2003

Solid-phase DNA sequencing reactions performed in micro-capillary reactors and solid-phase reversible immobilization in microfluidic chips for purification of dye-labeled DNA sequencing fragments

Yichuan Xu

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Chemistry Commons

Recommended Citation

https://digitalcommons.lsu.edu/gradschool_dissertations/3478

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
SOLID-PHASE DNA SEQUENCING
REACTIONS PERFORMED IN MICRO-CAPILLARY REACTORS
AND
SOLID-PHASE REVERSIBLE IMMOBILIZATION
IN MICROFLUIDIC CHIPS FOR PURIFICATION OF DYE-
LABELED DNA SEQUENCING FRAGMENTS

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in
The Department of Chemistry

by
Yichuan Xu
B.S., Fudan University, 1983
M.S., Fudan University, 1986
May, 2003
Acknowledgements

I would like to take this opportunity to acknowledge those individuals who, over the years, have continually supported me and all of my endeavors and have inspired me to always look forward.

To Dr. Steven Soper, thank you for your constant support of my research and overall professional development.

To my wife, Yonghong Shang, my daughter, Mengze M. Xu, and my son, Kenneth X. Xu, without your constant love, understanding and support, this work would not have been possible.

To Drs. Kermit Murray, Williams Daly, Robin McCarley, and Diane Williams for serving on my committee and your guidance.

To Dr. John Hopkins for your direction in my application for admission to Louisiana State University.

To Drs. Ding S. Shih, Richard C. Brush and Simon H. Chang for the knowledge and the direction you bestowed upon me.

To Drs. Daryl Williams and Huijun Dong for lighting the way before me.

To Marcus Nauman and Ron Finley for your assistance in electronics.

To Dr. Soper’s group, I will remember you all. From the beginning to the end there was always one of you there to help lift my spirits when things got rough, especially Wieslaw Stryjewski, Yun Wang, Li Zhu, Michelle Galloway, Yuling Zhang, Shize Qi, Xuezhu Liu, Hengwu Chen, Benxu Chen, Gloria Thomas, Shawn Liopis, Musundi Wabuyelle, Harrison Musyimi, Vera Verdree, and Suzanne Lassiter.
# Table of Contents

Acknowledgements ........................................................................................................ ii  

List of Tables ................................................................................................................ vi  

List of Figures ................................................................................................................ vii  

List of Abbreviations .................................................................................................... xi  

Abstract .......................................................................................................................... xii  

## Chapter 1. Fluorescence-based DNA Sequencing ..................................................... 1  
1.1. DNA Chemistry and Biology .................................................................................. 1  
  1.1.1. Genome ............................................................................................................. 1  
  1.1.2. The Structure of DNA ....................................................................................... 2  
  1.1.3. DNA Replication .............................................................................................. 6  
  1.1.4. The Polymerase Chain Reaction (PCR) ............................................................ 9  
  1.1.5. Sanger DNA sequencing ................................................................................. 12  
1.2. Gel Electrophoresis ............................................................................................... 14  
  1.2.1. Theory of Gel Electrophoresis ........................................................................ 14  
  1.2.2. DNA Migration Models in Hydrophilic Gel Matrices .................................... 18  
  1.2.3. Polyacrylamide Gels ....................................................................................... 21  
  1.2.4. Agarose Gel Electrophoresis ........................................................................... 23  
  1.2.5. Slab Gel Electrophoresis ................................................................................ 25  
  1.2.6. Capillary Electrophoresis ................................................................................ 26  
1.3. Fluorescence .......................................................................................................... 33  
  1.3.1. Singlet-Singlet Transition ............................................................................... 33  
  1.3.2. Triplet-Singlet Transition ................................................................................ 35  
  1.3.3. Quantum Yields .............................................................................................. 36  
  1.3.4. Lifetime ........................................................................................................... 37  
1.4. Fluorescent DNA Sequencing ................................................................................ 38  
  1.4.1. Fluorescence Detection ................................................................................... 38  
  1.4.2. Dye Primer/Terminator Chemistry and Fluorescence Detection Formats ..... 47  
1.5. References .............................................................................................................. 57  

## Chapter 2. Micro-Capillary Reactors for Direct Sample Introduction into Slab Gels for Solid-Phase DNA Sequencing .................................................. 61  
2.1. Introduction ............................................................................................................ 61  
2.2. Experimental Section ............................................................................................. 67  
  2.2.1 Instrumentation ................................................................................................ 67  
  2.2.2 Preparation of Biotinylated DNA .................................................................. 70  
  2.2.3. Immobilization of Biotinylated PCR Product to Capillary Wall ................. 73  
  2.2.4. Radioactive Labeling of Biotinylated DNA ................................................... 74  
  2.2.5. The Rate of DNA Immobilization ................................................................. 75  
  2.2.6. Near-IR Fluorescence Sequencing with Dye-labeled Primer ...................... 75
List of Tables

1.1. Approximate gel concentrations for separation of DNA linear fragments of various sizes ................................................................. 17

1.2. Binary coding scheme for two-color DNA Sequencing ................................. 54

3.1. Signal-to-noise ratio of near-IR dye M13mp18 universal primer with on-column laser-induced fluorescence detection ................................................................. 105

4.1. Optimal concentration ratios of dye-ddNTP/dNTPs in dye-terminator DNA sequencing ........................................................................................................ 129

4.2. Average lifetime values of heavy-atom modified tricarbocyanine dye-terminator sequencing electrophoresed in a capillary with POP6 as the separation medium ........................ 135

5.1. The amount of DNA immobilized to the PC surface ........................................ 155
List of Figures

1.1. Image of genome........................................................................................................... 1

1.2. Chemical structures of nucleotide base units: adenine, guanine, cytosine, and thymine .......................................................................................................................... 3

1.3. Chemical structure of deoxyribose sugar ........................................................................ 3

1.4. Chemical structure of deoxyribonucleoside triphosphate (dNTP)............................... 4

1.5. Structure of part of DNA chain ................................................................................... 4

1.6. The Watson-Crick model for the structure of DNA ................................................. 6

1.7. Model of DNA chain elongation .................................................................................. 8

1.8. Schematic diagram of a DNA replication fork ............................................................ 9

1.9. Schematic of the PCR process .................................................................................... 11

1.10. Structural ddNTP analog ........................................................................................ 12

1.11. DNA sequencing by the Sanger method ................................................................. 13

1.12. Representation of the different mechanisms of DNA migration in an array .......... 20

1.13. Polymerization of acrylamide and bisacrylamide ...................................................... 22

1.14. Dependence of pore size on acrylamide concentration ........................................... 23

1.15. Agarose structural unit ............................................................................................ 24

1.16. Photograph of agarose gel, showing a comb and a mold.......................................... 24

1.17. Structure of Ethidium bromide ................................................................................ 25

1.18. Schematic diagram of slab gel electrophoresis system .......................................... 26

1.19. Schematic representation of a capillary electrophoresis system ................................ 28

1.20. Schematic diagram of silanol dissociation process .................................................... 28
1.21. Schematic of electroosmotic flow (EOF) generated in a fused silica capillary when an electric field is applied................................................................. 29

1.22. Jablonski diagram. Block energy diagram showing the various radiative and non-radiative processes that occur to a molecule upon the absorption of a photo of light ............................................................................................... 34

1.23. (A) Chemical structures of the dyes used for labeling primers for four-color DNA sequencing................................................................................... 42

1.23. (B) Absorbance and emission spectra of the dyes shown in (A).............43

1.24. (A) Chemical structures of the dyes conjugated to ddNTPs......................... 44

1.24. (B) Absorbance and emission spectra of the dyes shown in (A). Also shown are the excitation laser wavelength (488 nm) and the filter set used to isolate the fluorescence onto the two detection channels................................. 45

2.1. Chemical structures of immobilization layer used to tether biotinylated PCR products to fused silica capillaries (A). Also included in this figure are the sequences of the chemical linker of biotin to the PCR product (B)........ 65

2.2. A schematic drawing of the basic principle of the solid-phase DNA sequencing..... 66

2.3. Flexigene thermal cycler ........................................................................... 67

2.4. A gross gel electrophoresis system which includes a gel casting tray, a gel 68

2.5. The Li-COR 4000 automated DNA sequencer ........................................... 68

2.6. Expandal view of electrophoresis apparatus .............................................. 69

2.7. The scanning platform in Li-COR 4000 automated DNA sequencer............ 71

2.8. Sequences of PCR primers ........................................................................ 71

2.9. Gel image of agarose gel electrophoresis ................................................... 73

2.10. Structure of IRD800 dye labeled sequencing primer .................................. 76

2.11. Structure of NIR dye-labeled ddGTP .......................................................... 78

2.12. Slab gel electrophoresis of sequencing fragments of a 1 kb PCR amplified λ-DNA immobilized to the wall of a micro-capillary reactor ..................... 80
2.13. Sequencing trace (called bases 1-589) of a 1000 bp PCR amplified λ-bacteriophage template immobilized to the wall of a capillary micro-reactor .... 81

2.14. Gel image of C-tract subjected to 15 sequencing rounds on a single A-bacteriophage template immobilized to the solid-phase capillary micro-reactor .... 83

2.15. Radioactivity (cpm) of $^{32}$P labeled DNA immobilized to the wall of a micro-capillary reactor as a function of denaturation time................................................................. 84

2.16. The rate of DNA immobilization onto capillary wall........................................ 85

2.17. The rate of streptavidin immobilization onto capillary wall............................ 86

2.18. Gel image of a single G tract with the dye-terminator chemistry...................... 88

3.1. Excitation and emission spectra of the IRD800 dye primer at a concentration of 2 µM in H$_2$O .............................................................................................................. 98

3.2. Transmission spectra of filter system used for near-IR sequencing experiments..... 99

3.3. The picture of the rapid air thermal cycler housing the capillary nanoreactor used to prepare sequencing fragments .......................................................... 100

3.5. Schematic diagram of CE system with nanoreactor and capillary gel column..... 101

3.6. Near-IR system for CGE analysis........................................................................ 103

3.7. CGE separation and on-column detection of a 5 X 10$^{-12}$ M solution of IRD800 labeled M13mp18 universal primer ................................................................. 105

3.8. Capillary gel electropherograms of C-terminated Sanger sequencing fragments produced in the nanoreactor possessing a volume of 62 nL ........................................ 109

4.1. Block diagram of near-Infrared time-correlated single-photon counting instrument. .................................................................................................................. 121

4.2. Structures of heavy-atom modified tricabocyanine dye-labeled (A) I-ddATP (B) Br-ddCTP, (C) Cl-ddGTP (D) F-ddUTP ................................................................. 123

4.3. Front panel displays of the data analysis packages for analyzing time-resolved data form sequencing gels................................................................. 128

4.4. Slab gel electrophoresis of sequencing fragments of a single stranded M13mp18 DNA using M13(-40) forward primer ....................................................... 130
4.5. Sequence output from Li-COR 4000 automated DNA sequencer................................. 131

4.6. CGE electropherograms along with the associate decay profiles of DNA sequencing fragments with (A) NIR-I ddATP, (B) NIR-Br ddCTP, (C) NIR-Cl ddGTP and (D) NIR-F ddUTP.................................................................................................................. 133

4.7. Electropherogram of two dye-terminator sequencing fragments in a capillary. ..... 136

5.1. (A) A possible mechanism of photochemical modification of PC in the mid-UV, Photo-Fries Rearrangement. (B) One reaction of the excited O₂ – complex is to form hydroperoxide via a caged mechanism........................................ 144

5.2. (A) Schematic of the integrated microchannel-CGE system. (B) Schematic diagram of the microchannel. (C) Optical micrograph of PC-SPRI capture bed and its dimensions .............................................................................................................. 150

5.3. (A) Photoactivated PC (through UV photomask) imaged with a near-IR microscope. (B) Fluorescence image of the slide in (A) following incubation with the dye-labeled M13 sequencing ladder. The fragments were produced using cycle sequencing and a near-IR dye-labeled terminator. Excess terminator was removed using off-chip SPRI cleanup prior to incubation of the labeled fragments with the photoactivated PC surface. (C) Fluorescence image of the slide in (B) following washing with distilled water. In both cases, the surfaces were imaged with a near-IR confocal microscope (780 nm excitation) operated in a scanning mode. The fluorescence was processed on a single photon avalanche diode transducer........................................................................... 154

5.4. (A) Slab gel electrophoresis of sequencing fragments of a single stranded M13mp18 DNA using M13(-40) forward primer. (B) Sequence output from Li-COR 4000 automated DNA sequencer .............................................................. 156

5.5. Gel image of G and C tracts subjected to four purification rounds on a same PC channel................................................................................................................................. 157

5.6. Capillary gel electropherograms of near-IR dye A-terminated sequencing fragments purified in the PC microchannel .................................................................................. 161

6.1. Integrated modular microfluidic system for the automated processing of sequencing templates ........................................................................................................................... 173
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Ammonium Persulfate</td>
</tr>
<tr>
<td>APTS</td>
<td>Aminopropyltriethoxy silane</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary Electrophoresis</td>
</tr>
<tr>
<td>CGE</td>
<td>Capillary Gel Electrophoresis</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine Triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxycytosine Triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>Deoxyguanosine Triphosphate</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide Triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>Deoxytyrosine Triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>Dideoxyadenosine Triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>Dideoxyctosine Triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>Dideoxyguanosine Triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>Dideoxytyrosine Triphosphate</td>
</tr>
<tr>
<td>ddUTP</td>
<td>Dideoxyuridine Triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>EOF</td>
<td>Electroosmotic Flow</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>LIF</td>
<td>Laser Induced Fluorescence</td>
</tr>
<tr>
<td>LPA</td>
<td>Linear Polyacrylamide</td>
</tr>
<tr>
<td>MAPS</td>
<td>Methacryloxypropyltriethoxy Silane</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimator</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infrared</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>SGE</td>
<td>Slab Gel Electrophoresis</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SPRI</td>
<td>Solid Phase Reversible Immobilization</td>
</tr>
<tr>
<td>ss</td>
<td>single stranded</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
Abstract

The research presented in this dissertation involves micro-capillary reactors for solid phase DNA sequencing, the identification of dye terminator sequencing fragments with time-resolved methods, and purification of dye-labeled DNA fragments using solid-phase reversible immobilization in microfluidic chips.

Solid-phase micro-reactors have been prepared for DNA sequencing applications using slab gel electrophoresis. A PCR product was immobilized to the interior wall of a fused-silica capillary tube via a biotin-streptavidin linkage. Solid-phase sequencing was carried out in micro-capillary reactors using a four-lane, single color dye primer chemistry strategy. The read length was found to be 589 bases. The complementary DNA fragments generated in the small volume (~62 nL) reactor were directly injected into the gel-filled capillary for size separation with detection accomplished using near-IR laser-induced fluorescence.

A set of terminators labeled with near-IR heavy-atom modified tricarbocyanine dyes were investigated for a terminator sequencing protocol in conjunction with slab gel electrophoresis. This protocol gave 605 bp read lengths. A one color, two lifetime format of DNA sequencing was implemented. A pixel-by-pixel analysis was employed to identify each of the bases in the run. The resulting read accuracy for the two-dye capillary run was 90.6%.

The use of photoactivated polycarbonate (PC) for purification of dye-labeled terminator sequencing fragments using solid-phase reversible immobilization (SPRI) was investigated. SPRI cleanup of dye-terminator sequencing fragments using a
photoactivated PC microchannel and slab gel electrophoresis produce a read length of 620 bases with a calling accuracy of 98.9%. The PC-SPRI cleanup format was also integrated to a capillary gel electrophoresis system. In this case, the immobilization microchannel contained microposts to increase the loading level of DNAs to improve signal intensity without the need for pre-concentration.
Chapter 1

Fluorescence-Based DNA Sequencing

1.1. DNA Chemistry and Biology

1.1.1. Genome

Initiated formally in 1990, the U.S. Human Genome Project was a 13-year effort coordinated by the U.S. Department of Energy and the National Institutes of Health. Project goals were to identify all of the approximate 30,000 genes in human DNA and to determine the sequences of the 3 billion chemical base pairs that make up human DNA. The genome is an organism’s complete set of DNA (see Figure 1.1). Genomes vary widely in size: the smallest known genome for a free-living organism (a bacterium)

Figure 1.1. Image of genome. Genome is the complete gene complement of an organism, contained in a set of chromosomes (in eukaryotes), in a single chromosome (in bacteria), or in a DNA or RNA molecule (in viruses).
contains about 600,000 DNA base pairs, while human and mouse genomes have around 3 billion. Except for mature red blood cells, all human cells contain a complete copy of the genome.

DNA in the human genome is arranged into 24 distinct chromosomes—physically separate molecules that range in length from about 50 million to 250 million base pairs. A few types of major chromosomal abnormalities, including missing or extra copies or gross breaks and rejoinings (translocations), can be detected by microscopic examination. Most changes in DNA, however, are subtler and require a closer analysis of the DNA molecule to find perhaps single-base differences. Each chromosome contains many genes, the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. Genes comprise only about 2% of the human genome; the remainder consists of noncoding regions, whose functions may include providing chromosomal structural integrity and regulating where, when, and in what quantity proteins are made. The human genome is estimated to contain 30,000 to 40,000 genes.

1.1.2. The Structure of DNA

DNA is a linear polymer that is made up of nucleotide units. The nucleotide unit consists of a nitrogenous base, a sugar, and a phosphate. The nitrogenous base is a derivative of purine or pyrimidine. The purines in DNA are adenine (A), guanine (G), and the pyrimidines are thymine (T), cytosine (C) (see Figure 1.2). The sugar is deoxyribose (see Figure 1.3).

A nucleoside consists of a purine or pyrimidine base bond to a sugar. The four nucleosides in DNA are called deoxyadenosine, deoxyguanosine, deoxycytidine and
deoxythymidine. A nucleotide is a phosphate ester of a nucleoside. Such a compound is called a nucleoside-5’-phosphate. (see Figure 1.4)

![Chemical structures of nucleotide base units: adenine, guanine, cytosine, and thymine.](image)

**Figure 1.2.** Chemical structures of nucleotide base units: adenine, guanine, cytosine, and thymine.

![Chemical structure of deoxyribose sugar.](image)

**Figure 1.3.** Chemical structure of deoxyribose sugar.
The backbone of DNA, which is invariant throughout the molecule, consists of deoxyriboses linked by phosphate groups. Specifically, the 3\(^{-}\)-hydroxyl of the sugar moiety of one deoxyribonucleotide is joined to the 5\(^{-}\)-hydroxyl of the adjacent sugar by a phosphodiester bridge. The variant part of DNA is its sequence of four bases (A, G, C, and T). The corresponding molecule units are called deoxyadenylate, deoxyguanylate, deoxycytidylate and deoxythymidylate. The structure of a DNA chain is shown in Figure 1.5.
Utilizing X-ray diffraction data obtained from crystals of DNA, James Watson and Francis Crick proposed a model for the structure of DNA.\(^1\) This model (subsequently verified by additional data) predicted that DNA would exist as a helix of two complementary antiparallel strands wound around each other in a rightward direction and stabilized by H-bonding between bases in adjacent strands. In the Watson-Crick model, the bases are in the interior of the helix aligned at nearly 90 degree angle relative to the axis of the helix. Purine bases form hydrogen bonds with pyrimidines in the crucial phenomenon of base pairing. Experimental determination has shown that, in any given molecule of DNA, the concentration of adenine (A) is equal to thymine (T) and the concentration of cytidine (C) is equal to guanine (G).\(^1\) This means that A will only base-pair with T, and C with G. According to this pattern, known as Watson-Crick base-pairing, the base-pairs composed of G and C contain three H-bonds, whereas those of A and T contain two H-bonds. This makes G-C base-pairs more stable than A-T base-pairs. The antiparallel nature of the helix stems from the orientation of the individual strands. From any fixed position in the helix, one strand is oriented in the 5' --- 3' direction and the other in the 3' --- 5' direction. On its exterior surface, the double helix of DNA contains two deep grooves between the ribose-phosphate chains. These two grooves are of unequal size and termed the major and minor grooves. The difference in their size is due to the asymmetry of the deoxyribose rings and the structurally distinct nature of the upper surface of a base-pair relative to the bottom surface (see Figure 1.6).
Figure 1.6. The Watson-Crick model for the structure of DNA.

1.1.3. DNA Replication

Replication of DNA occurs during the process of normal cell division cycles. Because the genetic complement of the resultant daughter cells must be the same as the parental cell, DNA replication must possess a very high degree of fidelity. The entire process of DNA replication is complex and involves multiple enzymatic activities. Replication of DNA occurs during the process of normal cell division cycles. Because the genetic complement of the resultant daughter cells must be the same as the parental cell, DNA replication must possess a very high degree of fidelity. The entire process of DNA
replication is related to its subcellular localization as well as its primary replicative activity. The polymerase of eukaryotic cells that is the equivalent to E. coli pol III is pol-α. The pol I equivalent in eukaryotes is pol-β. Polymerase-γ is responsible for replication of mitochondrial DNA. The ability of DNA polymerases to replicate DNA requires a number of additional accessory proteins. The combination of polymerases with several of the accessory proteins yields an activity identified as DNA polymerase holoenzyme. These accessory proteins include primase processivity accessory proteins, single strand binding proteins, helicase, DNA ligase, topoisomerases and uracil-DNA N-glycosylase.

