FINAL REPORT - DOE HUMAN GENOME PROJECT
DE-FG02-90ER-6085
June 1, 1990 – June 30, 2002
PROJECT TITLE: DNA Sequencing Using Capillary Electrophoresis

Barry L. Karger
Barnett Institute
Northeastern University
Boston, MA 02115

The overall goal of this program was to develop capillary electrophoresis as the tool to be used to sequence for the first time the Human Genome. Our program was part of the Human Genome Project. In this work, we were highly successful and the replaceable polymer we developed, linear polyacrylamide, was used by the DOE sequencing lab in California to sequence a significant portion of the human genome using the MegaBase multiple capillary array electrophoresis instrument. In this final report, we summarize our efforts and success. Reference numbers in the text in bold refer to the list of publications from this work listed at the end of the report.

We began our work by separating by capillary electrophoresis double strand oligonucleotides using cross-linked polyacrylamide gels in fused silica capillaries (**1**). This work showed the potential of the methodology. However, preparation of such cross-linked gel capillaries was difficult with poor reproducibility, and even more important, the columns were not very stable. We improved stability by using non-cross linked linear polyacrylamide. Here, the entangled linear chains could move when osmotic pressure (e.g. sample injection) was imposed on the polymer matrix. This relaxation of the polymer dissipated the stress in the column. Our next advance was to use significantly lower concentrations of the linear polyacrylamide that the polymer could be automatically blown out after each run and replaced with fresh linear polymer solution. In this way, a new column was available for each analytical run. Finally, while testing many linear polymers, we selected linear polyacrylamide as the best matrix as it was the most hydrophilic polymer available. Under our DOE program, we demonstrated initially the success of the linear polyacrylamide to separate double strand DNA (**2, 3, 7**). We note that the method is used even today to assay purity of double stranded DNA fragments.

Our focus, of course, was on the separation of single stranded DNA for sequencing purposes. In one paper, we demonstrated the success of our approach in sequencing up to 500 bases (**4**). Other application papers of sequencing up to this level were also published in the mid 1990's. (**5, 9**).

A major interest of the sequencing community has always been read length. The longer the sequence read per run the more efficient the process as well as the ability to read repeat sequences. We therefore devoted a great deal of time to studying the factors influencing read length in capillary electrophoresis, including polymer type and molecule weight, capillary column temperature, applied electric field, etc. In our initial optimization, we were able to demonstrate, for the first time, the sequencing of over 1000 bases with 90% accuracy (**6**). The run required 80 minutes for separation. Sequencing of 1000 bases per column was next demonstrated on a multiple capillary instrument (**10**).

Our studies revealed that linear polyacrylamide produced the longest read lengths because the hydrophilic single strand DNA had minimal interaction with the very hydrophilic linear polyacrylamide. Any interaction of the DNA with the polymer would lead to broader peaks and lower read length. Another important parameter was the molecular weight of the linear chains. High molecular weight (> 1 MDA) was important to allow the long single strand DNA to reptate through the entangled polymer matrix. In an important paper (**11**), we showed an inverse emulsion method to prepare reproducibility linear polyacrylamide polymer with an average MWT of 9MDa. This approach was used in the polymer for sequencing the human genome.

Another critical factor in the successful use of capillary electrophoresis for sequencing was the sample preparation method. In the Sanger sequencing reaction, high concentration of salts and dideoxynucleotide remained. Since the sample was introduced to the capillary column by electrokinetic injection, these salt ions would be favorably injected into the column over the sequencing fragments, thus reducing the signal for longer fragments and hence reading read length. In two papers (**12, 13**), we examined the role of individual components from the sequencing reaction and then developed a protocol to reduce the deleterious salts. We demonstrated a robust method for achieving long read length DNA sequencing.

Continuing our advances, we next demonstrated the achievement of over 1000 bases in less than one hour with a base calling accuracy of between 98 and 99% (**14**). In this work, we implemented energy transfer dyes which allowed for cleaner differentiation of the 4 dye labeled terminal nucleotides. In addition, we developed improved base calling software to help read sequencing when the separation was only minimal as occurs at long read lengths.

Another critical parameter we studied was column temperature. We demonstrated that read lengths improved as the column temperature was increased from room temperature to 60°C or 70°C (**8**). The higher temperature relaxed the DNA chains under the influence of the high electric field. The increased thermal motion allowed chains to display slight size differences as the DNA rotated through the entangled polymer chains. At room temperature, the long DNA fragments would align axially with the applied electric field, significantly decreasing read length.

As a final outcome of all this work, we showed under optimized conditions, the achievements of 1300 bases read length in 2 hours (**15**). To this day, this has been the longest read length achieved. Alternatively, we showed 1000 bases separated in under 40 minutes when DSMO was used as the denaturant in the capillary over area (**16**).

As noted in the beginning of this report, our studies played an important role in the development of DNA sequencing. Many factors affected sequencing and read length. We elucidated the most important factors which could be used for optimum read length or speed sequencing. The research was clearly successful. Of course, today new approaches to sequencing are being developed, leading to a second (even third) generation systems that are faster and less expensive. Nevertheless, these new developments would never have been possible without the research on the Human Genome Project. We are proud to have been part of that effort.

PUBLICATIONS

1. Chen, J.W., Cohen, A.S. and Karger, B.L., "Identification of DNA Molecules by Pre-column Hybridization Using Capillary Electrophoresis," J. Chromatogr., 559, 295 (1991).

