fundamental differences between conventional instruments and the microchannel instrument:

- By replacing the baseline correction algorithms with a more adaptable and robust morphology-based algorithm, the noisy microchannel peaks retained their resolution, while losing most of the overlying high frequency noise. The

morphological approach is non-linear and allows dynamic shape-based computations throughout the data trace.

- We have developed tools to generate dye-molecule mobility correction to account for the unique media/dye combination and channels of different length. We are now able to generate customized compensation tables for various types of lanes and obtain improved performance. These were applied separately to long and short lanes, with performance results shown in the chart below.
- Because the microchannel plate has lanes that vary in length, the electric field and thus velocity of the specimen vary as a function of position along each channel. As a result, the

single look-up-table in plan/phred that corrects for non-uniform signal spacing along the length of a lane were not flexible enough to accommodate the range of lane lengths in the microchannel data, and we implemented a dynamic spacing correction for microchannel lanes.

- Lastly, we found that the plan/phred software does not trim the beginning of the signal adequately for microchannel data and so we trimmed the beginning of the sequences by about 100 bases before analysis.

The improvement realized by these software changes is illustrated in the performance chart in Figure 5.

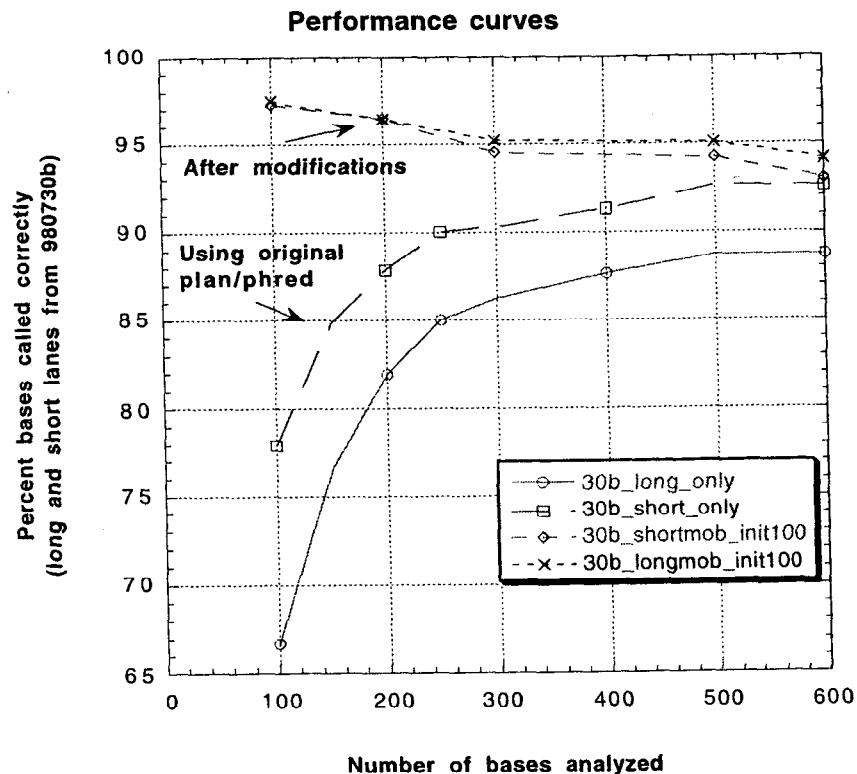


Figure 5. Performance improvement by modifying plan/phred software measured as an increase in the percent of correctly called bases from running known DNA samples in the microchannel sequencer. The 2 lower curves represent results from the original plan/phred software; the improved performance shown by the 2 upper curves, are the results of the software modifications described in this paper.

4. Conclusions

We have developed a microchannel-based electrophoresis system for high throughput DNA sequencing. High density microfluidic channels in borosilicate glass up to 58 cm in size that are reproducible, inexpensive, and physically robust have been fabricated. We have demonstrated DNA sequencing with up to 550 base resolution (using automated DNA base calling) in a 96-microchannel plate-based system that has convenient in situ

recycling and regeneration. Similar plates with 384 channels per plate have just been fabricated (Figure 6). Since the microchannel shape and size is very comparable to the 96-channel plate, these plates should provide comparable base-calling quality at four times the throughput. This type of high throughput system is needed for the JGI DNA sequencing goals.