DNA is synthesized by DNA polymerases. A single unpaired strand is required to act as template, and a primer is needed to provide a free 3’ end to which new nucleotide units are added. Each incoming nucleotide is selected by virtue of base pairing to the appropriate nucleotide in the template strand. The fundamental reaction is a nucleophilic attack by the 3’-hydroxyl group of the nucleotide at the 3’ end of the growing strand on the 5’-α-phosphorus of the incoming deoxynucleoside 5’-triphosphate (see Figure 1.7).

The process of DNA replication begins at specific sites termed origins of replication, requires a primer bearing a free 3’-OH, proceeds specifically in the 5’ ----> 3’ direction on both strands of DNA concurrently and results in the copying of the template strands in a semiconservative manner. The semiconservative nature of DNA replication means that the newly synthesized daughter strands remain associated with their respective parental template strands. The DNA replication includes the following steps. A portion of the double helix is unwound by a helicase. A molecule of a DNA polymerase binds to one strand of the DNA and begins moving along it in the 3’ to 5’ direction, using it as a template for assembling a leading strand of nucleotides and reforming a double
helix. In eukaryotes, this molecule is called DNA polymerase delta (δ). Because DNA synthesis can only occur 5' to 3', a molecule of a second type of DNA polymerase (epsilon, ε, in eukaryotes) binds to the other template strand as the double helix opens. This molecule must synthesize discontinuous segments of polynucleotides (called Okazaki fragments). Another enzyme, DNA ligase, then stitches these together into the lagging strand (see Figure 1.8).

Figure 1.7. Model of DNA chain elongation.
Kary Mullis conceived the idea for the polymerase chain reaction in the spring of 1983 while an employee of Cetus Corporation, a biotechnology firm located near Berkeley, California. Mullis and his assistant Fred Faloona tried to get it to work later in the year, and were soon joined by other Cetus scientists who saw the great potential of this method. In this later group were people in Henry Erlich's lab, who had been working on methods to identify mutations in human genomic DNA as part of a DNA diagnostics group at Cetus. Description of PCR was terse in their first publication, a 1985 article in Science, on detection of the mutation causing sickle cell anemia in whole genomic DNA. The details of the PCR method and its uses were covered more fully in articles published in the next two years. A general narrative of the invention of PCR, the internal politics of Cetus Corporation, and the philosophy and anthropology of scientific
invention can be found in the book by Rabinow. The DNA polymerase originally used for the PCR was extracted from the bacterium *T. aquaticus*. However, after each cycle of DNA synthesis, the reaction must be heated to denature the double stranded DNA product. Unfortunately, heating also irreversibly inactivated this polymerase, so new enzyme had to be added at the start of each cycle. The bacterium *T. aquaticus* lives in hot springs and produces a DNA polymerase which is not irreversibly inactivated at high temperature. David Gelfand and his associates at Cetus, purified, and subsequently cloned this polymerase, allowing a complete PCR amplification to be done without opening the reaction tube. However, the thermostable enzyme was found to be much more than just a convenience. The DNA synthesis step could now be done at a higher temperature than was possible with the E. coli enzyme, and it was discovered that the template DNA strand was now copied with high fidelity, eliminating the nonspecific products that had plagued earlier attempts at amplification.

The PCR is a powerful technique used to amplify DNA millions fold by repeated replication of a template in a short period of time. The process utilizes sets of specific in vitro synthesized oligonucleotides to prime DNA synthesis. The design of the primers is dependent upon the sequences of the DNA that is desired to be analyzed. The technique is carried out through many cycles (usually 20 - 50) of melting the template at high temperature, allowing the primers to anneal to complementary sequences within the template and then replicating the template with DNA polymerase. The process has been automated with the use of thermostable DNA polymerases isolated from bacteria that grows in thermal vents in the ocean or hot springs. During the first round of replication a single copy of DNA is converted to two copies and so on resulting in an exponential
increase in the number of copies of the sequences targeted by the primers. After just 20 cycles a single copy of DNA is amplified over 2,000,000 fold (see Figure 1.9).

![Figure 1.9. Schematic of the PCR process.](image)

The polymerase chain reaction can be used to amplify both double and single stranded (eg the products of a reverse transcription reaction, RT-PCR) DNA. The template is mixed with specific or degenerate primers, dNTPs, polymerase buffer including MgCl₂ and thermostable DNA polymerase. The template is denatured at high temperature (eg 95°C) and then cooled to a temperature that will allow optimal primer binding. The reaction temperature is then raised to that optimal for the DNA polymerase (eg 72°C) whereby the primers are extended along the template. This series of steps is carried out 20 - 30 times leading to exponential amplification of the target template. The
amplification is so great that the reaction products can be visualized following gel electrophoresis.

### 1.1.5. Sanger DNA Sequencing

Two basic methods are used for DNA sequencing, chemical degradation method of Maxam and Gilbert\(^9\) and the ddNTP-mediated chain-termination method of Sanger.\(^{10}\) The original technique for sequencing, Maxam and Gilbert sequencing,\(^9\) relies on the nucleotide-specific chemical cleavage of DNA and is not routinely used.

The enzymatic technique, Sanger sequencing,\(^{10}\) involves the use of dideoxynucleotides (2',3'-dideoxy) (see Figure 1.10) that terminate DNA synthesis and is, therefore, also called dideoxy chain termination sequencing.

![Figure 1.10. Structure ddNTP analog.](image)

The Sanger DNA sequencing protocol utilizes dideoxynucleotides (ddNTPs) to terminate chain elongation during the in vitro synthesis of DNA from a cloned template. Synthesis is initiated using a specific oligonucleotide primer. Four separate reactions are carried out simultaneously, each of which contains all 4 dNTPs and a single ddNTP. The higher the concentration of ddNTP the more frequently chain elongation will terminate. Therefore, one can regulate the extent of sequence information obtainable by varying the dNTP/ddNTP ratio. Following the extension reactions, the products are resolved by electrophoresis in a denaturing (urea) polyacrylamide gel. The results are obtained when the gel is dried and exposed to x-ray film. Bands near the bottom of the gel represent
short reaction products (ie closest to the 3'-end of the primer) and those near the top the longest products (see Figure 1.11). The Sanger method is in more widespread use because it has proven to be technically easier. It involves the enzymatic synthesis of a DNA strand complementary to the strand to be analyzed.

**Figure 1.11.** DNA sequencing by the Sanger method.
1.2. Gel Electrophoresis

The introduction of the technique of electrophoresis in acrylamide or agarose gels was a major advance in nucleic acid technology. Previously, large and expensive ultracentrifuge equipment was required to separate macromolecules on the basis of sedimentation coefficients or buoyant densities, and scintillation counters were required to measure the radioactivity of the physically separated fractions. The advent of gel electrophoresis, ethidium bromide staining and autoradiography essentially brought an end to the era of intensive centrifugal analysis of nucleic acids. The development of these techniques was combined with the use of restriction enzymes and cloning in bacterial plasmids or phages to obtain large quantities of pure fragments of DNA molecules. Gel electrophoresis is simple, rapid, nondestructive and inexpensive, and, when combined with autoradiography, blot hybridization and elution of specific fragments from the gel, is an extremely powerful tool in the arsenal of recombinant DNA technology both for preparative and for analytical purposes.

1.2.1. Theory of Gel Electrophoresis

The commonality in all electrophoresis formats is the use of an electric field to shuttle the DNAs through a maize, which consists of a polymer, either static or dynamic, which possesses pores of various sizes. In all electrophoresis experiments, the mobility of the molecule (\( \mu \), cm\(^2\)/Vs, defined as steady state velocity per unit electric field strength) in an electric field is determined by;

\[
\mu = \frac{q}{f}
\]  
\( (1.1) \)

where \( q \) is the net charge on the molecule and \( f \) is the frictional property of the molecule and is related to its conformational state as well as the molecular weight (MW) of the
molecule. For example, proteins can be considered solid spheres and \( f \sim (\text{MW})^{1/3} \). In the case of DNAs, either single stranded or double stranded, \( f \sim (\text{MW})^1 \) since the DNA molecule acts as a free-draining coil. Since \( q \) is also related to the length of the DNA molecule, one finds that \( \mu \sim N_b^0 \), where \( N_b^0 \) indicates that the mobility (in free solution) is independent of the number of bases comprising the DNA molecule. Because of this property of DNAs, the electrophoresis step must include some type of sieving medium, which can be a polymer consisting of pores with a definitive size.

In reality, electrophoresis in a gel matrix can be thought of as a type of "gel filtration" where the driving force is the electric field. Penetration of the gel matrix by a mixture of nucleic acid molecules of different sizes results in the retardation of larger molecules. The existence of a gel network contributes significantly to the electrophoretic migration pattern (ie, electrophoretic mobility) observed in the sequencing process. Any DNA fragments to be fractionated will inevitably encounter the gel network of polymer threads. This encounter increases the effective friction and consequently, lowers the velocity of movement of the molecules. It is obvious that the retardation will be most pronounced when the mean diameter of gel pores is comparable to the size of the DNA fragments. Size therefore plays a critical role in determining the relative electrophoresis mobility and the degree of separation of different DNA fragments. It is this sieving effect that partly determines the resolution obtained with polyacrylamide gels. The gel network also minimizes convection currents caused by small temperature gradients.

The main characteristic property of DNA is that the value of the negative charge on the molecule is almost independent of the pH of the medium. Therefore, their electrophoretic mobility is due mainly to differences in molecular size not in charge. For
a particular gel, the electrophoretic mobility of a DNA fragment is inversely proportional to the logarithm of the number of bases up to a certain limit. Since each DNA fragment is expected to possess a unique size, its electrophoretic mobility is unique and it migrates to a unique position within the electric field in a given length of time. Therefore, if a mixture of DNA fragments is subjected to electrophoresis each of the fragments would be expected to concentrate into a tight migrating band at unique positions in the electric field.

Exclusion of nucleic acid macromolecules from the gel occurs if the average pore size is smaller than the size of the macromolecule. However in the case of duplex DNA, extremely large molecules can penetrate a gel by "snaking" through the pores in an extended fashion. This behavior gives rise to the nonlinear migration and pileup of large DNA molecules, which is normally observed at the top of an agarose gel; this phenomenon has been used in the development of an elegant technique for the separation of extremely large duplex linear DNA molecules by field inversion and orthogonal field electrophoresis.\textsuperscript{11,12}

As mentioned above, large molecules which cannot penetrate the gel pores in a normal fashion pile up at a short distance from the origin. Small molecules, which migrate through the gel almost without retardation, pile up in a smear at the bottom of the gel. Molecules of intermediate size migrate through the gel with a mobility approximately proportional to the log of their molecular size. This region of linearity varies with the gel concentration and must be empirically determined for each mixture of nucleic acid molecules for which molecular sizes are desired. The values presented in Table 1.1 can be used to select a gel for a desired separation. For accurate size
determination, a mixture of fragments of known size is usually run in one or more gel lanes. Several algorithms are available for determination of the size of unknown fragments by least squares curve fitting procedures.\textsuperscript{13-15}

**Table 1.1.** Approximate gel concentrations for separation of DNA linear fragments of various sizes\textsuperscript{14}

<table>
<thead>
<tr>
<th>Gel Concentration</th>
<th>Separation Range (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>agarose 0.3%</td>
<td>60,000 - 5,000</td>
</tr>
<tr>
<td>agarose 0.7%</td>
<td>20,000 - 800</td>
</tr>
<tr>
<td>agarose 0.9%</td>
<td>7,000 - 500</td>
</tr>
<tr>
<td>agarose 1.2%</td>
<td>6,000 - 400</td>
</tr>
<tr>
<td>agarose 1.5%</td>
<td>4,000 - 200</td>
</tr>
<tr>
<td>agarose 2.0%</td>
<td>3,000 - 100</td>
</tr>
<tr>
<td>agarose 4.0%</td>
<td>500 - 10</td>
</tr>
<tr>
<td>acrylamide 4%</td>
<td>1,000 - 800</td>
</tr>
<tr>
<td>acrylamide 10%</td>
<td>500 - 25</td>
</tr>
<tr>
<td>acrylamide 20%</td>
<td>50 - 1</td>
</tr>
</tbody>
</table>

For any type of analytical separation, resolution is a key parameter which is optimized to improve the performance of the separation. The resolution (R) for electrophoretic separations can be calculated from the simple relation;

$$R = \frac{N^{1/2}}{4 \Delta \mu_{\text{app}}/\mu_{\text{app,avg}}}$$  \hspace{1cm} (1.2)

where $\Delta \mu_{\text{app}}$ is the difference in mobility between two neighboring bands, $\mu_{\text{app,avg}}$ is the average electrophoretic mobility for two neighboring bands and N is the plate number and represents the efficiency (band width) for the electrophoresis. As can be seen from this equation, the resolution can be improved by increasing the difference in the mobility of the two bands (increase selectivity, gel property) or increasing the plate numbers (narrower bands). As a matter of reference, when $R = 0.75$, two bands are baseline
resolved. For DNA sequencing, the accuracy in the base call depends intimately on the resolution obtained during gel fractionation.

1.2.2. DNA Migration Models in Hydrophilic Gel Matrices

Much research has been dedicated to the effects of electrophoretic parameters on DNA separations in entangled polymer networks and their connections to theoretical models. As a result, DNA migration in polymer solutions has been described by various molecular mechanisms. Two main theories describe the migration of a flexible macromolecule through a polymer network: the Ogston model, and the reptation model.

The Ogston model treats the polymer network as a molecular sieve. In the Ogston model, it is assumed that the matrix consists of a random network of interconnected pores having an average pore size. The migrating solute behaves as an undeformable particle of radius. Smaller molecules migrate faster because they have accessibility to a larger fraction of the available pores, providing the following expressions;

\[ \log \mu = \log \mu_0 - \pi L (r + R)^2 T \]  

(1.3)

where \( \mu \) is the electrophoretic mobility of the analyte in the matrix, \( \mu_0 \) is its free solution electrophoretic mobility, \( L \) is the gel polymer length per unit volume, \( r \) is the polymer radius, \( R \) is the radius of the migrating molecule, and \( T \) is the total polymer concentration. This model does not consider the electric field (E) and it’s effects on \( R \). Therefore, (1.3) is only true as \( E \rightarrow 0 \). Additionally, a plot of the log of mobility as a function of polymer concentration, a Ferguson plot, should give a linear relationship; however, for longer DNAs, a significant deviation from linearity occurs, which supports an Ogston-reptation transition.
The reptation model is based on the assumption that the migrating molecule does not behave as an immutable sphere with fixed radius, but instead behaves as a free draining coil.\textsuperscript{20} According to the reptation model, randomly coiled DNAs are too large to fit through a pore while maintaining a coiled conformation will migrate headfirst, snakelike, through “tubes” formed by the pore network of the gel. No lateral motion is allowed within the tube and the DNAs are thought to alternately stretch and relax as they slither through the tube. Thus, the reptation model assumes that large DNAs, instead of migrating as undeformable particles with a fixed radius of gyration, can deform and stretch according to local conditions including overly high voltages, too low of a pore size, or a combination of both. The mobility of the molecule by this process is inversely proportional to its molecular size.

When high electric fields are applied, electric field-induced orientation extends the stretching periods of DNA, causing their random walk to become strongly biased in the forward direction so that DNA is stretched to a rod-like conformation. The biased reptation model accounts for this process.\textsuperscript{21} In a fully biased reptation regime, the mobility increases to saturation and size-based separation is lost with all large fragments migrating at the same rate. Some effects of biased reptation can be reduced by pulsing, reversing the electric field for a given period of time to relax the large DNA back into their original conformation before continuing the separation. Lumpkin et al.\textsuperscript{21} developed an expression that explains this process;

\[
\mu = \frac{k}{N} + bE^2 \tag{1.4}
\]
where $N$ is the DNA molecular weight, $K$ is a constant, $b$ is a function of the mess size of the polymer network as well as the charge and the length of the migrating solute and $E$ is the electric field. At low field strength, the first term in equation (1.40 dominates and is dependent on the molecule size of the migrating molecule. At high field strength, the second term in the equation dominates and is basically independent of the molecular weight of the molecule.

In Figure 1.12 are shown the DNA migration mechanisms in hydrophilic gel matrices.

![Figure 1.12. Representation of the different mechanisms of DNA migration in an array of fixed obstacles. A) Ogston sieving. B) Reptation without orientation. C) Reptation with orientation (alignment in electric field).](image-url)
1.2.3. Polyacrylamide Gels

Polyacrylamide gels are formed by the copolymerization of acrylamide, a water-soluble monomer, with a cross-linking agent to form a three-dimensional lattice. The cross-linking agent of choice for most applications is N,N’-methylene bisacrylamide (BIS), which is shown below.

\[
(CH_2=CH–CO–NH–CH_2–NH–CO–CH=CH_2)
\]

BIS contains two double bonds, which in polymerization reactions cross-links with adjacent chains of polyacrylamide. Acrylamide, if polymerized in the absence of a cross-linking agent, forms only linear polymers, resulting in viscous aqueous solutions rather than a gel. The polymerization reaction occurs by a free radical chain mechanism; the free radicals are generated either by photolysis of a labile compound (e.g., riboflavin) or by chemical decomposition of a labile compound. We describe the latter method of radical generation with ammonium persulfate and tetramethylethylenediamine (TEMED). Ammonium persulfate is the di-sulfate ester of hydrogen peroxide, and readily homolyzes into unstable radicals. TEMED is a tertiary amine that reacts with these radicals to form TEMED free radicals, which in turn react with acrylamide to induce polymerization as shown in Figure 1.13.

The size of the pores in the net is determined by two parameters: (1) the amount of acrylamide used per unit volume of reaction medium and (2) the degree of cross-linkage. Regardless of the total amount of acrylamide per unit volume, the average pore size reaches a minimum when 5% of the total acrylamide used is N,N’-methylene-bis(acrylamide). Therefore, in many formulations the bis(acrylamide) content is fixed at 5% of the total acrylamide and is not altered as a means of controlling pore size. Pore
Figure 1.13. Polymerization of acrylamide and bisacrylamide.\textsuperscript{22} A) Free radical polymerization scheme of acrylamide. B) Crosslinking polymerization reaction of acrylamide and bis-acrylamide. Ammonium persulfate becomes a free radical when dissolved in water and when added to an acrylamide solution in the presence of the catalyst tetra-methylenethyldiamine, TEMED, the reaction proceeds.
size is manipulated by varying the total content of acrylamide. Figure 1.14 indicates an approximation of the pore size yielded from various concentrations (w/v) of acrylamide.

![Graph showing the average pore size versus percentage acrylamide.](image)

**Figure 1.14.** Influence of gel concentration (percentage weight/total gel volume) on the average pore size. The bis(acrylamide) content is fixed at 5% of the total acrylamide.²²

### 1.2.4. Agarose Gel Electrophoresis

The standard method used to separate, identify, and purify DNA fragments is electrophoresis through agarose gels.²³⁻²⁶ Agarose, which is extracted from seaweed, is a linear polymer whose basic structure is shown in Figure 1.15. Bands of DNA in the gel are stained with the intercalating dye ethidium bromide for visualization. As little as 1 ng of DNA can be detected by direct examination of the gel in ultraviolet light.²³

Agarose gels are cast by melting the agarose in the presence of the desired buffer. The melted solution is then poured into a mold and allowed to harden. Upon hardening,
the agarose forms a matrix, the density of which is determined by the concentration of the agarose. When an electric field is applied across the gel, DNA, which is negatively charged at neutral pH, migrates toward the anode (see Figure 1.16).

![Figure 1.15. Agarose structural unit.](image)

The most convenient method of visualizing DNA in agarose gels is by use of the fluorescent dye ethidium bromide\(^{23,24} \) (2,7-Diamino-10-ethyl-9-phenyl-phenanthridinium bromide). This substance (see Figure 1.17) contains a planar group that intercalates between stacked bases of DNA. The fixed position of this group and its close proximity to the bases causes dye bound to DNA to display an increased fluorescent yield compared to dye in free solution. UV-irradiation absorbed by the DNA at 260 nm and transmitted to the dye, or irradiation absorbed at 300 nm and 360 nm by the bound dye itself, is emitted at 590 nm in the red-orange region of the visible spectrum. Ethidium bromide can be used...
to detect both single- and double-stranded nucleic acids. However, the affinity of the dye for single-stranded nucleic acid is relatively low and the fluorescent yield is poor.

![Structure of Ethidium bromide](image)

**Figure 1.17.** Structure of Ethidium bromide.

1.2.5. Slab Gel Electrophoresis

DNA sequencing is usually performed with slab gels. In slab gel electrophoresis, polymerization of acrylamide as well as the electrophoresis are conducted in a mold formed by two glass plates with a thin spacer (see Figure 1.18). When analytical electrophoresis is performed, several samples are usually run simultaneously in the same gel. Since all samples are present in the same gel, the conditions of electrophoresis are quite constant from sample to sample. In slab gel electrophoresis, the sample is loaded using a pipette into wells formed into the gel during polymerization. The field strength that can be applied to these types of gels ranges from 50-80 V/cm, with the upper limit determined by heating (Joule) caused by current flow through the gel. Due to the thick nature of the gel, this heat is not efficiently dissipated causing convective mixing and
thus, zone broadening, which limits the upper level on the electric field strength that can be effectively used.

Figure 1.18. Schematic diagram of slab gel electrophoresis system.