2. Pariat, Y.F., Berka, J., Heiger, D.N., Schmitt, T., Vilencjik, M., Cohen, A.S., Foret, F. and Karger, B.L., "Separation of DNA Fragments by Capillary Electrophoresis Using Replaceable Linear Polyacrylamide Matrices," J. Chromatogr., 652, 57 (1993).

3. Berka, J., Y.F., Műller, O., Hebenbrock, K., Heiger, D.N., Foret. F. and Karger, B.L., "Sequence Dependent Migration Behavior of dsDNA in Capillary Electrophoresis", Electrophoresis 16, 377 (1995).

4. Ruiz-Martinez, M.C., Carrilho, E., Berka, J., Keileczawa, J., Miller, A.W., Foret, F., Carson, S. and Karger, B.L., DNA Sequencing by Capillary Electrophoresis Using Short Oligonucleotide Primer Libraries", BioTechniques, 20, 1058-1069, (1996).

5. Karger, B.L., Foret, F., Berka, J., "Application of Polymer Matrices to DNA and Protein Analysis by Capillary Electrophoresis", Methods in Enzymology Vol. 271:High Resolution Separation and Analysis of Biological Macromolecules, 293-319 (1996).

6. Carrilho, E., Reui-Martinez, M., Berka, J, Smirnov, I., Goetzinger, W., Foret, F., Miller, A.W., Brady, D. and Karger, B.L., "Rapid DNA Sequencing of More than 1,000 Bases per Run by CE Using Replaceable Linear Polyacrylamide Solutions", Anal. Chem., 68, 3305-3313 (1996).

7. Muth, J., Műller, O., Berka, J., Leonard, J.T. and Karger, B.L., "DNA Sequence Analysis of Prinker-Modified Restriction Fragments after Collection from Capillary Electrophoresis with Replaceable Polymer Matrices", J. Chromatogr. A, 744, 303-310, (1996).

8. Kleparnik, K., Foret, F., Berka, J., Goetzinger, W., Miller, A.W., Karger, B.L., "The Use of Elevated Column Temperature to Extend DNA Sequencing Read Lengths in Capillary Electrophoresis with Replaceable Polymer Matrices", Electrophoresis, 17, 1860-1866 (1996).

9. Cai, W., Cao, W., Wu, L., Exley, G.E., Waneck, G.L., Karger, B.L. and Warner, C.M., "Sequence and Transcription of Qa-2 Encoding Genes in Mouse T Cells and Blastocysts", Immunogenetics, 45, 97-107, (1996).

10. Carriilho, E., Miller, A.W., Ruiz-Martinez, M.C., Kolter, L., Kesilman, J., and Karger, B.L., "Factors to be Considered for Robust High-Throughput Automated DNA Sequencing Using a Multiple-Capillary Array Instrument", in Ultrasensitive Biochemical Diagnostics II, Gerald E. Cohn, Steven A. Soper, Editors, Proceedings of SPIE Vol. 2985, 4 (1997).

11. Goetzinger, W., Kolter, L., Mitnik, L. Carrilho, E., Ruiz-Martinez, M., Salas-Solano, O. and Karger, B.L., "Characterization of High Molecular Weight Linear Polyacrylamide Powder Prepared by Emulsion Polymerization as a Replaceable Polymer Matrix for DNA-Sequencing by Capillary Electrophoresis", Electrophoresis, 19, 242-248 (1998).

12. Ruiz-Martinez, M.C., Salas-Solano, O., Carrilho, E. and Karger, B.L., DNA Sequencing Analysis for CE Using Replaceable Polymer Solutions. Part A: Sample Purification Method for Rugged and High Performance Sequence Analysis", Anal. Chem., 70, 1516-1527, (1998).

13. Salas-Solano, O., Ruiz-Martinez, M.C., Carrilho, E., Kolter, L. and Karger, B.L. "DNA Sequencing Analysis for CE Using Replaceable Polymer Solutions. Part B: Quantitative Determination of the Role of Sample Matrix Components on Sequencing Analysis", Anal. Chem., 70, 1528-1535 (1998).

14. Salas-Solano, O., Carrilho, E., Kolter, L., Miller, A.W., Goetzinger, W., Sosic, Z. and Karger, B.L., "Routine DNA Sequencing of 1000 bases in Less than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions", Anal. Chem., 70, 3996-4003, (1998).

15. Zhou, H., Miller, A.W., Sosic, Z., Barron, A.E. Buchholz, B., Kolter, L. and Karger, B.L., "DNA Sequencing up to 1300 Bases in Two Hours by Capillary Electrophoresis with Mixed Replaceable Linear Polyacrylamide Solutions" Anal. Chem., 72, 1045 (2000).

16. He, H., Buchholz, B.A., Kolter, L., Miller, A.W., Barron, A.E. and Karger, B.L., " DNA Sequencing with Hydrophilic and Hydrophobic Polymers at Elevated Column Temperatures", Electrophoresis, 23, 1421-1428 (2002).

17. Kolter, L., He. H., Miller, A.W. and Karger, B.L., "DNA Sequencing of Close to 1000 Bases in 40 Minutes by Capillary Electrophoresis Using Dimethyl Sulfoxide and Urea as Denaturants with Replaceable Linear Polyacrylamide Solutions", Electrophoresis, 23, 3062-3070 (2002).