On February 1, 1999, the Joint Genome Institute began a pre-production trial to evaluate

commercial and in-house electrophoresis instrumentation for DNA sequencing. The trial's first phase requires the instruments to sequence three microtiter plates (96-well) each day for two weeks. The two commercial instruments were not operational in time for the trial. When samples were available, the microchannel instrument successfully sequenced the 3 plates each day within an 8-hour shift. Run duration was held at 2 hours and operator hands-on time averaged 40 minutes for a typical run. Using our current conservative Phred scoring metric, we averaged 230 bases with a Phred score greater than 20 (roughly a 1 in 10,000 probability of error).

Acknowledgements

The authors acknowledge the support and funding provided by the NIH and DOE programs in human genomics as well as the LLNL LDRD project for DNA Base Calling.

References

- [1] Manz, et. al., *J. Chromatogr.* 593, p. 253, 1992.
- [2] D. J. Harrison, et. al., *Anal. Chem.*, 64, p. 1926, 1992.
- [3] D. J. Harrison, et. al., "Micromachining a Miniaturized Capillary Electrophoresis-Based Chemical Analysis System on a Chip," *Science* 261, pp. 895-897, August 13, 1993.
- [4] BOROFLOAT™, Schott Corporation, Technical Glass Division, 3 Odell Plaza, Yonkers, N.Y. 10701.
- [5] S. Swierkowski, et. al., "Large Microchannel Array Fabrication and Results for DNA Sequencing," *Proc. Of SPIE* Vol. 3606, January 25-26, 1999.
- [6] Ultrasonic milling and mandrel by Bullen Ultrasonics, Inc., 4613 Camden Road, Eaton, Ohio 45320.
- [7] B. Ewing, et. al., "Base-Calling of Automated Sequencer Traces Using Phred," *Genome Research* 8, pp. 175-185, 1998.

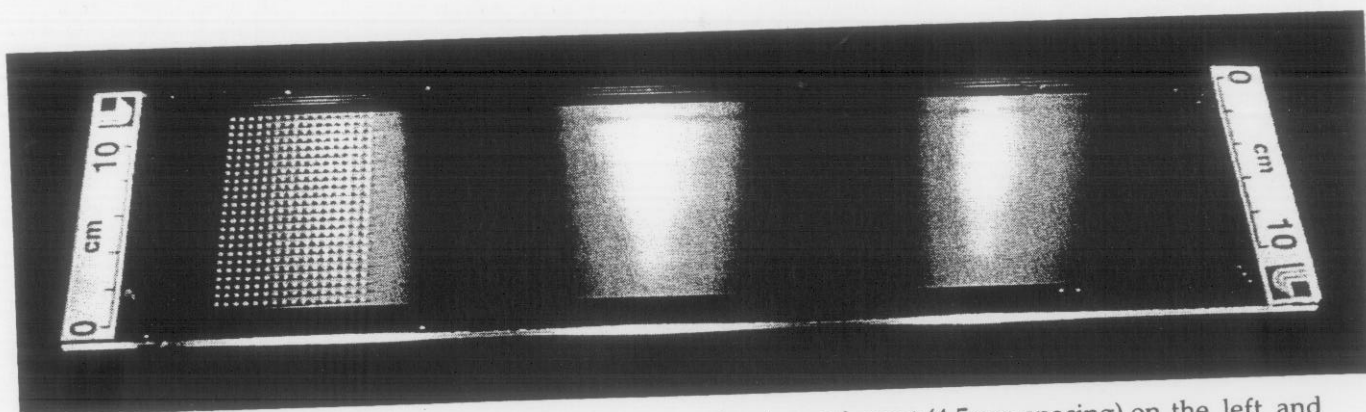


Figure 6. A 384-microchannel glass plate with microtiter input format (4.5mm spacing) on the left and the detection zone and fill slot on the right.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.