1.2.6. Capillary Electrophoresis

1.2.6.1. Basic Principles

Electrophoresis, defined as migration of analytes within an electrolyte solution under the influence of an electric field, has been a mainstay of biochemical separation methods since the mid-1930s when it was first described by Tiselius.\textsuperscript{27} Capillary electrophoresis (CE) developed as a technique in the late 1980s and has grown steadily in popularity.\textsuperscript{28-31} In its simplest and most commonly used form, it is a throwback to the original Tiselius method, but its popularity is based not only on the inherent simplicity but on the additional advantages of speed, versatility and low running costs. Essentially as an analytical method, it has found application in the separation of biopolymers such as peptides, proteins, oligonucleotides, metal ions and inorganic ions as well as in the analysis of pharmaceuticals and the monitoring of water quality.\textsuperscript{32} Although the basic
methodology involves the separation of molecules based on their charge to mass ratio, there are straightforward modifications to the procedure, borrowed from existing well-established techniques, which allow separations based on size or isoelectric point, or which permit the separation of non-charged molecules.

As the name suggests the distinctive feature of the method is that separations are carried out in a capillary tube (Figure 1.19), the cross-section of which means that temperature control is readily achieved. The length of the capillary differs in different applications, but is normally in the region of 20 to 75 cm. The capillary is filled with running buffer and the sample is introduced by dipping one end into the sample and applying an electric field (electrokinetic injection) or by applying gas pressure (pressure injection). Migration through the capillary is driven, directly or indirectly, by an electric field, and analytes are detected as they pass the window at the far end. Detection is normally by absorbance or, less frequently, fluorescence, and given the short path length requires sensitive monitoring equipment. Limits of detection can be improved by introducing a bend in the capillary such that detection is along a short length of the tubing, but this is at the expense of some resolution.

1.2.6.2. Electroosmotic Flow

The walls of a fused silica capillary contain silanols, which ionise in contact with a high pH electrolyte solution. This dissociation produces a negatively charged wall. A layer of cations is then established at the wall to preserve electroneutrality (see Figure 1.20). When a voltage is applied these cations and their associated solvating water molecules migrate towards the cathode. This movement of ions and their associated water molecules results in a flow of solution towards the detector. This flow effectively pumps
Figure 1.19. Schematic representation of a capillary electrophoresis system.

Figure 1.20. Schematic diagram of silanol dissociation process.
Figure 1.21. Schematic of electroosmotic flow (EOF) generated in a fused silica capillary when an electric field is applied.

Solute ions along the capillary towards the detector and could be considered as an "electrically-driven pump" (see Figure 1.21.). At low pH the silanols are unionised and therefore the flow rate is much reduced or can become zero at very low pH values.

The magnitude of the EOF can be expressed mathematically in terms of mobility (\(\mu_{EOF}\)) by:

\[
\mu_{EOF} = \left( \frac{\varepsilon \zeta}{4\pi \eta r} \right)
\]

(1.5)

where \(\eta\) is viscosity, \(\zeta\) is zeta potential, \(r\) is capillary radius, and \(\varepsilon\) is the dielectric constant (C^2/Joule·cm). This term can be treated in vector terms in an additive or subtractive nature depending on the direction of the EOF. The electrophoretic mobility of a solute in the presence of EOF is termed apparent mobility \(\mu_{app}\) and is calculated as:

\[
\mu_{app} = \mu_{EOF} + \mu_e
\]

(1.6)

where \(\mu_e\) is the electrophoretic mobility of the analyte.
1.2.6.3. Capillary Zone Electrophoresis (CZE)

This is the simplest form of the technique and is essentially "free solution" electrophoresis. The migration of analytes in solution under the influence of an electric field is affected by the charge/size ratio, with the field-induced mobility promoted by the charge being opposed by the frictional drag through the electrolyte solution. This is a complex function with contributions from the size of the ion, including any hydration shell, and the viscosity of the electrolyte. Prediction of mobility is therefore not simple in practice, but a useful relationship has been developed empirically by Rickard et al.,\textsuperscript{33} who has shown a good correlation between the mobility of peptides with the function $q/M^{2/3}$, where $q$ is the charge carried by an ion of mass $M$.

1.2.6.4. Capillary Gel Electrophoresis (CGE)

Capillary electrophoresis has been extensively investigated as a practical tool for DNA sequencing. Initially introduced as capillary gel electrophoresis, it has proven to be a successful analytical technique for DNA sequencing, capable of fast and sensitive separations. Due to the higher surface-to-volume ratio afforded by the thin glass capillary, large electric fields can be applied, which results in shorter electrophoresis development times and also, enhanced plate numbers compared to slab gel electrophoresis. An electroosmotic flow can interfere with the electrophoresis of the DNAs. Therefore, in DNA separations using glass capillaries, the wall is coated with neutral polymers (for example a linear polyacrylamide) to suppress this electroosmotic flow. After coating the wall, the capillary can then be filled with the sieving gel.
Early work reported successful application of crosslinked polyacrylamide and melted agarose to capillary electrophoresis for the separation of DNAs.\textsuperscript{34-36} Crosslinked polyacrylamide gel filled capillaries can be prepared by polymerizing the acrylamide monomer in situ. Although these chemical gels provide high resolution of fragments, covalent bonding to the capillary wall limits the lifetime of the capillary in the event of bubble formation, polymer degradation or column fouling. On the other hand, physical gels that contain no crosslinking between polymer strands permit replenishment of the matrix after each separation and recycling of the capillary.\textsuperscript{37,38} This typically extends the life of the capillary, prevents cross-contamination of samples and allows the use of higher separation temperatures.

Low viscosity polymer solutions have been shown to provide high-resolution separations of dsDNA mixtures.\textsuperscript{37,39} Since an entangled mesh requires no specific crosslinking or gelation, a wide range of polymers may be easily employed for DNA size separations using electrophoresis. Polymers must only be water soluble and uncharged. The earliest work used linear polyacrylamide (LPA).\textsuperscript{34-36} More recent reports utilize dextran, poly (ethylene glycol)\textsuperscript{40} and glucomannan\textsuperscript{41} as the sieving medium. A very commonly used entangled polymer solution is cellulose and its derivatives.\textsuperscript{42}

The higher plate numbers that is obtained in capillary gel electrophoresis compared to slab gel is a direct consequence of the ability to use higher electric fields. Since the development time is significantly shorter in capillary gel electrophoresis due to the ability to apply higher electric fields, band spreading due to longitudinal diffusion is reduced resulting in higher plate numbers. The ability to use higher electric fields in

31
capillary gel electrophoresis results from the fact that Joule heating is suppressed since the heat can be effectively dissipated by the high surface-to-volume ratio capillary.

1.2.6.5. Injection Bias

One disadvantage of CE involves injection related artifacts encountered during electrokinetic injection. Injection biases toward high electrophoretic mobility ions and species of high concentration also occur during sample introduction; smaller fragments and those of high concentration are loaded onto the capillary in larger quantities than longer, slower moving fragments or those of low concentration. The transference number (T) provides a measure of the fraction of the total electric current carried by each species in solution during the injection process. Assuming a sample containing n solutes in free solution, the transference number of any given DNA fragment during the electrokinetic injection from free solution into a capillary filled with polymer solution can be calculated as:

\[
T_{DNA} = \frac{C_{DNA} z_{DNA} \mu_{DNA}}{\sum_{i=1}^{n} C_i z_i \mu_i}
\]  

where \(C_i\) is the concentration of an ion (i), \(z_i\) is the charge on the ith ion, and \(\mu_i\) is the free solution mobility of the ith ion. Additionally, chloride ions migrate faster than DNA in free solution, a difference that should be enhanced in polymer solutions. This can be particularly problematic in separations of DNA fragments of low abundance in the presence of salts and an excess of primers, as in mutation screening applications.
1.3. Fluorescence

All molecules absorb light. However, only a relatively low number of molecular species (usually rigid conjugated polyaromatic hydrocarbons or heterocycles) emit light as a result of absorption of light from some other source. If the emission is immediate or from the electronically excited singlet state, the phenomenon is called fluorescence.

1.3.1. Singlet-Singlet Transition

Fluorescence is the emission of light from a molecule in which an electronically excited state has been populated. The emission of the light is usually in the ultraviolet to visible portion of the spectrum, sometimes in the near-infrared (see Figure 1.22). The energy levels of the fluorophore are represented by horizontal lines and are grouped in bands. The lowest band is associated with the ground electronic state, $S_0$. The lowest energy of the excited singlet state is represented by $S_1$, whereas $T_1$ is the lowest level of the excited triplet state.

Upon excitation, fluorophores in the ground state absorb a photon and jump to higher vibrational energy levels of the electronically excited singlet state, shown as absorption of light, $A$. The photon of excitation is supplied by an external source, such as an incandescent lamp or a laser. The transition from $S_0$ to higher excited levels of $S_1$ is responsible for the visible and ultraviolet absorption spectra observed in fluorophores.

Excitation is followed by a return to the lower vibrational levels of the electronically excited state. This relaxation occurs in about a picosecond. The excited state itself exists for a finite time. Typical values of excited-state lifetimes are in the range of nanoseconds.
Figure 1.22. Jablonski diagram. Block energy diagram showing the various radiative and non-radiative processes that occur to a molecule upon the absorption of a photo of light.

From the singlet state, the fluorophore returns to the electronic ground state with the emission of the photon, shown as fluorescence F, but to higher vibrational levels of this state. In fluorescence emission, the spin multiplicities of the ground and excited states are the same. One can measure either a steady state spectrum of emitted light or the actual decay kinetics of emission.

Photoemission is unimolecular. It is a first order process in the concentration of the excited state. The rate constants for fluorescence are typically of the order of $10^8 \text{ s}^{-1}$. The energy of the photon that is emitted as the electron decays to the ground state depends on the energy difference between the excited and ground state at the time of emission. The rapid decay of excited vibrational states implies that the state from which
the fluorophore decays is independent of the excitation wavelength. However, the state to which the fluorophore decays is not always the lowest vibrational state of the ground state, but it is an equilibrium distribution of vibrational levels. Therefore, the emission spectra of fluorescent molecules show fine structure. The probability of decay from the excited state to each vibrational level of the ground state is what determines the shape of the fluorescence spectrum.

1.3.2. Triplet-Singlet Transition

During its lifetime, the excited state is subject to a variety of possible interactions with its molecular environment. Processes such as;

- partial dissipation to relaxed singlet excited state;
- collisional quenching;
- fluorescence energy transfer;
- intersystem crossing;

may depopulate the singlet state and thus the energy is lost for fluorescence.

Some fluorophores may leave the excited state via two main processes other than fluorescence, especially at high levels of excitation power. They can either;

- irreversibly switch to non-fluorescent species by non-radiative transitions, which is called bleaching;
- or they can undergo intersystem crossing into the triplet state, T₁ from which the fluorophore returns to the ground state by emission of light as phosphorescence, shown as P.

Intersystem crossing occurs when a triplet state lies just below the excited singlet electronic state, i.e. there is a near coincidence of two vibrational levels in the excited
singlet state and triplet states. In the triplet state, the spins of the excited and ground state electrons are no longer paired. This involves the flipping of one of the electron spins so that unpaired electron spin results. Thus, in phosphorescence emissions, the spin multiplicities of the ground and excited states are different. Because the change in spin violates the quantum mechanical spin conservation rules, the decay from the triplet to ground state is very slow and occurs only if there are no other allowed energy paths open. Common rate constants for phosphorescence are in the range $10^2$-$10^5 \text{s}^{-1}$.

There are other mechanisms that can delay the fluorescence emission to very long periods. This can occur if the energy follows some circuitous path in the excited state before returning to the lower vibrational levels of the electronically excited state. Such a delayed fluorescence is different from true phosphorescence, which is derived from the triplet state. Delayed fluorescence results from two intersystem crossings, first from the singlet to the triplet, then from the triplet back to the singlet.

1.3.3. Quantum Yields

The fluorescence quantum yield is the ratio of the number of photons emitted to the number of photons absorbed. The emissive rate of the fluorophore ($\Gamma$) and the rate of radiationless decay to $S_0$ ($k_{nr}$) both depopulate the excited state. The fraction of fluorophores that decay by emission, defined by the quantum yield, is given by;

$$Q = \frac{\Gamma}{\Gamma + k_{nr}} \quad (1.9).$$

The quantum yield can be close to unity if the radiationless decay, that is $k_{nr} \ll \Gamma$. The energy yield of fluorescence is always less than unity because of Stokes’ losses.
1.3.4. Fluorescence Lifetime

The lifetime of the fluorophore is the average value of the time a fluorophore spends in the excited state before it returns to the ground state. The expression for the time-dependent fluorescence emission \( I(t) \) can be described as:

\[
I(t) = \sum_{i=1}^{n} A_i e^{-t/\tau_f(i)}
\]

(1.10)

where \( n \) represents the number of components in the decay, \( A \) is a preexponential factor, \( t \) is the time, \( \tau_f \) is the fluorescence lifetime. Because the fluorescence lifetime is proportional to the fluorescence quantum yield, the fluorescence lifetime can be determined experimentally by measuring the time it takes the fluorescence intensity to fall to 1/e of its initial value following excitation.

The value of the fluorescence lifetime of a molecule depends on the radiative and nonradiative decay processes that are involved and can be expressed through:

\[
\tau_f = \frac{1}{k_r + k_{nr}} = \frac{1}{k_m}
\]

(1.11)

where \( k_r \) is the radiative rate constant (s\(^{-1}\)), \( k_{nr} \) is the nonradiative rate constant (s\(^{-1}\)), and \( k_m \) is the total decay rate (s\(^{-1}\)).

There are two principal methods to measure fluorescence decay parameters: the time- and phase-resolved methods. In the time resolved method, the sample is excited by a short pulse of light, which may have a width at half-maximum of 10 nanosecond or less. The fluorescence decay (intensity) is recorded as a function of time from which the lifetime can be calculated by deconvolution, assuming one to three exponentials in the decay curve. The decay function is monoexponential for simple fluorophores. Often
however, the decay of fluorescence cannot be described by a single first-order rate process because two or more species are present with different lifetimes. In this case, the quantum yield still can be determined from the corrected spectrum, but the relationship between the quantum yield and fluorescence lifetime is complex. In general, the time dependence of the fluorescence decay can be written as a multiexponential decay curve in terms of an average relaxation time.

In the phase-resolved method, the sample is excited by a sinusoidically demodulated light. The finite lifetime of fluorescence causes it to be delayed in phase and to be modulated to a smaller degree. The phase shift is expressed as a phase angle from which the lifetime can be determined using simple relationships between the modulation frequency and the degree of demodulation. The modulation frequencies are of the order of 10-50 MHz for most lifetimes. Shorter lifetimes require higher modulation frequencies up to 2 GHz. Using a single excitation frequency, multiexponential decays cannot be easily resolved. By obtaining phase data at different frequencies, such as in multi-frequency phase fluorometry, one can obtain all information needed to resolve complex decay curves. Multi-frequency modulation is also referred as the frequency domain method. Both the pulse and phase shift methods give essentially identical results.

1.4. Fluorescent DNA Sequencing

1.4.1. Fluorescence Detection

Following electrophoresis, the individual DNA bands separated on the gel must be detected and subsequently analyzed. One of the earlier methods implemented to detect DNA bands in gels was autoradiography. In this mode, one of the phosphates of an individual nucleotide is replaced with a radioisotope, typically $^{32}$P ($\tau_{1/2} = 14$ days) or
$^{35}\text{S} (\tau_{1/2} = 87 \text{ days})$, both or which are radioprobes that emit $\beta$-particles. When Sanger methods are used to prepare the sequencing reactions, either the primer or the dideoxynucleotide can contain the radiolabel. The labeling is done using an enzyme (T4 polynucleotide kinase), which catalyzes the transfer of a $\gamma$-phosphate group from ATP to the 5'-hydroxy terminus of a sequencing primer. After the electrophoresis has been run, the gel is dried and then, situated on an X-ray film. The film is developed (exposure to radiation from radioprobes) and dark bands are produced on the film where the DNA was resident. This is then followed by reading the sequence from the gel manually.

The primary advantage of this approach is the inexpensive nature of the required equipment to perform the measurement. It basically requires only a gel dryer, film holder and film. The difficulties associated with this approach are numerous. One important issue is the fact that radioisotopes are used, and therefore, waste disposal becomes a difficult problem to contend with. Throughput issues (data production rates) are also a primary concern when using autoradiographic detection. For example, radiography can sometimes require several days to expose the film in order to get strong signals to read the bases from the gel. In addition, the detection is done after the electrophoresis and not during the electrophoretic run. Also, since there is no means to identify the individual bases using radioprobes, each base must be analyzed in a different lane of the gel. And finally, the bases must be called manually, which many times leads to frequent errors in the sequence reconstruction. Therefore, the inability to obtain data production rates sufficient to accommodate large sequencing projects has made radiographic detection obsolete for high throughput applications.
For most DNA sequencing applications, irrespective of the separation platform used, fluorescence is the accepted detection protocol for several important reasons. Fluorescence allows one to perform the base calling and detection in an automated fashion and alleviates the need for manual base calling. In addition, fluorescence can be carried out during the separation, eliminating long film development times. More importantly, due to the ability to implement multiple probes possessing unique spectral properties, the four bases comprising the DNA molecule can be identified in a single gel lane, potentially increasing throughput by a factor of four compared to radiographic detection. All of these important advantages associated with fluorescence allow for higher throughputs in DNA sequencing applications. As such, fluorescence can be considered as one of the most important recent technical innovations in DNA sequencing and has made it feasible to consider tackling large genome sequencing projects, such as the human genome.

The first demonstrations on the use of fluorescence in DNA sequencing came with the work of Smith et al., Probe et al. and Ansorge et al. In these works, slab gels or large gel tubes were used to fractionate the DNA ladders produced during enzymatic polymerization using Sanger sequencing strategies. The fluorescence detection was accomplished using four spectroscopically unique probes, which allowed the DNA sequence reconstruction to be done in a single electrophoresis lane of the gel. The chemical structures of the dye labels used in the Smith and Prober’s experiments are shown in Figures 1.23 and 1.24, respectively. As can be seen, the dyes were either attached (covalently) to the sequencing primer or to the dideoxynucleotides. The advantage of using dye-labeled dideoxynucleotides is that the sequencing reactions can
be performed in a single reaction tube, whereas the dye-labeled primer reactions must be performed in 4 separate tubes and pooled prior to electrophoresis. In the case of the dye-labeled terminators, succinylfluorescein analogs were used with slight structural modifications to alter the absorption/emission maxima. The dyes were attached either to the 5 position of the pyrimidine bases or the 7 position of the 7-deazapurines, both of which are non-hydrogen bonding sites on the nucleotide base. The linker structure is also important, which in this case was a propargylamine, since the presence of the dye onto the terminator radically affects its ability to be incorporated by the polymerase enzyme. For dye-labeled primers, the oligonucleotides possessing the appropriate sequence were prepared on a standard DNA synthesizer. For Smith et al., a thymidine derivative was prepared, which contained a phosphoramidite at the 3' carbon and a protected alkyl amino group at the 5' carbon (typically a 6-carbon linker structure). During the final addition cycle of the oligonucleotide prepared via solid-phase synthesis using phosphoramidite chemistry, the thymidine residue is added and following deprotection of the alkyl amino group and cleavage from the support, a free primary amine group results, which can be reacted with any amino-reactive fluorescent dye to produce the oligonucleotide derivative.

Also shown in Figures 1.23 and 1.24 are the absorption and emission profiles for the dye sets used in these experiments. The major attributes of the dye sets are that they can be efficiently excited with either 488 and/or 514.5 nm lines of the Ar ion laser. In addition, there is minimal separation between the emission maxima of the dyes, which allows processing of the fluorescence on as few detection channels as possible. However, there is significant overlap in the emission spectra of the four dyes, producing
severe spectral leakage into other detection channels, which must be corrected by software.

![Chemical structures of the dyes used for labeling primers for four-color DNA sequencing.](image)

**Figure 1.23.** (A) Chemical structures of the dyes used for labeling primers for four-color DNA sequencing. 

\[ X = \text{DNA primer} \]
Figure 1.23. (B) Absorbance and emission spectra of the dyes shown in (A).
Figure 1.24. (A) Chemical structures of the dyes conjugated to ddNTPs.
Figure 1.24. (B) Absorbance and emission spectra of the dyes shown in (A). Also shown are the excitation laser wavelength (488 nm) and the filter set used to isolate the fluorescence onto the two detection channels.
For the Smith et al. experiment, a single laser was used as well as a single detection channel, which in this case consisted of a multi-line Ar ion laser and a conventional photomultiplier tube (PMT). Placed in front of the laser and PMT were filter wheels to select the appropriate excitation wavelength (488 or 514.5 nm) and emission color. The filter pairs used during fluorescence readout were 488/520 nm; 488/550 nm; 514/580 nm; 514/610 nm. In the case of the Prober et al. experiment, due to the narrow distribution between the excitation maxima of the dyes, only excitation using the 488 nm line from the Ar ion laser was required as well as two PMT tubes to process the emission from the four colors. Discrimination of the four colors was accomplished by monitoring the intensity of each dye on both detectors simultaneously. By histograming the ratio of the fluorescence intensity of each dye (produced from an electrophoresis band) on the two detection channels, a discrete value was obtained that allowed facile discrimination of the four different fluorescent dyes (ie, terminal base). In order to determine the limit of detection of these fluorescence systems, injections of known concentrations of dye labeled sequencing primers were electrophoresed. In both cases, the mass detection limit was estimated to be $10^{-17}$ – $10^{-18}$ moles.

Unfortunately, reading the sequence directly from the raw gel data becomes problematic due to several non-idealities, including signal from a single dye appearing on multiple detection channels due to the broad and closely spaced emission bands, dye-dependent electrophoretic mobility shifts, and non-uniformity in the intensity of the electrophoresis bands due to the enzymatic reaction used to construct the individual DNA size-ladders. As such, several post-electrophoresis processing steps were required to
augment sequence reconstruction of the test template. In the case of the Smith et al. example, these steps involved:

1) High frequency noise removal using a low-pass Fourier filter.

2) A time-delay between measurements at different wavelengths corrected by linear interpolation between successive measurements.

3) A multi-component analysis performed on each set of four data points, which produced the amount of the four dyes present in the detector as a function of time.

4) The peaks present in the data stream located.

5) The mobility corrected for the dye attached to each DNA fragment. In this case, it was empirically determined that fluorescein and rhodamine-labeled DNA fragments moved as if they were 1 base longer than the NBD-labeled fragments and the Texas Red fragments moved as if they were 1 ¼ bases longer.

The important performance criteria in any type of automated DNA sequencer is its throughput, the number of bases it can process in a single gel read and its accuracy in calling bases. In terms of base calling accuracy, these early instruments demonstrated an error rate of approximately 1% with a read length approaching 500 bases. The throughput of the instrument described by Prober et al. was estimated to give a raw throughput of 600 bases per hour (12 electrophoresis lanes). Interestingly, many present-day commercial automated sequencers still use similar technology in their machines and the throughput can be as high as 16,000 bases per hour (96 electrophoresis lanes).

1.4.2. Dye Primer/Terminator Chemistry and Fluorescence Detection Formats

In most sequencing applications, dye-labeled primers are used for accumulating sequencing data using automated instruments. This stems from the fact that dye-labeled
primers are typically less expensive to use compared to their dye-labeled terminator counterparts. Also, in most applications, small pieces of DNA (1-2 kbp in length) are cloned into bacterial vectors for propagation (to increase copy number), such as M13s, which have a known sequence and serve as ideal priming sites. However, dye-labeled primers do present one with problems, for example, the sequencing reactions must be run in four separate tubes during polymerization and then pooled prior to the gel electrophoresis. In addition, unextended primer can result in a large electrophoretic peak (ie, high intensity) which can often times mask the ability to call bases close to the primer annealing site.

Dye-labeled terminators can be appealing to use in certain applications, for example cases where high quality sequencing data is required and also, in primer walking strategies. In primer walking, the sequence of the DNA template is initiated at a common priming site using a primer that is complementary to that site. After reading the sequence at that site, the template is subjected to another round of sequencing, with the priming site occurring at the end of the first read. In this way, a long DNA can be sequenced by walking in a systematic fashion down the template. Dye terminators are particularly attractive since primers need to be synthesized frequently in primer-walking strategies and the need for non-labeled primers simplifies the synthetic preparation of these primers. Dye terminators improve the quality of sequencing data in many cases since the excess terminators are removed prior to electrophoresis (using size exclusion chromatography) and as such, give clean gel reads free from intense primer peaks. However, it should be noted that in most cases, terminators can produce uneven peak
heights (broad distribution of fluorescence intensities) due to the poor incorporation efficiency of dye-terminators by polymerase enzymes.

When dye-labeled primers are used, several different formats can be implemented to reconstruct the sequence of the template when using fluorescence detection. In the case of spectral discrimination, these formats may vary in terms of the number of dyes used, the number of detection channels required, or the need for running 1-4 parallel electrophoresis lanes. For example, if the sequencing instrument possesses no spectral discrimination capabilities, the electrophoresis must be run in four different lanes, one for each base comprising the DNA molecule. However, if four different dyes are used the electrophoresis can be reduced to one lane, and as a consequence, the production rate of the instrument goes up by a factor of 4. The fluorescence-based formats that will be discussed here includes (# dyes / # electrophoresis lanes), single dye/four lane; single dye/single lane; two dye/single lane and finally, four dye/single lane strategies.

The most pressing issue in any type of DNA sequencing format is the accuracy associated with the base call, which is intimately related to a number of experimental details, for example the number of spectral channels used in the instrument as well as the signal-to-noise ratio in the measurement. The information content of a signal, $I$, can be determined from the simple relation: \[ I = n \log_2(SNR) \] (1.12)

where $n$ is the number of spectral channels and SNR is the signal-to-noise ratio associated with the measurement. The term $I$ is expressed in bits and typically 2 bits are necessary to distinguish between four different signals, but only if there is no spectral overlap between the dyes used for identifying the bases. Unfortunately, in most multi-color
systems, the spectra of the dyes used in the sequencing device show significant overlap and as such, many more bits will be required to call bases during the sequencing run.

While the above equation can provide information on how to improve the accuracy of the base call, it does not provide the sequencer with information on the identity of the individual electrophoretic peaks (base call) nor the quality of a base call within a single gel read. For example, if four color sequencing is used with dye-primer chemistry, how should one process the data and what is the confidence to which an electrophoretic peak is called an A, T, G or C? In order to provide such information, an algorithm has been developed to not only correct for anomalies associated with fluorescence-based sequencing, but also assign a quality score to each called base. The typical algorithm that is used is called the Phred scale and it uses several steps to process the sequencing data obtained from fluorescence-based, automated DNA sequencers.50

The data input into Phred consists of a trace, which is electrophoretic data processed into four spectral channels, one for each base. The algorithm consists of four basic steps:

1. Idealized electrophoretic peak locations (predicted peaks) are determined. This is based on the premise that most peaks are evenly spaced throughout the gel. In regions where this is not the case, typically during the early and late phases of the electrophoresis, predictions are made as to the number of correct bases and their idealized locations. This step is carried out using Fourier methods as well as the peak spacing criterion and helps to discriminate noise peaks from true peaks.

2. Observed peaks are identified in the trace. Peaks are identified by summing trace values to estimate the area in regions which satisfy the criterion, $2 \times v(i) - v(i+1)$.
\[ v(i) = v(i-1) + v(i-1), \text{ where } v(i) \text{ is the intensity value at point } i. \] If the peak area exceeds 10% of the average area of the preceding 10 accepted peaks, and 5% of the area of the immediate preceding peak, it is accepted as a true peak.

3. Observed peaks are matched to the predicted peak locations, omitting some peaks and splitting others. In this phase of the algorithm, the observed peak arises from one of four spectral channels, and thus, can be associated with one of the four bases. It is this ordered list of matched observed peaks which determines a base sequence for the DNA template in question.

4. Observed peaks that are uncalled (unmatched to predicted peaks) are processed. In this step, an observed peak that did not have a complement in the predicted trace is called and assigned a base and finally inserted into the read sequence.

As can be seen, this algorithm mainly deals with sorting out difficulties associated with the electrophoresis by identifying peaks in the gel traces, especially in areas where the peaks are compressed (poor resolving power) or where multiple peaks are convolved due to significant band broadening produced by diffusional artifacts.

Many times, pre-processing of the traces is carried out prior to Phred analysis to correct for dye-dependent mobility shifts. In most cases, these mobility shifts are empirically determined by running an electropherogram of a single dye-labeled DNA ladder (for example, T-terminated ladder) and comparing the mobilities to the same ladder, but labeled with another dye of the set. This type of analysis can be very complex and involved, since the mobility shift is not only dependent upon the dye and linker structure, but also upon the separation platform used. For example, dyes which show uniform mobility shifts in slab gel electrophoresis may not show the same effect in
capillary gel electrophoresis. In addition, these mobility shifts can be dependent upon
the length of the DNA to which the dye is attached. These type of mobility shifts have
been ascribed to not only differences in the net charge of the dye-label, but also potential
dye-DNA base interactions. These interactions, predominately driven by hydrophobic
interactions, may cause loops or hairpin structures on the 5' end of the dye-DNA
complex. These structures would cause a faster migration rate compared to fully
extended structures produced by most dye-DNA complexes.

1.4.2.1 Single Color/Four Lane

In this processing format, only a single fluorescence detection channel is required
to analyze the signal from the labeling dye since only a single dye is used to detect the
sequencing fragments produced following chain extension. However, since no color
discrimination is implemented, the electrophoresis must be run in four lanes, one for each
base, similar to the format used in traditional autoradiographic detection. While this is a
reasonable approach for slab gel separations, it is not a viable strategy in capillary gel
applications due to the poor run-to-run reproducibility in the migration rates of the
fragments traveling through the different capillaries. This is due to differences in the gel
from capillary to capillary as well as differences in the integrity of the wall coatings used
to suppress the electroosmotic flow. In the slab gel format, reproducibility in the
migration times becomes less of a problem since all of the lanes are run in the same gel
matrix.

1.4.2.2. Single Color/Single Lane

In this sequencing approach, only a single fluor is used and as such, only a single
laser is required to excite the fluorescence and only a single detection channel is needed
to process the fluorescence. The advantage of this approach is that instrumentally, it is very simple since the hardware required for detection is simple. In addition, since the sequence is reconstructed from a single electrophoresis lane and not four, the throughput can be substantially higher compared to a single fluor/four lane method.

The bases are identified by adjusting the concentration ratio of the terminators used during DNA enzymatic polymerization to alter the intensity of the resulting electrophoretic bands. Therefore, if the concentration of the terminators used during DNA polymerization was 4:2:1:0 (A:C:G:T), a series of fluorescence peaks would be generated following electrophoretic sizing with an intensity ratio of 4:2:1:0 and the identification of the terminal bases would be carried out by categorizing the peaks according to their heights. In order to accomplish this with some degree of accuracy in the base calling, the ability of the DNA polymerase enzyme to incorporate the terminators must be nearly uniform. This can be achieved using a special DNA polymerase, which in this case is a modified T7 DNA polymerase. This enzyme has been modified so as to remove its proof reading capabilities by eliminating its 3' → 5' exonuclease activity. Since this method requires uniform incorporation of the terminators, it is restricted to the use of dye-primer chemistry. In addition, since the T7 enzyme is not a thermostable enzyme, cycle sequencing cannot be used.

1.4.2.3. Two Color/Single Lane

In order to improve on the base calling accuracy associated with the single color/single lane strategy without having to increase the instrumental complexity of the fluorescence readout device associated with sequencing instruments, one may employ a two color format to identify the four terminal bases in sequencing applications. In this
approach, one or two lasers are used to excite one of two spectrally distinct dyes used for labeling the sequencing primers and the fluorescence is processed on one of two detection channels, consisting of bandpass filters and photon transducers.

Table 1.2. Binary coding scheme for two-color DNA Sequencing. The (1) indicates the presence of the dye-labeled primer during DNA polymerization and the (0) represents the absence of the dye-label.

<table>
<thead>
<tr>
<th></th>
<th>FAM</th>
<th>JOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A schematic of a two-color scanning instrument was used to excite the fluorescence of the labeling dyes, FAM and JOE. Due to their similar absorption maxima, a single laser (488 nm) could efficiently excite the fluorophores. In addition, this dye-pair was selected since they produced sequencing fragments which co-migrated and thus, no mobility correction was required. The bases were identified using a binary coding scheme, which is shown in Table 1.2. During DNA polymerization, four separate reactions were run with the A-reaction containing an equimolar mixture of the FAM and JOE labeled sequencing primers. In the case of G, only the JOE labeled primer was present, while for T, only the FAM labeled primer was present and for C, no dye labeled primer was used. By ratioing the signal in the red (JOE) to green (FAM) channel, a value was obtained which could be used to identify the terminal base. The attractive feature associated with this protocol is that while the absolute intensity of the bands present in the electropherogram may vary by a factor of 20 due to sequence dependent termination, the ratio only varies by a factor of 1.7. The read length using this binary
coding system was up to approximately 350 bases, with the number of errors ~15 (accuracy = 95.7%). The majority of the errors were attributed to C determinations, since a null signal was used to indicate the presence of this base.

In order to alleviate the errors in the base calling associated with identifying bases using a null signal, a two-dye, two-level approach can be implemented.\(^{59,60}\) In this method, the bases using a common dye-label have the concentration of the ddNTPs adjusted during chain extension to alter the intensity of the fluorescence peaks developed during the electrophoresis. Also required in this approach is uniform peak heights, requiring the use of the modified T7 DNA polymerase in the presence of manganese ions. For example, Chen et al. used a FAM-labeled primer for marking Ts and Gs, with the concentration ratio of the ddNTPs adjusted to 2:1 (T:G). Likewise, the As and Cs were identified using a 2:1 concentration ratio of the terminators with the labeling dye in this case being TAMRA.\(^{59}\) Sequencing data produced an effective read length of 350 bases, with an accuracy of 97.5%. When the concentration ratios of the terminators sharing a common labeling dye was increased to 3:1, the read length was extended to ~400 bases with a base-calling accuracy >97%. As can be seen, the elimination of null signals to identify bases can improve the base calling accuracy in these types of sequence determinations.

While most fluorescence labeling strategies for DNA sequencing which depend on differences in intensities of the electrophoretic peaks to identify bases use dye-labeled primers, internal labeling, where the fluorescent dye is situated on the dNTP, can also be used.\(^{61}\) The advantages associated with using dye-labeled dNTPs are: (1) ability to use a wide range of primers since no dye-labeled primer is required; (2) incorporation of the
dye-labeled dNTP can be much more uniform than dye-labeled ddNTPs; (3) dye-labeled dNTPs are much less expensive compared to dye-labeled primers and terminators. Using a tetramethylrhodamine-labeled dATP and a fluorescein labeled dATP, a two-color/single lane sequencing assay has been reported. For internal labeling, a two step polymerization reaction was used, in which the template was annealed to the sequencing primer (unlabeled) along with the dye-dATP and the four unlabeled dNTPs as well as the polymerase enzyme. The extension reaction was incubated at 37°C for 10 mins after which the appropriate terminator was added and the reaction allowed to proceed for an additional 10 mins. The initial extension reaction extended 6-8 nucleotides to a quartet of As, with 80-90% of the fragments containing a single dye-labeled dATP. Since only two dyes were used in this particular example, the concentration ratio for a pair of terminators sharing a common dye was adjusted (3:1) so as to allow discrimination based upon the intensity of the resulting electrophoretic peaks. Analysis of the sequencing data indicated that the read length was found to be 500 bases with an accuracy of 97%.

1.4.2.4. Four Color/Single Lane

The commonly used approach in most commercial DNA sequencing instruments using fluorescence detection is the use of a four color/single lane strategy for identifying the terminal bases in sequencing applications. The primary reasons for using a four color/single lane approach is that it provides high accuracy in the base calling, especially for long reads and also, the throughput can be high due to the fact that all bases comprising the template DNA can be called in a single gel tract. Unfortunately, a four color detector requires extensive optical components to sort the fluorescence and also, in some cases, multiple excitation sources are needed in order to efficiently excite the
fluorophores used to label the individual sequencing ladders. In addition, post-electrophoresis software corrections may be required to account for spectral leakage into detection channels.

Most dye-terminator reads are used with this four color strategy, since the data analysis (base calling) does not depend on uniform incorporation efficiencies, which are hard to achieve using dye-labeled terminators. The same type of instrumentation that is used for four color/dye-primer reads can also be used for four color/dye-terminator reads as well. The only difference one finds is in terms of the sample preparation protocols and also, the software corrections in the sequencing data, such as different mobility correction factors. In most cases, a size exclusion step is used following DNA polymerization to remove excess dye-labeled terminators, since they are negatively charged and can mask the sequencing data due to the presence of a large dye-terminator band in the gel.

1.5. References


Chapter 2

Micro-Capillary Reactors for Direct Sample Introduction into Slab Gels for Solid-Phase DNA Sequencing

2.1. Introduction

Over the past few years, novel DNA sequencing methods and new types of automated detection strategies have been devised. These greatly reduce the effort required for DNA sequencing and data entry. The 373 DNA Sequencer (Applied Biosystems) uses continuous real-time laser-based detection of fluorescent dye-labeled reaction products on the gel and enters sequencing data directly into a computer, thus eliminating the tedious task of reading autoradiographs. A second generation instruments, the ABI 377 DNA Sequencer, Li-COR 4000, the Pharmacia ALF, and Millipore Basestation, use slightly different technology that allows more rapid electrophoresis and data collection than the 373 DNA Sequencer does. Instruments with unispectral detection use either fluorescein-based dyes or in the case of the Li-COR instrument, infrared-emitting dyes excited by a diode laser.

The attractive feature associated with fluorescence in the near-IR (\(\lambda_{ex} > 700\) nm) includes smaller backgrounds observed during signal collection and the rather simple instrumentation required for carrying out ultrasensitive detection. In most cases, the limit of detection for fluorescence measurements is determined primarily by the magnitude of the background produced from scattering and/or impurity fluorescence. This is particularly true in DNA sequencing since detection occurs within the gel matrix, which can be a significant contributor of scattering photons. In addition, the use of denaturants in the gel matrix, such as urea (7 M) or formamide, can produce large
amounts of background fluorescence. The lower background that is typically observed in the near-IR can be attributed to the fact that few species fluoresce in the near-IR. In addition, the $1/\lambda^4$ dependence of the Raman cross section also provides a lower scattering contribution at these longer excitation wavelengths.

An added advantage of near-IR fluorescence is the fact that the instrumentation required for detection can be rather simple and easy to use. A typical near-IR fluorescence detection apparatus can consist of an inexpensive diode laser and single photon avalanche diode (SPAD). These components are solid-state allowing the detector to be run for extended periods of time requiring little maintenance or operator expertise.

Near-IR fluorescence can be a very attractive detection strategy in gel sequencing because of the highly scattering medium in which the separation must be performed. Due to the intrinsically lower backgrounds that are expected in the near-IR compared to the visible, on-column detection can be performed without sacrificing detection sensitivity. Near-IR has been demonstrated in sequencing applications using slab gel electrophoresis where the detection sensitivity has been reported to be ~2000 molecules.$^{1,2}$ The improved detection limit resulted primarily from the significantly lower background observed in the near-IR.

Solid-phase DNA sequencing strategies have developed into important processes for high throughput applications. Some of the important attributes associated with solid-phase approaches that make them particularly appealing include: (1) The solid support can act as a capture medium for the double-stranded template generated using PCR and once captured, the complement can be effectively removed using heat or NaOH melting. In fact, elution of the complement can allow for bidirectional sequencing of one PCR
(2) The removal of excess primers, salts or dNTPs can easily be accomplished by capture of the sequencing template on the solid-support without the need for centrifugation or precipitation steps. Many times the removal of these interferences can improve banding in the electrophoresis step of the analysis, especially in capillary gel electrophoresis. In capillary electrophoresis, electrokinetic injection is biased toward high electrophoretic mobility ions. To increase the amount of DNA injected into capillary column, an effective removal of small ionic species is required. Template DNA has also been shown to interfere with the analysis of sequencing fragments. The large amount of unextended primer left in the sequencing cocktail can produce a bleeding effect of dye into the gel tract, complicating data analysis. (3) Solid-phase sequencing is conducive to automation. Fluid handling is greatly simplified in most solid-phase approaches and many of the steps can simply be handled by conventional robotic workstations.

The common support medium used in many solid-phase approaches is magnetic beads covalently-coated with streptavidin or to ultra-small polystyrene beads. By producing a PCR product using a primer containing a biotin molecule on the 5’ end, the PCR product can be tethered to the bead with magnetic capture to permit isolation of the template for further processing. The beads are typically placed in titer plates or micro-centrifuge tubes and then, situated on a magnet to trap the beads. An alternative method is to use streptavidin agarose gels to immobilize the DNAs.

The detection of the DNA sequencing ladders can be easily accomplished using either fluorescence or radiography. If dye-primer chemistry is implemented, the sequencing reactions must be performed in separate tubes or wells of the titer plate and then combined prior to loading of the gel. However, Schofield and co-workers have also
shown that dye-terminator chemistry can be used,\textsuperscript{22} which permits generating the four terminated products in one reaction vessel, minimizing post-sample handling prior to gel loading. The importance of this is that it reduces the amount of sequencing reagents and magnetic beads required to perform the analysis. In addition, these researchers were able to perform primer walking using solid-phase sample preparation techniques.

Recently, we have extended the solid-phase approach to micro-capillary tubes, which offer the additional advantage of being able to scale the reaction to nanoliter volumes with a concomitant savings arising from the reduction in consumable consumption during DNA polymerization.\textsuperscript{23} In our work, the micro-capillary tube was functionalized with biotin to serve as the anchoring scaffold for the DNA template (see Figure 2.1 A, B). This was followed by incubation of the tube with streptavidin and then, utilizing another binding site of the tetra-dentate streptavidin for attaching the PCR-product containing a biotin handle to the wall of the micro-capillary tube (biotin:streptavidin:biotin layer). The advantage of tethering the biotin directly to the wall of the capillary and not streptavidin is that one can simply strip the DNA from the capillary reactor using heat and formamide without removing the biotin layer. This makes the capillary available for re-immobilization of new sequencing templates by incubating the capillary with streptavidin. The sequencing then followed conventional solid-phase steps (see Figure 2.2.) with the exception that strand melting was accomplished using heat (~95°C) and not NaOH, since strong base was found to remove the biotin scaffold from the wall of the capillary. Using this format, we have demonstrated that cycle sequencing using \textit{Taq} polymerase and a single dye tract can be accomplished in a volume of ~64 nL with the generated sequencing products fractionated
in a capillary gel column directly interfaced to the capillary solid-phase reactor. The detection of the sequencing fragments was accomplished using dye primer chemistry, with the labeling dye consisting of a near-IR dye, which absorbs radiation at 780 nm and fluoresces at 810 nm.

Figure 2.1. Chemical structures of immobilization layer used to tether biotinylated PCR products to fused silica capillaries (A). Also included in this figure are the sequences of the chemical linker of biotin to the PCR product (B).
In this work, we would like to demonstrate the use of this micro-capillary solid-phase DNA sequencing reactor directly interfaced to conventional slab gel electrophoresis. Since the slab gel format can accept larger volumes of materials as compared to capillary gel electrophoresis (sample volume \( \cong 10 \) nL), the reactor volume was scaled-up by simply increasing the length of the reactor and internal diameter. In this case, the reactor was 15 cm in length and 100 \( \mu \)m in diameter, producing a total reaction volume of 1.2 \( \mu \)L making it appropriate for the automated feeding of multi-well slab gels. Following generation of the sequencing ladder, the entire contents of the capillary reactor were pressure injected into a conventional slab gel for fractionation of the sequencing products. In addition, we will show that the anchoring system is stable to
typical cycle sequencing conditions, allowing the implementation of multiple sequencing rounds on a tethered template, making it amenable to primer-walking strategies. In addition, data will be included to demonstrate the feasibility of dye-terminator chemistry as well, which will allow four-color generation of sequencing ladders in a single micro-capillary reactor.

2.2. Experimental Section

2.2.1 Instrumentation

The PCR reactions were performed in a Flexigene thermal cycler (Techne, Minneapolis, MN) (see Figure 2.3). The Flexigene is programmed by means of an integral keypad and LCD display. The agarose gel electrophoresis system (BIO-RAD, Hercules, CA) consists of a gel casting tray, a gel comb, an electrophoresis tank and a Power PAC 300 power supply (see Figure 2.4).

Figure 2.3. Flexigene thermal cycler
Figure 2.4. The agarose gel electrophoresis system which includes a gel casting tray, a gel comb, an electrophoresis tank and a Power PAC 300 power supply.

The Li-COR 4000 automated DNA sequencer (Li-COR, Lincoln, NE) consists of a detection system, electrophoresis apparatus and computer system (see Figure 2.5). Figure 2.6 is an expandal diagram showing the parts of the electrophoresis apparatus.

Figure 2.5. The Li-COR 4000 automated DNA sequencer.
Figure 2.6. Expandal view of electrophoresis apparatus. 1. Upper buffer tank lid; 2. silicone tubing gasket material; 3. upper buffer tank; 4. sequencing comb; 5. 0.25 mm thickness space; 6. rectangular glass front plate; 7. notched glass front plate; 8. left rail assembly; 9. right rail assembly; 10. lower buffer tank lid; 11. lower buffer tank; 12. high voltage cable.
The detection system includes laser diode, detector and scanning platform. The laser diode emits radiation at 780 nm with a peak power rating of 30 mw. Radiation emitted by the excited fluorophores is collected by a microscope detector. Radiation is filtered by a bandpass filter to remove scattered radiation and the Raman scattering due to the water present in the gel. The scanning platform is shown in Figure 2.7. The scan motor drives a belt that moves the microscope assembly from side to side across the entire width of the scanning window. The focus motor moves the microscope perpendicular to the gel apparatus and is used to position the focal plane of the detection microscope precisely onto the gel. The pivot motor swings the entire platform back and forth around the pivot point, which serves to keep the microscope moving in a path to the gel. The output current from the avalanche photodiode (APD) is fed into a lock-in amplifier that discriminates the signal. Electronic filtering options are available to optimize the signal-to-noise ratio. Base ImagIR is a software set containing several distinct programs that are used to collect, analyze DNA sequencing data from the Model 4000 Automated DNA Sequencer.

2.2.2 Preparation of Biotinylated DNA

• DNA Amplification

A 1 kbp target was amplified by PCR from bacteriophage DNA in the Flexigene thermal cycler (Techne, Minneapolis, MN). The PCR primers were designed (see Figure 2.8) to amplify bases 6631-7630 of the λ-DNA target. The PCR mix contained 1 µL of λ-DNA (5 pmol), 10 µL of 10x PCR buffer (20 mM TRIS-HC1; pH = 8.3; 50 mM KC1; 15 mM MgCl₂), 5 µL of a 24 bp biotinylated forward primer (10 µM), 5 µL of a 24 bp reverse primer (10 µM), 0.5 µL of AmpliTaq polymerase (Perkin-Elmer, Gaithersburg,
MD) and 70.5 µL of double-distilled water. Thirty-five PCR cycles were performed using the following temperature (1) denature dsDNA at 94°C for 45 s; (2) anneal primers to template at 66°C for 30 s; (3) extend at 72°C for 90 s.

Biotin-GGA TAT GGG CCG CAG TGA GGA GAA

(A)

GGT TAT CGA AAT CAG CCA CAG CGC

(B)

Figure 2.7. The scanning platform in Li-COR 4000 automated DNA sequencer.

Figure 2.8. Sequences of PCR primers. (A) PCR forward primer (Complement of (–) strand, bases 6631 – 6654). (B) PCR reverse primer (Complement of (+) strand, bases 7538 – 7558).
• **Agarose Gel Electrophoresis**

Agarose (0.4 g) was added into 40 ml of TAE 1X buffer and heated to allow all the agarose (Sigma, Saint Louis, MO) to dissolve. The gel solution was cooled to 60°C and added with Ethidium Bromide (EtBr) (Promega, Madison, WI) to a final concentration of 0.5 µg/mL. The warm gel solution was poured and the comb was placed into the casting tray. The gel was allowed to polymerize for 30 min and placed in the electrophoresis chamber with TAE 1X buffer. Samples with 2 µL loading dye were loaded into the wells. The 100 V was applied to the gel for 30 min. After the electrophoresis was completed, the gel was placed on a UV transilluminator for visualization.

• **PCR Purification**

Following PCR amplification, the biotinylated PCR products were purified using micro-concentrators. This purification step removed biotinylated primer from the PCR product, which would effectively compete for the limited number of binding sites in the capillary reactor. Amicon micro-concentrators (Amicon, Beverly, MA) with molecular cut-offs at 50 kDa was employed. The microconcentrators trap large molecules, > 50 kDa, and allow small molecules, primers, dNTPs and salts, to pass through. The sample reservoir was inserted into a vial. The PCR solution was pipetted into sample reservoir (0.5 ml maximum volume) and sealed with the attached cap. The assembly was placed in a centrifuge and centrifuged at 2900 rpm (14,000 x g) for 15 min. The 0.5 ml of ddH₂O was pipetted into sample reservoir and the assembly was spun for 15 min. To transfer concentrate to the vial, the sample reservoir was placed upside down in a new vial and spun briefly.
The quality of purified PCR product was checked by agarose gel electrophoresis, followed by visualization by UV irradiation of the gel (see Figure 2.9). The concentration of purified PCR product was determined using a Beckman DU-60 spectrophotometer (Beckman Instruments, Fullerton, CA) with the absorption measured at 260 nm. The PCR product was stored in water at -20°C until required for use.

Figure 2.9. Gel image of agarose gel electrophoresis. One percent agarose gel and 100 V potential were used for separation. The ΦX174 Hae III ladder serves as a marker by which the size of the PCR product can be judged.

2.2.3. Immobilization of Biotinylated PCR Product to Capillary Wall

The PCR product was immobilized to the wall of a fused-silica capillary using a biotin:streptavidin:biotin system. Fused silica capillary tubes (100 µm i.d.; 360 µm o.d.) were cut into 100 cm sections and rinsed successively with 1 M NaOH, ddH₂O, 1M HCl for 30 min each. This was followed by purging the capillary with air for 10 min and finally oven-baking the capillary at 200°C for several hours to remove all traces of water which may affect the siloxane formation between the capillary wall and anchoring layer. After drying, the capillary was filled with a solution of 5% (3-aminopropyl)triethoxysilane (Sigma Chemical, St. Louis, MO) in acetone. The capillary was allowed to stand for 30 min and then purged with air for 3 min and finally incubated for ~24 hrs. The capillary was then filled with a bicarbonate solution (50 mM; pH = 8.3)
containing 5.0 mg/mL NHS-LC-biotin (Sigma) for 4 hr at room temperature. Following this treatment, the capillary was briefly rinsed with ddH₂O and gravity filled with a 4.0 mg/mL solution of streptavidin (Sigma) prepared in 50 mM phosphate buffer (pH = 7.4) at 4°C. Finally, the streptavidin-coated capillary was rinsed with water and treated (by gravity flow) with the biotinylated PCR product (~1.0 µM) at 4°C. Excess PCR product was removed using a final water wash and the DNA-coated capillary column was stored in a refrigerator at 4°C until required for use.

### 2.2.4. Radioactive Labeling of Biotinylated DNA

The PCR product was radiolabeled at the 3' terminus with [γ⁻³²P]dATP (Promega, Madison, WI). Ten pmoles of DNA was suspended in 57 µL of ddH₂O and 20 µL of 10X terminal deoxynucleotidyl transferase (TDT) buffer. To the DNA-buffer solution, 8 µL of [γ⁻³²P]dATP (800Ci/mmole, 10mCi/mL) and 5 µL of the TDT (15 U/µL) were then added. The reaction mixture was incubated at 37°C for 30 min. The action of TDT enzyme was halted by heating the mixture to 65°C for 10 min. Unincorporated labeled dATP was removed using MicroSpin™ G-25 columns (Amersham Pharmacia Biotech, Piscataway, NJ). Sample was applied to the center of the angled surface of the compacted resin bed. The column was spun for 2 min. The purified sample was collected. The efficiency of ³²P was determined by an assay using a DNA binding filter. 1 µL of the ³²P DNA was spotted onto a circular filter. The filter was dried under a heat lamp and was washed twice in 0.5 M sodium phosphate to remove the unincorporated nucleotides. The radioactivity was measured by submerging the filter into a vial containing scintillation fluid (3 mM p-terphenylin in toluene) and monitoring the
radioactivity using a LS6000IC series scintillation counter (Beckman Instruments, Fullerton, CA).

2.2.5. The Rate of DNA Immobilization

In order to measure the rate of DNA immobilization, the $^{32}$P DNA was used. The immobilization protocol was the same as described above. A 100 cm fused silica capillary tube (100 µm i.d.; 360 µm o.d.) tethered with streptavidin was filled with $^{32}$P biotinylated DNA by pressure. Five cm sections were cut from the long column at different times. The rate of streptavidin immobilization was also measured by treating biotin coated capillary tubes with streptavidin for different times. The streptavidin coated columns then were filled with $^{32}$P biotinylated DNA overnight. To measure the rate of protein denaturation, the $^{32}$P DNA coated capillary tubes were treated with formamide at 95°C, then rinsed with water.

Scintillation measurements of the $^{32}$P labeled DNA in the micro-capillary reactor were performed by submerging the sealed reactor in scintillation liquid (3 mM p-terphenyl in toluene). Measurements were performed on a LS6000IC series scintillation counter (Beckman Instruments, Fullerton, CA).

2.2.6. Near-IR Fluorescence Sequencing with Dye-labeled Primers

Extension of the immobilized PCR product was accomplished directly within a 15 cm section of micro-capillary (total volume = 1.2 µL). Since the original PCR product was immobilized as the duplex, it was thermally denatured at 95°C for 5 min with constant gravity flushing to remove the complement. For dye-primer chemistry, four reactors were used, one for each base. Each solid-phase capillary reactor was filled with the appropriate extension mix, which consisted of 1.0 µL of a 21 bp IRD-800 labeled
sequencing primer (LI-COR, INC., Lincoln, NE); 5 µL of ddH₂O and 2 µL of the A, C, G, or T extension mix (Thermo sequenase primer cycle sequencing kit with 7-deaza-dGTP, Arnersham, Arlington Heights, IL). The sequencing primer was designed to hybridize to bases 7538-7558 of bacteriophage λ-DNA (see Figure 2.10).

![Figure 2.10](image)

**Figure 2.10.** Structure of IRD800 dye labeled sequencing primer.

To provide a sufficient population of extension products to aid in fluorescence detection, the capillary reactor was cycled through 20 temperature cycles, with each cycle consisting of the following steps: (1) 95°C for 30 s; (2) 68°C for 30 s; (3) 72°C for 90 s. The thermal cycling was accomplished by submerging the capillary reactor, which was capped on both ends with a quartz capillary plug (Micro-quartz, Phoenix, AZ), in a temperature-controlled water bath. Following thermal cycling, the reactor was immediately heated to 95°C to remove extension products from the wall immobilized
sequencing template and the reaction mixture from each capillary was loaded directly into one of the wells of a polyacrylamide slab gel using pressure injection.

2.2.7. The Stability of the Template Anchored to the Capillary Wall

In order to investigate the stability of the anchor scaffold to typical thermal cycling conditions, several capillary micro-reactors were subjected to 8 thermal cycles using the temperature program described above. The tethered template was subjected to 15 rounds of sequencing with each round analyzed on a LI-COR automated DNA sequencer. For these experiments, only a single terminator tract was analyzed (ddCTP).

2.2.8. Fluorescence Dye-terminator Sequencing in a Micro-capillary Reactor

A near-IR dye terminator was kindly prepared for us by Shaheer Kahn of PE-Applied Biosystem. The dye label (see Figure 2.11) in this case was a sulfonated tricarbocyanine that was prepared in our laboratories. This particular dye label has absorption and emission maxima that are easily accommodated by the detection optics and diode laser situated in the LI-COR automated sequencer.

The capillary reactor was filled with the extension mix which in this case consisted of 1.0 µL of a 21 bp sequencing primer (1.0 µM), 2 µL of the reaction buffer, 1.0 µL of the dNTP mix, 1.0 µL of the thermal sequenase enzyme (dye terminator cycle sequencing kit, Amersham) and finally, 1.0 µL (50 µM) of the dye-labeled terminator (ddGTP). The reaction mixture was pressure injected into the capillary and then, the capillary was sealed with the quartz plugs and subjected to the following thermal cycling conditions (20 cycles); (1) 95°C for 30 s; (2) 60°C for 4 min. Following thermal cycling, the contents of the capillary reactor was pressure injected into a micro-centrifuge tube and subjected to a cold ethanol precipitation to remove any excess dye-labeled ddGT.
After decanting off the ethanol, the DNA was spun and dried in a centro-vap and finally, taken up in 2 µL of a formamide loading buffer. One µL of the solution was loaded onto a polyacrylamide slab gel.

**Figure 2.11. Structure of NIR dye-labeled ddGTP**

2.2.9. Slab Gel Electrophoresis and Data Collection

The sequencing was performed on a Li-COR model 4000 automated DNA sequencer (Li-COR, Lincoln, NE). The sieving matrix was an 8%T (w/v) cross-linked gel (FMC Bioproducts Long Ranger, Rockland, ME) that contained 7.0 M urea as the denaturant and 1X TBE (pH 8.0). Polymerization of 30 ml of the 8% gel was initiated by adding 200 µl of 10% (w/v) ammonium persulfate and 20 µl 0f TEMED, and a comb was inserted between the glass plates to form wells. After 3 hrs, the comb was removed and the plates placed in the sequencing instrument and the buffer reservoirs filled with 1X TBE. The gel was heated to 50°C and prerun at 1650 V for 30 min. One µl of each
sample was loaded into the sample wells and the electrophoresis was run at 1650 V for 8 hours. Data acquisition and base-calling were performed using the LI-COR software loaded onto a Gateway computer.

2.3. Results and Discussion

Sequencing was carried out using the micro-capillary solid-phase approach using single dye-primer chemistry and four separate lanes of the gel, since our automated DNA sequencer can process only one color. In this case, the single base tracts prepared in each of four capillaries were pressure injected into the gel lanes of the sequencer. In Figure 2.12 is shown the slab gel electropherogram of the sequencing fragments. A trace of the sequencing data as well as the called bases is shown in Figure 2.13. Using the base caller for this machine, we were able to read 589 bases with the number of ambiguities from this trace determined to be < 10. Even after base 589, strong signal intensity was observed. In a similar fashion, we generated sequence data on this same template using sequencing reactions prepared in conventional micro-centrifuge tubes and a block thermal cycler. The results obtained (data not shown) indicated a similar read length with the errors (in this case, defined as the number of ambiguities) equal to ~20. Therefore, the capillary micro-reactor can produce high quality sequencing data, which is comparable to conventional sample preparation formats. Improvements could easily be achieved using this micro-capillary solid-phase sequencing format by improving electrophoresis conditions, either through improvements in the sieving gel or the electrophoresis operating parameters, since the signal intensity was strong even beyond the last readable base in the electropherogram.
Figure 2.12. Slab gel electrophoresis of sequencing fragments of a 1 kb PCR amplified λ-DNA immobilized to the wall of a micro-capillary reactor. Gel was 8% (w/v) polyacrylamide gel with 1.0 µl of sample loaded into each well of a Li-COR 4000 automated sequencer. The electrophoresis was run at 1650 V for 8 hours.
Figure 2.13. Sequencing trace (called bases 1-589) of a 1000 bp PCR amplified λ-bacteriophage template immobilized to the wall of a capillary micro-reactor. The sequencing was performed using an IRD800 NIR fluorescent label covalently attached to the custom-designed 21mer primer. In this sequencing run, a single dye was used with four sequencing lanes. The electrophoresis was performed on a Li-COR 4000 automated slab gel electrophoresis device with near-IR laser-induced fluorescence detection.
Our next set of experiments was focused on investigating the stability of the anchoring system and typical cycle sequencing conditions. In Figure 2.14 are shown images of single-base tracts (C-tract) taken from our automated sequencer as a function of sequencing round. For each round, a total of 8 thermal cycles were performed on a single capillary solid-phase reactor as described in the experiment section. Between each sequencing round, the capillary reactor was washed with copious amounts of buffer to remove any excess reagents potentially remaining from a previous sequencing round. As seen from this data, even after 15 sequencing rounds the signal is clearly observable. In panel B of Figure 2.14 is shown the signal intensity (peak height, see Figure 2.14 for band monitored) as a function of sequencing round. The relative signal intensity was found to drop by only ~40% after 120 thermal cycles (15 sequencing rounds and 8 thermal cycles per round). The impressive chemical stability of this tethering system arises from both the stability of the biotin:streptavidin couple and also, the ability of the siloxane bonds used to attach biotin to the wall of the capillary reactor to withstand the heating conditions required for cycle sequencing. We have also found that while this tethering system is stable toward typical thermal conditions used in melting duplex DNAs, it shows degradation when subjected to NaOH melting due to hydrolysis of the siloxane bond with high pH. In addition, we have found that the streptavidin, immobilized template DNA can be stripped from the reactor by subjecting the capillary to heat (95°C and 90% formamide) (see Figure 2.15). Under these conditions, the streptavidin protein is severely denatured, disrupting the biotin:streptavidin complex. However, the wall-immobilized biotin is not removed from the capillary using these conditions, since incubation with fresh streptavidin and addition of biotinylated DNA will
'reactivate' the capillary reactor. After 15 rounds, we also performed a four-lane sequence analysis on the immobilized template and found the effective read length was comparable to that obtained on round 1.

Figure 2.14. Gel image of C-tract subjected to 15 sequencing rounds on a single A-bacteriophage template immobilized to the solid-phase capillary micro-reactor. In the accompanying plot, the normalize fluorescence signal intensity from the band marked (see asterick) in the gel image is monitored a function of sequencing round.
Figure 2.15. Radioactivity (cpm) of $^{32}$P labeled DNA immobilized to the wall of a micro-capillary reactor as a function of denaturation time. A 5 cm 100 µm capillary tube coated with $^{32}$P labeled DNA was treated with a 98% formamide solution at 95 °C.

We also investigated the kinetics of the DNA and streptavidin immobilizations. This has been substantiated by using $^{32}$P-labeled DNA and performing scintillation measurements at different times. The amount of immobilized DNA was determined from the scintillation intensity and the specific activity of $^{32}$P-labeled DNA. The specific activity was determined to be $2.0 \times 10^6$ cpm/pmol by measuring the scintillation intensity of known DNA concentration. The In Figure 2.16 is shown the rate of DNA immobilization onto the capillary wall. The rate of streptavidin immobilization is shown in Figure 2.17. As can been seen, the maximum coverage was achieved in ~1 hr for both DNA and streptavidin immobilizations. As mentioned above, we can successfully strip the streptavidin from the capillary reactor using heat and formamide without removing the covalently linked biotin from the wall of capillary, which takes 2 min. This makes it
available for immobilization of a new DNA template. From the kinetics of the DNA and streptavidin immobilization, we find it takes 2 hours to immobilize the new template onto the reactor. The innate reason for the strong interaction between biotin and streptavidin is not yet known. It is a reasonable hypothesis that biotin and streptavidin participate in hydrophobic interaction or charge-transfer complex. The immobilization processes should be diffusion control. The smaller diameter capillary is expected a higher immobilization rate. Therefore it would take less than two hours to immobilize a new DNA template onto the wall of a capillary.

![Graph](image-url)

**Figure 2.16.** The rate of DNA immobilization onto capillary wall. A 100 cm capillary tube (100 µm) tethered with streptavidin was filled with \(^{32}\text{P}\) biotinylated DNA by pressure. The 5 cm columns were cut from a long column at different times and inserted into the scintillation vial for determining the radioactivity. The left Y-axis represents the count of radioactivity of the \(^{32}\text{P}\) DNA immobilized onto the capillary wall. The right Y-axis represents the amount of DNA immobilized onto the capillary wall.
Figure 2.17. The rate of streptavidin immobilization onto capillary wall. A 100 cm capillary tube (100 µm) tethered with biotin was filled with streptavidin solution by pressure. The 5 cm columns were cut from a long column at different times. The column was filled with $^{32}$P DNA overnight and inserted into the scintillation vial for determining the radioactivity. The Y-axis represents the count of radioactivity of the $^{32}$P DNA immobilized onto the capillary wall.

One of the disadvantages of the dye-primer chemistry used in this particular format is that the DNA polymerization must be done in four separate reactors and pooled prior to the electrophoresis if four unique fluorescent probes can be utilized (four colors). Therefore, it would be particularly advantageous if dye-terminator chemistry could be implemented allowing the sequencing reaction to be performed in a single capillary micro-reactor. We thus initiated experiments on evaluating the feasibility of using near-IR labeled dye terminators for this application. These experiments were carried out by
accumulating data for a single tract (G-tract). The results of this sequencing reaction are depicted in Figure 2.18. As can be seen from this data, the quality of the G tract is high and the effective read length (determined by counting the number of bands in the electropherogram and multiplying by 4) was found to be ~566 bases. However, this should be taken as only a conservative estimate, since the use of a four-color sequencing run may likely lower the read length compared to that projected here. In this case, we found it necessary to remove excess dye-terminator prior to gel load using a cold ethanol precipitation since the dye-terminator migrated at a rate equivalent to the 110 bp DNA fragment. This produced a large band in the electropherogram, which masked the signal from several bands in the DNA ladder. However, the need for an ethanol precipitation step could be alleviated if all extension products could remain captured on the wall of the capillary micro-reactor (only one thermal cycle). Simple pumping of clean buffer into the capillary prior to heat denaturation would allow effective removal of the excess dye-labeled ddNTP prior to gel loading. Improving the loading level of templates to the capillary, either through increasing the surface area or increasing the immobilization efficiency, would allow for lowering the number of thermal cycles during the DNA polymerization process.

2.4. Conclusions

We have demonstrated the ability to perform micro-sequencing in capillary tubes containing wall-immobilized DNA templates (solid-phase sequencing). Using a four-lane, single color dye primer sequencing run, a read length was obtained (589 bases) that was comparable to a conventional sample prep format. One of the advantages of this format is that the reactor can be scaled to ultra-small volumes, making it appropriate for
Figure 2.18. Gel image of a single G tract with the dye-terminator chemistry. The terminator (ddGTP) was labeled with a near-IR fluorescent dye. The electrophoresis was performed on the Li-COR 4000 automated DNA sequencer.
loading sequencing samples directly to micro-separation platforms, such as ultra-thin slab gels or capillary gels. This could potentially produce a significant reduction in consumable consumption, reducing the cost of acquiring sequencing data. In addition, the stability of the anchoring system toward typical thermal cycling conditions makes the template available for multiple sequencing rounds. This will allow the implementation of primer walking across a single DNA template. While we have demonstrated the ability to perform cycle sequencing in these micro-capillaries, the use of a single thermal cycle will offer some potentially attractive features, especially when using dye-tenninator chemistry. For example, after a single thermal cycle, excess reagents, in particular dye-ddNTPs, can be removed from the sequencing products without requiring the need of a cold ethanol precipitation, simplifying sample preparation. Another attractive feature of this solid-phase sequencing strategy is the potential to integrate the micro-reactor to capillary-based gel loaders. Recently, a pneumatic device has been described, which uses arrays of capillaries to transfer samples from titer plates to the wells of high density gels to automate the loading process of slab gels\textsuperscript{26}. One can envision the use of a capillary containing immobilized single-stranded templates that would not only load the wells of the automated DNA sequencer, but also prepare high quality sequencing ladders, expanding on the existing capabilities of the capillary-based gel loader.

In addition, the capillary-based micro-reactor can be integrated to other separation platforms as well such as capillary gel electrophoresis, which will offer some particularly attractive features to this new technology. For example, a controlled amount of DNA (determined by the size of the capillary micro-reactor) can be loaded onto the gel column to prevent the detrimental effects (column failure) associated with overloading. Also,
tethering the template to the wall of the micro-reactor would eliminate the loading of template onto the gel column, which has been shown to severely degrade the quality of sequencing data obtained in capillary gel electrophoresis.⁶

2.5. References


Chapter 3

DNA Sequencing Reactions Performed in a Nanoreactor Directly Coupled to Capillary Gel Electrophoresis

3.1. Introduction

Capillary electrophoresis (CE) has become a valuable tool for the high speed and efficient separation of such analytes as pharmaceuticals, \(^1\), \(^2\) environmental pollutants, \(^3\), \(^4\) oligonucleotides \(^5\)-\(^10\) and proteins. \(^3\), \(^11\) The popularity of CE results primarily from the high speed in which separations can be carried out as well as the high plate numbers that can be generated. The speed of the separation is produced by using high electric fields and, in some cases, short column lengths. In order to further increase the speed of the separation, miniaturization of the column, both in length and internal diameter, will be required. As a result, injection volumes will need to be reduced as well in order to minimize zone variance introduced by the finite injection volumes placing severe demands on detection. In order to meet the challenges placed on detection by the small injection volumes required for CE, fluorescence, in particular laser-induced fluorescence (LIF), has become the detection protocol of choice.

The pioneering work of LIF in CE was carried out by Zare and Gassman \(^12\) who separated and detected femtomole quantities of racemic mixtures of amino acids labeled with a fluorescent dansyl group. In this work, a He-Cd laser (442 nm excitation) was used to excite the fluorescence of the labeled amino acids. Since then, many researchers have demonstrated the utility of LIF detection in CE. The common LIF strategy is to use labeling dyes that absorb radiation in the visible region of the spectrum (350-650 nm), due primarily to the readily available instrumentation, such as excitation sources like He-
Cd, Ar ion, He-Ne and Kr ion lasers. In addition, most photon transducers (photomultiplier tubes, PMTs) possess high single photon detection efficiencies in the visible region. And finally, fluorogenic labels that absorb radiation in the visible range for a variety of different functional groups (primary amines, alcohols, and thiols) are readily available from commercial sources.

Some of the disadvantages associated with visible LIF detection include the expensive and sophisticated instrumentation required to carry out detection. For example, ion lasers, such as Ar or Kr, are large lasers that can place heavy demands on utilities and have relatively short operational lifetimes, requiring the replacement of expensive ion tubes. Another issue is the signal-to-noise ratio (SNR) in the measurement. Many times, the use of visible excitation can produce large backgrounds in the form of impurity fluorescence (autofluorescence) or scattering from the sample matrix. This is particularly true in capillary gel electrophoresis, where the sieving gel can produce a large scattering background. Indeed, formats using off-column detection such as the sheath flow cell have been implemented to reduce this background problem.

Recently, researchers have been interested in the near-infrared (NIR) region for LIF detection because the near-IR possesses many advantages over visible fluorescence primarily due to photoprocesses of absorption and emission occurring above 700 nm. In the near-IR, background interference from impurity molecules can be reduced or even eliminated because few molecules demonstrate intrinsic fluorescence in this region. Also, since the amplitude of Raman or Rayleigh scattering is inversely proportional to the power of the excitation wavelength \(1/\lambda^4\), enhanced sensitivity can be achieved because of significant reductions in scattering effects. Moreover, with a major push
towards miniaturization of CE instrumentation, the NIR is a viable detection strategy because of the emergence of semiconductor lasers and detectors, which have a compact size and lower cost compared to gas ion lasers and PMTs. Additionally, semiconductor lasers provide sufficient light power for ultra sensitive measurements.

Applications using near-IR-LIF detection in CE have grown over the past decade due to the increased available chromophores and semiconductor devices appropriate for near-IR fluorescence. While many commercial sources currently exist from which laser diodes and diode detectors can be purchased to allow construction of near-IR detectors appropriate for CE, work continues toward increasing the availability of labeling dyes that target particular functional groups of the analyte. In addition, modification in the CE operating conditions to accommodate the photophysical and chemical properties of the near-IR labeling has been and continues to be a focus area.15

LIF detection coupled to CGE has become an attractive detection strategy for many DNA sequencing applications due to its ultrahigh sensitivity.6,18,19 Detection limits ranging from 6,000-600,000 molecules have been reported using visible LIF detection.9,20 Near-IR fluorescence offers a viable alternative to visible fluorescence detection in CGE due to the intrinsically lower background levels that are observed in the near-IR when compared to the visible. To highlight the intrinsic advantages associated with the use of near-IR fluorescence detection in capillary gel DNA sequencing applications, a direct comparison between laser-induced fluorescence detection at 488 nm excitation and 780 nm excitation has been reported.21 In this study, a sequencing primer labeled with FAM or a near-IR dye was electrophoresed in a capillary gel column and the detection limits calculated for both systems. The results indicated that the limits of detection for
the near-IR case were found to be $3.4 \times 10^{-20}$ moles, while for 488 nm excitation, the limit of detection was $1.5 \times 10^{-18}$ moles. The improvement in the limit of detection for the near-IR case was observed in spite of the fact that the fluorescence quantum yield associated with the near-IR dye was only 0.07, while the quantum yield for the FAM dye was $\sim 0.9$. To demonstrate the utility of near-IR fluorescence in actual DNA sequencing runs using CGE, Williams and Soper\textsuperscript{10} performed CGE using an M13mp18 plasmid and a single-lane, single-fluor peak height identification protocol with ddNTP concentration ratios of 4:2:1:0 (A:C:G:T) used during DNA polymerization. In this study, a near-IR dye-labeled primer was used with excitation at 780 nm.

Several groups have made advances in data throughput by increasing the speed of DNA separations using capillary gel electrophoresis (CGE).\textsuperscript{22-25} The principle advantages associated with these electrophoretic formats is that the surface-to-volume ratio is much larger than in conventional slab gel electrophoresis (SGE), and as a result, higher electric field strengths can be used to drive the separation, decreasing the time required to effectively separate the DNA. Improvements in resolution are also encountered when separations are performed in capillaries versus slab gels. Reports have detailed a nearly 3-fold improvement in resolution and a 25-fold increase in the speed of the separation in CGE compared to SGE due primarily to the ability of CGE to more effectively dissipate Joule heat, allowing operation at high electric field strengths.\textsuperscript{18} However, in both capillary gel and microchip electrophoresis techniques, the amount of sample required for analysis ranges in the nanoliter regime, which could potentially offer significant savings in the use of costly reagents required for preparing the sequencing fragments.
While advances in separation, detection, and base calling have been numerous, they have outpaced the development of methodologies for preparing DNA-sequencing fragments in a volume more commensurate with the microcolumn separation techniques used to fractionate these DNAs. Standard DNA-sequencing protocols generate products in the microliter range, therefore, not exploiting the low (nL) sample requirements for capillary and microchip separations. Low consumable costs for any sequencing protocol are of critical importance, especially if it is to be financially feasible to sequence large genomes, such as the human genome.

Solid-phase DNA-sequencing methods possess several advantages which could potentially facilitate their integration into a miniaturized DNA-sequencing system. Solid-phase DNA sequencing micro-capillary reactors have been described in Chapter 2. The work presented herein describes the operational characteristics and use of a miniaturized solid-phase DNA polymerase nanoreactor which provides a significant reduction in the amounts of reagents and other consumables used in the preparation of Sanger, dideoxynucleotide-terminated DNA-sequencing fragments. In addition, the nanoreactor can be directly coupled to a capillary gel column for DNA separation, providing on-line analysis of sequencing fragments in an automated fashion on a nanoliter scale. The reactor consists of a conventional fused-silica capillary tube ($V = 62$ nL), in which biotin is covalently attached to the wall with the subsequent addition of streptavidin to provide an anchoring point for a biotinylated DNA template prepared using the polymerase chain reaction (PCR). The preparation of sequencing fragments directly in the nanoreactor will also be demonstrated. To provide the necessary temperature conditions for denaturation, annealing, and chain extension, the nanoreactor was placed in a thermal cycler which
consisted of a small chamber with heating and cooling accomplished using air. Due to the low thermal mass associated with the capillary tubes and the rapid convection associated with moving air, efficient thermal equilibration was achieved in short time periods allowing fast temperature transitions reducing the time necessary to prepare the DNA sequencing ladders.26-29

3.2 Experimental Section

3.2.1. Spectroscopic Analysis of IRD800 Dye Primer and Filters

The excitation and emission spectra of IRD800 dye primer were collected on a Spex fluorometer (Spex, Edison, NJ) at a dye primer concentration of 2 µM. The spectrofluorometer contained a red-sensitive photomultiplier tube (R636, Hamamatsu Corp.) and the emission gratings blazed for 780 nm. In the case of emission scans, the excitation wavelength was 780 nm. In the case of excitation scans, the emission wavelength was 810 nm (see Figure 3.1).

The transmission spectra of the longpass and bandpass filters were acquired on a Ultraspec 4000 UV/Visible Spectrophotometer (Biochrom Ltd., Cambridge, England) (see Figure 3.2).

3.2.2. DNA Thermal Cycler

The capillary oven was designed to provide a solution to two major problems: rapid thermal cycling; good temperature accuracy and stability. Since the working temperature range was typically between 55 °C to 95 °C, poly(vinylchloride) (PVC) was chosen as the oven chamber material since it provided reasonable thermal insulation, moderately low thermal capacity and was very machinable. The oven chamber measured ~7 cm in length and 7 cm in width (see Figure 3.3). The capillary tubes were fed through
Figure 3.1. Excitation and emission spectra of the IRD800 dye primer at a concentration of 2 µM in H₂O. The excitation is shown in black and the emission in red. The data were collected on a Spex Fluorometer using a Xe lamp for excitation. Excitation and emission slit widths: 2 mm.
Figure 3.2. Transmission spectra of filter system used for near-IR sequencing experiments. The bandpass filter for laser source is shown in red, the bandpass for fluorescence in blue and the longpass for fluorescence in black. The data were collected on a Ultraspec 4000 UV/Visible Spectrophotometer.
the chamber via holes drilled into two ends of the oven and could accommodate 12 capillaries. Temperature-controlled air moved through the floor of the chamber through a 2.54 cm diameter hole. The chamber also contained two baffles on either side of the capillaries, which allowed the air to circulate uniformly around the capillaries. The air exited the chamber through 1/8 in. holes drilled on opposite sides of the baffles.

**Figure 3.3.** The picture of the rapid air thermal cycler housing the capillary nanoreactor used to prepare sequencing fragments.

The chamber was heated or cooled via forced air, which was shuttled through the chamber using a high-speed fan. By running the forced air over heating coils (resistively heated), hot air was generated to heat the chamber and the chamber was cooled by running room temperature air into the insulated chamber. The temperature was sensed at the oven inlet with a miniature type K thermocouple. Due to the transport lag of the system, a proportional-plus-derivative control was used to prevent temperature overshoot and ringing. The thermal cycler was found to possess a heating rate of 15°C/s and a cooling rate of 12°C with minimal temperature overshoots.

During preparation of the sequencing fragments, the reactor was flanked on each side by two 20-cm lengths of fused-silica tubing which were linked to the reactor via zero
dead volume glass capillary connectors (MicroQuartz, Phoenix, AZ). The flanking capillaries were coated with a 2% linear polyacrylamide in order to minimize the electroosmotic flow, which allowed the reagents to be delivered into the reaction vessel from the cathodic end of the system. Following primer extension, the gel-filled separation column was attached directly to the reactor using the connector with the removal of one of the flanking LPA-coated tubes (see Figure 3.5). The DNA sequencing

![Diagram](image)

**Figure 3.5.** Schematic diagram of CE system with nanoreactor and capillary gel column. The nanoreactor was connected to the flanking LPA capillary, and the gel column using zero dead volume unions. The flanking capillary was coated with a 2% linear polyacrylamide to reduce the magnitude of the electroosmotic flow, was 20 cm in length, and possessed an inner diameter of 75 µm. The gel column was a 3%T/3% C cross-linked polyacrylamide gel with 7 M urea as the denaturant. The inner diameter of this column was 75 µm with a total length of 70 cm, 40 cm to the detection window.
fragments generated in the nanoreactor were then injected into the gel column for separation by applying a negative potential at the cathodic end of the LPA-coated flanking capillary (see Figure 3.5).

3.2.3. Near-IR-LIF System for CGE Analysis

Figure 3.6 shows a representation of the custom built laser-induced fluorescence capillary gel electrophoresis system. The 10 mW excitation beam (780 nm) was supplied by the GaAlAs diode laser (LT024MD/MF, THORLABS Newton, NJ) and spectrally filtered by a bandpass filter (CWL = 780 nm, HBW = 10 nm, Omega Optical, Brattleboro, Vermont). The laser beam was focused to a 10 µM diameter spot inside the capillary ($1/e^2$). The fluorescence was collected with a microscope objective (NA = 0.85) (Nikon, Natick, MA) and spatially filtered through a slit with a width set at 0.6 mm. After spatial filtering, the emission was spectrally filtered with a longpass filter (CWL = 800 nm, Omega Optical, Brattleboro, Vermont) and a bandpass filter (CWL = 820 nm, HBW = 10 nm, Omega Optical, Brattleboro, Vermont). The emission was then focused onto the face of the photodetector using a 20X microscope objective. The photodetector was a single photo avalanche diode (SPAD) (SPM 200 B Single Photo Detector, PicoQuant, Berlin, Germany). The output of the SPAD was passively quenched and interfaced to a PC which contained a counting board (CT101, Cyber Research, Brand, CT) with the data acquisition software written using Labview.
Figure 3.6. Near-IR system for CGE analysis. BP1, 780 nm bandpass; DC, dichoric; MO, microscope objective; C, capillary tube; BD, beam bump; SF, spatial filter; LP, 800 nm longpass filter; BF2, 820 nm bandpass filter; SPAD, single photo avalanche diode; DISC, discriminator; PC, computer.

3.2.4. Solid-Phase Nanoscale Sequencing Reactions

The DNA was immobilized to the wall of a fused-silica capillary tube using a biotin-streptavidin-biotin system by following the same procedure described in section 2.2.2. The extension of the immobilized DNA was performed directly inside the nanoreactor. The reactor was filled with extension mix, which consisted of 1.0 µL of a 21 bp IRD-800 labeled sequencing primer (LI-COR, INC., Lincoln, NE); 5 µL of ddH₂O and 2 µL of the C extension mix (Thermo sequenase primer cycle sequencing kit with 7-deaza-dGTP, Arnersham, Arlington Heights, IL). To build a sufficient population of extension fragments to aid in detection for the CGE experiments, the temperature of the reactor was cycled 10 times with each cycle consisting of 92°C for 30 s, 55°C for 30 s and 72°C for 90 s. The matrix used for separation in the capillary format was a 3% T/3%
C polyacrylamide gel containing 7 M urea as the denaturant. The field strength was set to 200 V/cm during electrophoresis and the running buffer consisted of 1× TBE (pH 8.6). The capillary gel column possessed a total length of 75 cm, with an effective length of 50 cm (injection to detection).

3.3. Results and Discussion

In order to measure the limits of detection of the LIF-CGE system for near-IR dye labeled primer, the IRD800 labeled primer was electrophoresed in a capillary column containing of a 3% T/3% C polyacrylamide gel. The resulting electropherogram is shown in Figure 3.7. To simplify the calculation, the field strength (200 V/cm) for separation was kept the same as that for injection. The apparent electrophoretic mobility can be calculated from the experimental data using Equation 3.1 which is shown here;

\[ \nu = \frac{L_{\text{eff}}}{t_{\text{mig}}} \]  (3.1)

where \( L_{\text{eff}} \) is the capillary effective length, \( t_{\text{mig}} \) is the migration time. The apparent mobility was found to be \( 1.95 \times 10^{-2} \text{ cm/s} \) (\( 1.95 \times 10^{-3} \text{ dm/s} \)). The injection volume can be calculated using Equation 3.2;

\[ V = \nu \cdot t_{\text{inj}} \delta \]  (3.2)

where \( V \) is injection volume, \( t_{\text{inj}} \) is injection time, and \( \delta \) is the cross-sectional area.

The summary of the data is presented in Table 3.1. The detection limit was found to be \( 2.2 \times 10^{-20} \text{ mol} \) (SNR = 3) for the near-IR dye labeled primer, while the online detection limit for the fluorescein analog was \( 1.5 \times 10^{-18} \text{ mol} \) (SNR = 3).\(^{10}\) Comparison of the on-column detection limits indicated a significant improvement for the near-IR dye primer. In spite of the lower quantum yield exhibited by the near-IR dyes, we were able...
Figure 3.6. CGE separation and on-column detection of a $5 \times 10^{-12}$ M solution of IRD800 labeled M13mp18 universal primer. The laser power was 5 mW. The sample was electrokinetically injected onto the gel column for 30 s at 15 KV. The field strength was 200 V/cm. The gel column consisted of a 3% T 3% C polyacrylamide matrix.

Table 3.1. Signal-to-noise ratio of near-IR dye M13mp18 universal primer with on-column laser-induced fluorescence detection

<table>
<thead>
<tr>
<th>Concentration (M)</th>
<th>$5 \times 10^{-12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration Time (s)</td>
<td>2560.6 (s)</td>
</tr>
<tr>
<td>Mobility (dm/s)</td>
<td>$1.95 \times 10^{-3}$</td>
</tr>
<tr>
<td>Injection Volume (L)</td>
<td>$2.58 \times 10^{-8}$</td>
</tr>
<tr>
<td>Moles Injected</td>
<td>$1.29 \times 10^{-19}$</td>
</tr>
<tr>
<td>Net Signal (cps)</td>
<td>$1810.8$ ($12318.3 - 10507.7$)</td>
</tr>
<tr>
<td>Background(cps)</td>
<td>10507.7</td>
</tr>
<tr>
<td>SNR $^c$</td>
<td>17.7</td>
</tr>
<tr>
<td>Detection Limit (moles, SNR = 3)</td>
<td>$2.2 \times 10^{-20}$</td>
</tr>
</tbody>
</table>

The primer was excited at 780 nm; gel column (3%T3%C). Sample was injected onto the column for 30 s at 15 KV and separated at a field strength of 200 V/cm. The split width was set at 0.6 mm with a laser power 5 mW. $^a$Net signal was calculated as peak height (cps) minus the mean of the average background.
to realize a ∼60 fold improvement in detection sensitivity, primarily resulting from the significantly smaller background observed in the near-IR region. This detection limit is comparable to our previous reported results.\textsuperscript{10}

Integration of the nanoreactor directly to a capillary gel column for sequencing requires careful consideration to zone broadening due to extracolumn effects because of the stringent requirements on separation efficiency in sequencing applications in order to achieve single-base resolution. In the present case, the finite volume of the reactor (injection volume, $V_{\text{inj}}$, for a 20 µm i.d. × 20 cm column, $V_{\text{inj}} = 62$ nL) and the dead volume associated with the connector can potentially result in the reduction in the plate numbers. To assess the contributions of these two variances to the total variance, experiments were carried out to evaluate the relative contribution from the connector ($\sigma^2_{\text{con}}$) and reactor volume ($\sigma^2_{\text{inj}}$). This extra column variance ($\sigma^2_{\text{xc}}$) can be calculated from the following expression, assuming that the zone variance from the finite injection volume could be calculated from $l^2/12$,\textsuperscript{32} where $l$ is the length of the injection plug;

$$
\sigma^2_{\text{xc}} = \frac{\mu_{\text{gel}}^2}{12\mu_{\text{fs}}^2} \frac{r^4_{\text{nr}}}{r^4_{\text{sc}}} l^2_{\text{nr}} + \sigma^2_{\text{con}}
$$

where $\mu_{\text{gel}}$ is the electrophoretic mobility of the oligonucleotide in the gel column, $\mu_{\text{fs}}$ is the free solution mobility of the oligonucleotide (4.0 × 10\textsuperscript{-4} cm\textsuperscript{2}/V·s),\textsuperscript{33} $r_{\text{nr}}$ is the radius of the nanoreactor, $r_{\text{sc}}$ is the radius of the separation capillary, and $l_{\text{nr}}$ is the length of the nanoreactor. The term in the brackets represents the zone variance arising from the finite injection volume ($\sigma^2_{\text{inj}}$) and assumes that complete radial diffusion occurs when injecting into a column with a larger radius. As can be seen from this expression, to minimize $\sigma^2_{\text{inj}}$, it is necessary to use a highly cross-linked gel or high % T linear gel (small $\sigma_{\text{gel}}$) and also
a small-diameter nanoreactor and larger diameter separation column. It should also be noticed from the equation that the zone variance from $\sigma_{\text{inj}}^2$ will be highly dependent upon the number of bases comprising the oligonucleotide, since the longer oligonucleotides have a smaller $\sigma_{\text{gel}}$. In the present case, $\sigma_{\text{gel}} < \sigma_{fs}$ and $r_{nr} < r_{sc}$, and therefore, zone compression should result when the contents of the nanoreactor are injected directly into the gel column. It should also be noted that since the separation column used 7 M urea as the denaturant, it is expected that zone compression should also result at the head of the column due to a conductivity difference between the column and the solution contained within the nanoreactor.

The zone variance from the connector was determined by loading the nanoreactor with dye-labeled primer and interfacing it via the glass connector to a gel-filled column followed by electrokinetically injecting the dye primer onto the gel column. The electropherogram generated using this arrangement was compared to one generated after performing a direct injection of near-IR dye primer onto the CGE column under similar electrokinetic injection conditions. In both cases, the injection volume was kept constant, which was accomplished, in the case of the connector experiment, by removing the separation capillary and placing it in 1× TBE once the appropriate injection volume had been reached. The plate numbers (data not shown) were found to be $2.49 \times 10^5$ plates for direct injection and $2.45 \times 10^5$ plates when injection occurred across the zero dead volume connector. Since the only additional contribution to the zone variance was that arising from the connector between these two cases, the difference in the total zone variance calculated from the electropherograms yielded $\sigma_{\text{con}}^2$, which was determined to be $1.5 \times 10^{-4}$ cm$^2$. Even though the connector is stated as possessing zero dead volume,
the loss in efficiency in this case is most likely due to the inability to polish the capillary ends correctly, producing a void volume at the interface.

The nanoreactor used in our sequencing experiments was selected to have a length of 20 cm \( (l_{nr}) \) and radius of 10 \( \mu \)m \( (r_{nr}) \), yielding a total volume of 62 nL. From previous measurements, we found that the mobility of a 150-mer in a 3% T/3% C polyacrylamide gel with a near-IR primer was \( 6.1 \times 10^{-5} \) cm\(^2\)/V·s.\(^{10} \) Using these numbers as well as the free solution mobility of an oligonucleotide with >20 bases yielded a value of \( 3.4 \times 10^{-3} \) cm\(^2\) for \( \sigma_{nn}^2 \). The total variance \( (\sigma_{tot}^2) \) for this 150-mer near-IR labeled single-stranded oligonucleotide using direct injection onto a cross-linked gel column was found to be \( 1.3 \times 10^{-3} \) cm\(^2\) \( (H_{tot} = 2.5 \times 10^{-5} \) cm, where \( H \) is the height equivalent of a theoretical plate) and the plate number was \( 2.0 \times 10^6 \).\(^{10} \) Adding the zone variance contributions from the connector and the injection volume produced by the nanoreactor would give an estimate of the efficiency of this injection system, which yielded a value of \( 4.8 \times 10^{-3} \) cm\(^2\) \( (H_{tot} = 8.6 \times 10^{-5} \) cm) and a plate number of \( 5.8 \times 10^5 \). Therefore, one could expect an approximate 65% reduction in separation efficiency using the injection system as configured above compared to direct injection.

In Figure 3.8 is shown the capillary electropherogram for C-terminated fragments generated in the nanoreactor with subsequent separation in a 3% T/3% C cross-linked polyacrylamide gel column (i.d. = 75 \( \mu \)m). From this electropherogram, we estimated an effective read length of \(~450\) bases, which was determined by peak counting and then multiplying by 4 to a point in the electropherogram where resolution permitted single-base identification \( (R > 0.7) \). However, it should be noted that since only the C-track was analyzed, the stated read length will most likely be reduced when the system is
Figure 3.8. Capillary gel electropherograms of C-terminated Sanger sequencing fragments produced in the nanoreactor possessing a volume of 62 nL. The fluorescence of the labeling dye was excited with 4.5 mW of laser power at 780 nm. The column length, from injector to detector, was 40 cm with a total length of 70 cm. The electrophoresis was carried out using a field strength of 200 V/cm. The graph inset shows an expanded view with the 160-mer with the plate numbers calculated for this peak. In the figure, only 5000 s of data is shown.

transferred to a single-lane, four-base protocol. Calculation of the plate number for the near-IR dye-labeled 160-mer depicted in the figure inset yielded a value of $2.05 \times 10^6$ plates compared to $4.51 \times 10^6$ plates for direct injection (data not shown), in fair agreement with the efficiency loss calculated above. An interesting feature of this injection format is the fact that biases due to electrokinetic injection are absent, since in the present case we are injecting onto the gel column a fixed volume, which is defined by the physical dimensions of the capillary nanoreactor. Therefore, any differences in peak
amplitude between the early- and latter-migrating components arises from the efficiency of the polymerase enzyme during chain extension.

### 3.4. Conclusions

The use of near-IR fluorescence in DNA sequencing applications using CGE can result in significant advancements in detection and instrumentation. The background resulting from scattering and fluorescence interferences are much smaller in the near-IR region, which results in impressive sensitivity without the need for off column detection which can add to the complexity of the instrument. Since many near-IR dyes have absorption maxima which match fundamental lasing lines of inexpensive semiconductor diode lasers, sensitive fluorescence detectors can be constructed in multiplexing devices appropriate for handling large-scale sequencing projects.

We have demonstrated the ability to effectively produce sequencing ladders on a nanoliter scale, a volume that exploits the small sample requirements associated with the microcolumn separation techniques. The volume used in the present system (62 nL) represents an approximate 300-fold reduction in sample size typically used in Sanger chain-termination protocols. The net result is a significant reduction in the amount of consumables required for generation of sequencing ladders.

Another advantage of the present system is its ability to operate in an automated fashion. Recently several research groups have discussed automated systems for the preparation of sequencing ladders using HPLC hardware and the associated pumps. In these examples, the volumes required for sequence analysis were on the same order as those typically used in slab gel applications and, also, the ancillary equipment can be somewhat prohibitive. In our system, ultrasmall volumes are utilized and, in addition, no
mechanical pumps are required, with fluid pumping accomplished by simple
electropumping. Therefore, by using electropumping to fill the nanoreactor with the
appropriate amount of sequencing reagents (primer, sequenase, ddNTP, dNTPs, buffers,
and salts) required for DNA polymerization, the reactor inlet can be inserted into a
microtiter well containing these reagents and an electric field applied to fill the reactor. In
this way, only a few nanoliters of materials are consumed, and since the template is not
added to the mix, the remaining contents of the microtiter well can be used for
subsequent sequencing rounds without having to discard them after one use. Another
advantage associated with this system is the fact that biases due to electrokinetic injection
are absent. As such, alterations in the dNTP/ddNTP ratios during DNA polymerization
are not required and standard mixes included in the sequencing kit can be utilized. In
addition, since we are immobilizing the target DNAs onto the wall of the capillary, long
DNAs (>10 kbp) could be immobilized allowing one to implement primer walking
strategies in this format, simplifying sequence reconstruction.

3.5. References


Chapter 4

DNA Sequencing with Near-IR Heavy-Atom-Modified Fluorescent Dye-Labeled Terminators and Time-Resolved Fluorescence Lifetime Determinations in Capillary Gel Electrophoresis

4.1. Introduction

Fluorescence sequencing can be divided into two categories: dye primer sequencing, in which the fluorescent dyes are attached to the 5’ end of the sequencing primer, and the dye terminator sequencing, in which the fluorescent dyes are attached to the 5 position of pyrimidines or the 7 position of purines or pyrimidines in dideoxynucleoside triphosphates. Dye primer and dye terminator systems have their own set of advantages and disadvantages. One advantage of the dye primer system is that any DNA polymerase can be used as long as it accepts dideoxynucleoside triphosphates as substrates. The main disadvantage is the requirement for four separate extension reactions and each reaction requires a different dye-labeled primer. Consequently, primer chemistry methods do not easily lend themselves to the high throughput demands of DNA sequencing.

Dye terminators offer the advantages of convenience. Syntheses of labeled primers are unnecessary, thereby allowing the use of any sequencing primer. Additionally, because only one extension reaction is needed for each sequence, the labor required to perform the extension reactions is reduced. Another advantage of this chemistry is that noise from ‘false terminations’, in which the oligonucleotide is terminated by a deoxynucleotide rather than a dideoxynucleotide, is eliminated, since such fragments are unlabeled and therefore, invisible. A disadvantage is that many DNA polymerase enzymes are very sensitive to the type of dye attached to the terminators.1
The result is that the peak height in terminator chemistry profiles are far less uniform compared to the profiles for primer chemistry methods.\cite{1}

The commonly used approach for single lane base-calling in DNA sequencing applications using fluorescence is spectral discrimination, where a set of four spectrally distinct chromophores, which can be attached to either the sequencing primer or the dideoxynucleotide, are identified on the basis of unique emission maxima.\cite{2} The dyes typically used are the fluorescein and/or rhodamine derivatives, which contain structural modifications to alter the absorption and emission maxima to allow efficient discrimination.\cite{2,4} An ideal property of the chromophore set is similar absorption maxima, to allow excitation with a single source, but widely spaced emission maxima, to permit efficient sorting onto the appropriate detection channel. A set of commercial dyes are available (6-FAM, NED, TAMRA, and ROX), which nearly possess the aforementioned characteristics.\cite{5} These dyes can be excited with the 488- or 514-nm lines from an argon ion laser and then spectrally isolated using a series of optical filters onto appropriate photodetectors.\cite{3,6} Some of the potential difficulties with spectral discrimination include the need for multiple excitation sources in some cases and multiple detection channels.

A method to eliminate the need for multiple detection channels has been to use rotating filter wheels.\cite{3} Here, only one detection channel is required, but, due to decreases in the duty cycle, losses in the signal-to-noise ratio can result, producing errors in base-calling. In addition, this method still may require the need for multiple excitation sources. To eliminate the need for multiple excitation sources, the use of fluorescence-energy-transfer probes have been utilized in single-lane DNA-sequencing applications.\cite{7,9} In this approach, a donor (FAM) is attached to the 5'-end of a sequencing primer, and an
acceptor (FAM, JOE, TAMRA, ROX) is covalently bound to a modified thymidine residue eight or nine bases down the sequence. Since the primers utilize Förster resonance energy transfer, only a single excitation source is required (488 nm), with the emission sorted onto appropriate detectors.

Recently, a set of electrophoretically uniform fluorescent dyes for DNA sequencing have been reported. The dyes, which are BODIPY derivatives, are attached to sequencing primers (5'-end) via a unique linker structure that produces excellent sequencing data without software correction for dye-dependent mobility shifts. In addition, the dye-primer set yields narrower spectral emission bandwidths compared to those of conventional dye-primer sets, resulting in smaller amounts of cross-talk between detection channels.

While most sequencings applications using fluorescence require spectral discrimination to identify the terminal base during electrophoretic sizing, an alternative approach is to use the fluorescence lifetime of the labeling dye to identify the terminal base. The principal advantages associated with lifetime discrimination for base-calling are the following: (1) since the calculated lifetime is immune to concentration differences, dye-labeled terminators can potentially be used as well as dye-primers, with a wide choice in polymerase enzymes to suit the particular sequencing application; (2) lifetime values can be determined with higher precision than fluorescence intensities under appropriate conditions, improving the accuracy in base-calling; (3) lifetime determinations do not suffer from broad emission profiles associated with spectral discrimination; and (4) fluorescence can potentially be processed on a single detection channel without the need for spectral sorting to multiple detection channels. Several
problems do arise when considering lifetime discrimination for DNA sequencing, especially when utilizing microseparation techniques such as capillary gel electrophoresis. The most pervasive problem is associated with the complex instrumentation required for lifetime determinations. For example, in time domain techniques, a pulsed laser is required with a fast detector, typically a microchannel plate photomultiplier tube, and sophisticated counting electronics. In addition, poor photon statistics (low number of photocounts in decay profile) produced from low loading masses and the transient nature of the signal, determined by the width of the electrophoretic band (1-10 s), can produce poor precision in the measurement. The poor precision would also be compounded by the presence of large amounts of scattering and impurity photons included in the decay profile. Finally, complex algorithms are required for abstracting the lifetime from the decay profile, making on-line determinations during electrophoresis difficult.

There are two different formats for measuring fluorescence lifetimes, time-resolved\textsuperscript{11-15} and frequency-resolved.\textsuperscript{16-19} Since the time-resolved mode is a digital (photon counting) method, it typically shows better signal-to-noise than a frequency-resolved measurement, making it more attractive for separation platforms that deal with minute amounts of sample. In addition, the use of time-resolved methods allow for the use of time-gated detection in which background photons, which are coincident with the laser pulse (scattered photons) can be gated out electronically, improving the signal-to-noise ratio in the measurement.

In the time-resolved mode, a laser that operates in a pulsed mode is used to excite the fluorescence of the dye molecules traveling through the excitation beam. Typically,
the laser is on for a short time and then off for an extended period of time. During this off state, the excited dye molecule can relax, emitting a fluorescent photon. In time-resolved fluorescence, the time difference between excitation and fluorescence emission is measured. These time-differences are histogrammed and when photons are accumulated over many excitation/emission cycles, an exponential function is produced from which the lifetime can be extracted.

Many of these concerns associated with lifetime determinations for base-calling in DNA sequencing have been addressed using near-IR fluorescence. For example, several groups have demonstrated that semiconductor diode lasers, which can be operated in a pulsed mode and lase between 680 and 800 nm in conjunction with single-photon avalanche diodes (SPADs), can produce a time-correlated single-photon-counting apparatus that is simple to operate, with performance characteristics comparable to those of visible devices, using mode-locked Nd:YAG lasers and microchannel plates. Soper et al. reported that lifetime measurements can be acquired in the near-IR using a solid-state device and counting electronics situated on a PC board. They have also shown that simple algorithms can be used to calculate fluorescence lifetimes on-line during free solution or gel capillary electrophoresis using near-IR time-resolved fluorescence. The standard deviation in the lifetime measurement of C-terminated fragments labeled with a near-IR dye was found to be approximately ±9 ps with decay profiles constructed from as few as 5,000 photocounts. The high precision resulted primarily from the fact that, in the near-IR, the low scattering cross sections and the minimal number of intrinsically fluorescent components produced low numbers of interfering photocounts in the decay.
Wolfrum and co-workers have demonstrated the use of a four-lifetime approach to calling bases in DNA sequencing applications.\cite{17} In their work, a series of rhodamine derivatives were prepared, which possessed absorption maxima at ~630 nm and fluorescence lifetimes that varied between 1.6 and 3.7 ns. Using appropriate linker structures, dye-dependent mobility shifts were minimized, eliminating the need for post-electrophoresis processing. In a single lane, four lifetime format with capillary electrophoresis as the separation platform, these researchers were able to demonstrate a read length of 660 bases with a probability of correct classification >90%.

Flanagan et al. prepared a series of near-IR fluorescent dyes (tricarbocyanines) which possess the capability to be used in time-resolved identification of terminal bases in Sanger sequencing strategies using either slab gel or capillary gel electrophoresis.\cite{24} These dyes are structurally very similar, thereby showing identical absorption (765 nm) and emission (794 nm) maxima and also, demonstrate uniform mobilities in gel electrophoresis applications, irrespective of the dye-primer linker structure. The fluorescence lifetime of the dye series was altered by incorporating a heavy atom modification into the molecular framework of the base chromophore. By strategically positioning the heavy atom within the dye's structure or changing the identity of the heavy atom modification, the fluorescence lifetime could be systematically changed. Incorporating a single halide (I, Br, Cl, or F) into the molecular structure, the lifetimes of the dye set varied from 889 ps to 688 ps when measured in a nonpolymerized acrylamide solution containing 40% formamide.\cite{24} While these differences were found to be somewhat small ($\Delta \tau_f = 70$ ps, ~8% relative difference), our data has shown that the calculated lifetime obtained during capillary gel electrophoresis of near-IR labeled
oligonucleotides can be determined with high precision using simple maximum likelihood estimators. This high precision in part is due to the use of near-IR excitation, which minimizes background contributions into the estimated lifetime.

In this work, we will report on the study of DNA sequencing with the near-IR heavy-atom modified dye-labeled terminators in conjunction with slab gel electrophoresis for highly efficient base calling. Sequencing of a M13mp18 template will be demonstrated. Sample preparation methods for this dye terminator set have been optimized to yield intense electrophoretic bands and will be reported in this manuscript as well. In addition, we will demonstrate the lifetime identification of DNA sequencing fragments using this dye terminator set in capillary gel electrophoresis.

4.2. Experimental Section

4.2.1. Instrumentation

The instrumentation used to acquire the fluorescence lifetimes on-line during gel fractionation consisted of a conventional time-correlated single photon counting apparatus. The instrument (see Figure 4.1) consisted of a pulsed diode laser (PicoQuant, Berlin, Germany). The laser light was directed into a fiber, which transported the laser light to a dichoric and then to the gel column. The use of the fiber optic minimized beam walk, which required constant realignment of the system. The light emanating from the fiber was collimated by a 10X microscope objective and focused onto the capillary tube using a single laser focusing lens (Melles Griot, Irvne,Ca). The fluorescence emission was collected using a 60X (NA = 0.85) microscope objective (Nikon, Tokyo, Japan). The collected radiation was directed into a single model fiber and then imaged onto a slit, spectrally filtered (interference band pass filter, center wavelength = 810 nm; half band
width =10 nm (Omega Optical, Brattleboro, VT) and reimaged onto the face of a photodetector. The detector consisted of a passively quenched single photon avalanche diode (SPAD), SPCM-PQ-200 (EG&G optoelectronics, Vandrieul, Canada). The output of the SPAD was amplified 20-fold with an in-house fabricated fast amplifier and conditioned by a constant fraction discriminator (CFD).

**Figure 4.1.** Block diagram of near-Infrared time-correlated single-photon counting instrument. Components: L, focusing lens; C, capillary; DBDR, double dichroic; SF, spatial filter; BF, bandpass filter; MO, objective; SPAD, single photon avalanche diode; DISC, discriminator; ADC, analog-to-digital converter; TAC, time-to-amplitude converter; MCA, multi-channel analyzer; PC, personal computer.
The synchronization pulse obtained from the pulsed-laser was also directed into the CFD. The time-to-amplitude converter (TAC) was operated in a gated mode, with the gate pulse produced by a photocount generated from the SPAD. The start pulse was supplied by the SPAD output while the laser synchronization pulse was used as the stop pulse. The PC contained a PicoQuant SPC-430 time-correlated single-photon-counting module. The SPC-430 provides up to 4096 time channels for each of the 128 separately measured decay curves and the channel resolution can be as small as 813 fs.

4.2.2. DNA Sequencing with Near-IR Dye-Labeled Terminators

Near-IR dye terminators (see Figure 4.2) were kindly prepared for us by Shaheer Khan of Applied Biosystem. The dyes were sulfonated tricarbocyanines that were prepared in our laboratories.24

The dye-terminator sequencing fragments were prepared with an ABI Prism dye-terminator cycle sequencing kit (Applied Biosystems, Foster, CA) using a M13 (-40) forward primer and M13mp18 single-stranded template. The sequencing cocktail consisted of 4 µl sequencing buffer, 1.25 µl of 0.2 µg/µl template, 0.75 µl of 5 µM primer, 1.0 µl of dNTP mix, 1.0 µl of Taq polymerase, 11.25 µl of distilled water and varying amounts of near-IR-I labeled ddATP, near-IR-Br labeled near-IR-Cl labeled ddCTP, near-IR-F labeled ddUTP. Cycle sequencing was completed in a Flexigene thermal cycler (Techne, Minneapolis, MN) using the following conditions (34 cycles): (i) 95°C for 30 s; (ii) 55°C for 15 s; (iii) 60°C for 4 min.

The reaction cocktails were then subjected to a cold ethanol precipitation that consisted of the addition of 7 µL of 7 M NH₄OAc and 68 µL of 100% cold ethanol. The solution was vortexed and placed in the freezer (-20°C) for 30 min, followed by
Figure 4.2. Structures of heavy-atom modified tricabocyanine dye-labeled (A) I-ddATP (B) Br-ddCTP, (C) Cl-ddGTP (D) F-ddUTP. (fig. Con’d.)
centrifuging for 30 min at 12,000 rpm at room temperature. The supernatent was removed by tapping the reaction tube on a counter top and then, the samples were air dried. Finally, the DNA pellet was reconstituted in loading dye, vortexed for 1 min and placed in the freezer until needed for injection.

4.2.3. Slab Gel Electrophoresis and Data Collection

The sequencing was performed on a Li-COR model 4000 automated DNA sequencer (Li-COR, Lincoln, NE). The gel plates were boro-float glass that measured 21 cm x 47 cm. The sieving matrix was an 8%T (w/v) cross-linked gel (FMC Bioproducts Long Ranger, Rockland, ME) that contained 7.0 M urea as the denaturant and 1x TBE (pH 8.0). Polymerization of 30 ml of the 8% gel was initiated by adding 200 µl of 10%(w/v) ammonium persulfate and 20 µl TEMED, and a comb was inserted between the glass plates to form wells. After 3 hrs, the comb was removed and the plates placed in the sequencing instrument and the buffer reservoirs filled with 1x TBE. The gel was heated to 50°C and prerun at 1650 V for 30 minutes. One µl of each sample was loaded into the sample wells and the electrophoresis was run at 1650 V for 8 hours. Data acquisition and base-calling were performed using the Li-COR software loaded onto a Gateway computer.

4.2.4. Capillary Gel Electrophoresis

Capillary gel electrophoresis was performed in a capillary (Polymicro, Phoenix, AZ) with an internal diameter of 75 µm. The optical window was produced using a heating wire to remove the polyimide coating. The distance from the injector to detector was set at 50 cm, with the total length of 75 cm. The separation matrix was a self-coating POP6 matrix (Applied Biosystems, Foster, CA) with urea as the denaturant. The high
voltage was supplied by a Spellman high-voltage power supply (CZ1000R, Plainview, NY) and operated in a reverse mode (cathode at injection end). The running buffer used was a 3700 running buffer with EDTA (Applied Biosystems, Foster, CA). The column was run at 5 KV for 30 min with the running buffer prior to performing the CGE analysis. All samples were electrokinetically injected onto the column at 12 KV for 30 s. The separation was performed using a field strength of 200 V/cm.

4.2.5. Lifetime Data Acquisition and Analysis

The fluorescence decay profiles were collected over the electrophoretic peak. The data acquisition software was written in Visual Basic (Microsoft, Seattle, WA) and consisted of several control and data acquisition functions, such as streaming data to the hard drive (both intensity and time-resolved data) and providing for real-time visualization of the acquired data. After the electrophoresis run was complete, the data was compressed and assembled into one contiguous file using a freeware multiplatform compression library written in C by Mark Adler and Jean-loup Gailly. This freeware is a created set of library routines that provide implementation of the deflate compression algorithm and can be used as the compression engine for Zip compatible archivers.

The algorithms utilized to calculate the lifetime of the molecule from the histogram employed the maximum likelihood estimator (MLE). Data analysis was directly integrated into the controlling software. The data in the compressed binary file with the pixel decay information could be displayed as an entire 2D image of both the fluorescence intensity or lifetime along the y-axis as a function of time (x-axis). After construction of the decay profile, the lifetime was calculated using equation 4.1 (see below);
\[ 1 + \left( e^{T/\tau_f} - 1 \right)^{-1} - m(e^{mT/\tau_f} - 1)^{-1} = N_t^{-1} \sum_{i=1}^{m} iN_i \]

where \( m \) is the number of time bins, \( N_t \) is the number of photocounts in the decay spectrum, \( N_i \) is the number of photocounts in time bin \( i \), \( T \) is the time width of each bin, and \( t \) is the lifetime.

The right hand side of the equation is calculated from experimental values and the left-hand side describes a monotonic, non-decreasing function of two variables (\( m, t \)). For the chosen number of time bins, \( m \), a binary search for the calculation of \( t \) in the left-hand side of the equation was performed. The binary search for \( t \) begins by selecting a region of possible \( t \)'s from 10 to 40,000 ps. The left-hand side of the equation is calculated at the ends of the region as well as in the middle of the region. Depending on the relation to the right hand side value, the appropriate half of the region is chosen and the calculations are repeated. In each step, the region of possible values of \( t \) satisfying the equation is narrowed by a factor of two. Computation continues until the search region of lifetimes is less than a predefined value, typically ~0.1 ps.

The lifetime data were obtained by a pixel-by-pixel analysis of the electropherograms. Sections of the electropherograms can be enlarged and selected for lifetime analysis. Using the computer mouse, a single pixel or multiple pixels from the electropherograms can be selected to form decay profiles. The decay profiles can be analyzed by selecting the channels over which the MLE calculations were to be performed. The calculated value was shown in the bottom left corner in the display screen (see Figure 4.3).
Figure 4.3. Front panel displays of the data analysis packages for analyzing time-resolved data form sequencing gels. The small section of an electropherogram of a single-base tract is shown on the top panel. Using the computer mouse, a single pixel or multiple pixels from the electropherogram can be analyzed by selecting the channels over which the MLE calculation is to be performed. The calculated value is shown in the bottom left corner. In this case, the displayed decay represents counts from multi-pixels. The region in the electropherogram selected for constructing the decay is marked in red.
4.3. Results and Discussion

Since the amount of dideoxynucleotide and deoxynucleotide in chain-termination sequencing determine the signal intensity and read lengths, the molar ratio between ddNTP and dNTP were examined. The template was a single-stranded M13mp18 DNA and the primer was M13 (-40) forward (5’-GTTTTCCCAGTCACGAC). The enzyme was AmpliTaq FS. We analyzed varying concentration ratios of dye-ddNTP/dNTPS. After thermal cycling, as described above, the resulting DNA fragments were separated by size using Long Ranger gel electrophoresis. The concentration ratios of dye-ddNTP/dNTPS were optimized to give a balanced distribution of signal between base 10 and 700 (see Table 4.1).

Table 4.1. Optimal concentration ratios of dye-ddNTP/dNTPs in dye-terminator DNA sequencing.

<table>
<thead>
<tr>
<th>Dye-ddNTP</th>
<th>dNTPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-ddATP</td>
<td>19 µM</td>
</tr>
<tr>
<td>Br-ddCTP</td>
<td>25 µM</td>
</tr>
<tr>
<td>Cl-ddGTP</td>
<td>25 µM</td>
</tr>
<tr>
<td>F-ddUTP</td>
<td>6.25 µM</td>
</tr>
</tbody>
</table>

Under the optimal conditions, sequencing was carried out using four centrifuge tubes and four separate lanes of slab gel. The single base tracts prepared in each of four tubes were loaded onto the gel lanes of the sequencer. In Figure 4.4 is shown the slab gel electropherogram of the sequencing fragments. A trace of the sequencing data as well as the called bases is shown in Figure 4.5. Using the base caller for this machine, we were able to read 605 bases with the number of ambiguities from this trace determined to be 9.
Figure 4.4. Slab gel electrophoresis of sequencing fragments of a single stranded M13mp18 DNA using M13(-40) forward primer. Near-IR heavy-Atom-modified fluorescent dye-Labeled terminators (near-IR-I ddATP, near-IR-Br ddCTP, near-IR-Cl ddGTP, near-IR-F ddUTP) were used. Gel was 8% (W/V) polyacrylamide gel with 1.0 µl of sample for each well on a Li-COR 4000 automated sequencer at 1650 V for 8 hours.
To investigate the feasibility of lifetime identification in dye-terminator sequencing, NIR-I ddATP and NIR-Br ddCTP fragments were electrophoresed in a capillary with POP6 as separation matrix. A section of the I-ddATP/Br-ddCTP electropherogram was selected for base identification. The pixel-by-pixel analysis was used in lifetime determination. Lifetime identification of 32 base calls resulted in 3 miscalls. The resulting read accuracy for the two-dye capillary run was 90.6%. Figure 4.6 shown the two-dye electropherogram and the section of the electropherogram with the bases and miscalls identified. The miscalls resulted from the small difference of the lifetime values of the two dyes and overlapping electrophoretic peaks.

Figure 4.5. Sequence output from Li-COR 4000 automated DNA sequencer. The sequencing was performed using a single stranded M13mp18 DNA, M13 (-40) forward primer and near-IR heavy-Atom-modified fluorescent dye-Labeled terminators (near-IR-I ddATP, near-IR-Br ddCTP, near-IR-C1 ddGTP, near-IR-F ddUTP). Read length was 605 with the number of ambiguities equal to 9.
To acquire lifetime values in real dye-terminator DNA sequencing using CGE separation, single tracts of four dye-terminators were separately electrophoresed in a capillary with POP6 as the separation medium. In Figure 4.6 is shown the capillary gel electropherograms for four dye-terminator sequencing fragments. From these electropherograms we estimated an effective read length of ~620 bases which was determined by peak counting. The decay profiles for all four dyes together with the instrument response (prompt) functions in each of the detection channel were also shown. The instrument response functions were collected over the gel prior to the migration of ladders. The decay profiles and prompt functions were constructed by integrating the total number of counts over equal number of pixels obtained from the intensity image of the gel. The average lifetime value and standard deviation were taken from multiple peaks within a run and between electrophoretic runs. The lifetime values and standard deviations of this set of dyes are shown in Table 4.2.

As can be seen in the lifetime values, the four heavy-atom modified tricarbocynine dyes have weaknesses such as relatively short lifetimes and closely spaced lifetime values. The closely spaced lifetime values make this set of dyes unsuitable for single color/four lifetime sequencing. The largest difference was found to be 42 ps between near-IR-I and near-IR-Br dyes. Using the mean and the standard deviation, a Student’s T test indicates that we could successfully discriminate these two dyes at a 95% confidence interval (n_{obs} = 41; t_{calc} = 1.68; t_{95\%} = 2.02).
Figure 4.6. CGE electropherograms along with the associate decay profiles of DNA sequencing fragments with (A) near-IR-I ddATP, (B) near-IR-Br ddCTP, (C) near-IR-Cl ddGTP and (D) near-IR-F ddUTP. (fig. Con’d.)
Table 4.2. Average lifetime values of heavy-atom modified tricarbocyanine dye-terminator sequencing electrophoresed in a capillary with POP6 as the separation medium.

<table>
<thead>
<tr>
<th></th>
<th>( \tau_f ) (ps)</th>
<th>( \sigma ) (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR-I ddATP</td>
<td>650</td>
<td>9</td>
</tr>
<tr>
<td>NIR-Br ddCTP</td>
<td>691</td>
<td>8</td>
</tr>
<tr>
<td>NIR-Cl ddGTP</td>
<td>678</td>
<td>10</td>
</tr>
<tr>
<td>NIR-F ddUTP</td>
<td>662</td>
<td>15</td>
</tr>
</tbody>
</table>

To investigate the feasibility of lifetime identification in dye-terminator sequencing, near-IR-I ddATP and near-IR-Br ddCTP fragments were electrophoresed in a capillary with POP6 as the separation matrix. In Figure 4.7 is shown the capillary electropherogram for the two dye labeled fragments. A section of the I-ddATP/Br-ddCTP electropherogram was selected for base identification. For the two dye electropherogram, knowing the lifetime values and the standard deviations of each dye, pixel-by-pixel analysis was employed to identify each of the bases in the run. In the case of very poor resolution, extra attention had to be paid to carefully select the pixels across the band. Judicious choice of pixels along the electrophoretic band can improve the accuracy in the determination. If partially overlapped peaks consisted of two dyes, selected pixels where bad overlapping happened, for example, the center pixel of each band, the calculated lifetime would be a weighted average of the two components. However, the pixels from the rising or falling sides, where minimum overlap happened, would give a more reliable lifetime value associated with the peak. In the case of band doublets composed of only one dye, the lifetime values across the bands would be found to remain the same. Lifetime identification of 32 base calls resulted in 3 miscalls. The resulting read accuracy
for the two-dye capillary run was 90.6%. Figure 4.7 shown the two-dye electropherogram and the section of the electropherogram.

**Figure 4.7.** Electropherogram of two dye-terminator sequencing fragments in a capillary. Run in POP6 at 150 V/cm. Pixel-by-pixel analysis was employed for base calling. Four miscalls are noted by W.
4.3. Conclusions

We have demonstrated the feasibility of DNA sequencing with dye-labeled terminators in conjunction with slab gel electrophoresis. A set of terminators were labeled with the near-IR heavy-atom modified dyes. Sample preparation methods for this dye terminator set have been optimized to yield intense electrophoretic bands. Best results were generated by fine-tuning the ddNTP/dNTP ratios. We were able to read 605 bases with the number of ambiguities from the trace determined to be 9. Since prematurely terminated chains were not labeled, stop artifacts and most background bands were eliminated. Only one excitation source and one detection channel was required to process fluorescence data since the set of dyes possess similar absorption and emission maxima. In addition, the uniform mobility eliminated the need for postrun corrections. The fluorescence lifetimes of the dye sequencing fragments labeled with these dyes were determined in capillary gel electrophoresis. A one color, two lifetime format of DNA sequencing was implemented. A pixel-by-pixel analysis was employed to identify each of the bases in the run. The resulting read accuracy for the two-dye capillary run was 90.6%.

4.4. References


Chapter 5

Solid-Phase Reversible Immobilization in Microfluidic Chips for Purification of Dye Labeled DNA Sequencing Fragments

5.1. Introduction

The construction of the first rough draft of the human genome sequence by both the private and government/academic sectors has been reported. While the human genome has been successfully sequenced, there continues to be a need toward reducing costs, increasing automation and simplifying technology so that sequencing in areas such as functional genomics, sequencing genomes of model organisms, comparative and/or population genomics, de novo sequencing and the detection of single nucleotide polymorphisms (SNPs) can become more practical. In addition, improvements in sequencing technology can potentially move large genomic projects into smaller laboratories rather than restricted to major genome sequencing factories. In the last decade, extensive research has been dedicated to developing miniaturized separation techniques to reduce the development time and improve automation of the electrophoretic process, an example being capillary gel electrophoresis (CGE) and recently, capillary array electrophoresis (CAE).

Electrophoresis on microchips is an emerging new technology that promises to lead the next revolution in DNA sequencing. Microelectrophoresis platforms have already been shown to separate DNA restriction fragments, PCR products, short oligonucleotides and short tandem repeats rapidly and with reasonable resolution. High-speed separations of DNA sequencing fragments have also been achieved using these chips. Recently, microfabricated capillary array electrophoresis (µCAE) chips
have also been fabricated to further demonstrate increased throughput using these devices.\textsuperscript{25-27}

While microchips may become the platform of choice in most sequencing applications using electrophoresis as the final step in the process, it is clear that challenges in data quality will continue to be an important issue when using these devices. For example, the short length of the separation channel limits the effective read length obtainable using these devices.\textsuperscript{24} Proper preparation of the sequencing ladders prior to electrophoresis will be an integral step in using these devices, since the quality of the read depends intimately on the quality of the sample input into the device, especially for sequencing applications. In 1998, Ruiz-Martinez et al. and Salas-Solano et al. carried out critical investigations on the effects of reaction matrix components on read length using capillary gel electrophoresis for the separation.\textsuperscript{28-30} Desalting and template removal were found to be essential for maintaining read lengths $>500$ bases. Purification of sequencing reaction products in these studies was predominantly accomplished by conventional methods, such as acetate-ethanol precipitation,\textsuperscript{31} phenol-chloroform extraction,\textsuperscript{32} or gel filtration in spin columns or microtiter plate formats.\textsuperscript{33-35} Unfortunately, these methods require large sample volumes and high-speed centrifugation, which may not be practical for integration into miniaturized sequencing platforms.

There have been examples that discuss fully automated systems using capillaries and microfluidics to process sequencing samples and send the processed samples to a capillary gel column for electrophoretic sorting.\textsuperscript{36-39} Electroosmotic driven flow, dye-terminator chemistry and cycle sequencing was used to sequence $p$GEM templates in a
capillary nanoreactor with the cycle sequencing products purified (remove excess salts and dye-ddNTPs) using free solution capillary zone electrophoresis prior to loading on the gel column.\textsuperscript{37} Paegel and co-workers recently implemented a gel immobilization strategy in a microfluidic chip, in which appropriately prepared PCR products and the associated microfluidic circuit were used for the purification of the DNA sequencing samples.\textsuperscript{40} The PCR primers contained complementary sequences to short capture probes tethered covalently to a gel matrix. Due to hybridization between the PCR-generated sequencing products and the tethered capture probes, the authors were able to purify the templates in less than 120 s. Template release to allow loading into the electrophoretic channel of the microchip was accomplished by simply heating the gel matrix to denature the duplexed DNA.

An attractive approach for purifying cycle sequencing reactions prior to electrophoretic sorting is solid-phase reversible immobilization (SPRI).\textsuperscript{41-45} Under conditions of high polyethylene glycol (PEG) and salt concentrations, DNAs of the appropriate size binds to the surface of carboxy-coated magnetic particles. Once bound, the DNA-bead complexes can be washed and the DNA released from the beads in water to yield highly purified sequencing fragments. Elkin et al. optimized the binding buffer components for magnetic bead purification of DNA fragments appropriate for capillary gel electrophoresis.\textsuperscript{41, 42} The optimum combination of tetraethylene glycol (TEG) and ethanol was found to be appropriate for capillary gel electrophoresis with 57.4% ethanol and 5% TEG. Based on this technique, Agencourt has developed a highly effective magnetic bead based dye-terminator purification kit, CleanSEQ\textsuperscript{TM} (Argencourt Bioscience Corporation, www.argencourt.com). In this technique, magnetic beads are
coated with a carboxylate layer and then, the DNA is suspended in a solution containing TEG, which permits selective precipitation and immobilization of the DNA to the magnetic beads. Once the appropriate sized DNA has been immobilized, a magnetic field is applied and the beads are washed with ethanol to remove salts, excess dye-labeled terminators and other soluble components. DNA is released from the magnetic support using ddH$_2$O. In contrast to most solid phase purification methods, SPRI does not require a streptavidin coated solid support, the use of biotinylated primers or probes (oligonucleotides) attached to a support media. The attractive features associated with the SPRI technique are that it is conducive to automation, can be scaled to small volumes to reduce the amount of reagents required for the sequencing reactions and also, can potentially be adapted to microfluidic devices.

The photochemistry of bisphenol-A polycarbonate (PC) has been widely studied in the mid-UV (254 – 300 nm)$^{46-50}$ and far-UV (180 -240 nm)$^{51-54}$ At wavelengths between 254 – 300 nm, bisphenol-A polycarbonate is known to undergo a photo-Fries reaction with a relatively high quantum yield (values between 0.03 – 0.4)$^{48,49}$ The photo-Fries reaction results in the formation of phenyl salicylates and hydroxybenzophenones (see Figure 5.1). The occurrence of an O$_2$–contact charge transfer complex as an initiation step in the photo-oxidation of both aliphatic and aromatic hydrocarbons has been demonstrated by Chien.$^{55}$ One potential reaction of the excited O$_2$–complex is formation of hydroperoxides via a caged mechanism and this complex may go on to form carboxylic acids, as illustrated in Figure 5.1. Thus, direct UV-exposure of PC creates a carboxylated surface that could serve as a capture medium for DNA fragments in a SPRI purification format.
Figure 5.1. (A) Possible mechanism for the photochemical modification of PC in the mid-UV, producing the Photo-Fries rearrangement. (B) One reaction of the excited O$_2^-$ complex is to form hydroperoxide via a caged mechanism.

In this work, we demonstrate the purification of cycle sequencing reactions using SPRI technology carried out in a microchip fabricated from PC. A microchannel containing microposts was used as the high-surface area immobilization bed. By exposing the PC surface to UV radiation, a photo-oxidation reaction took place that resulted in the formation of surface carboxylate groups. Results on using these PC microchannels for SPRI to purify dye-terminator sequencing fragments will be presented. DNA was precipitated onto the surface of the photoactivated PC microchannels, washed, and released in ddH$_2$O. The purified sequencing fragments were then loaded into a slab gel of an automated DNA sequencer. In addition, we show that the PC microchannel can
be subjected to multiple purification rounds, making it reusable. We also demonstrate that this SPRI format could be directly coupled to a capillary gel column for separation of the purified DNA sequencing ladders.

5.2. Experimental Section

5.2.1 Generation of Dye-Labeled DNA Sequencing Fragments

Near-IR dye terminators were kindly prepared for us by Shaheer Khan of Applied Biosystem (Foster City, CA) using a sulfonated tricarbocyanine dye that was prepared in our laboratories according to published procedures. The dye-terminator sequencing fragments were prepared with an ABI Prism dye-terminator cycle sequencing kit (Applied Biosystems, Foster, CA) using a M13 (-40) forward primer and M13mp18 single-stranded template. The sequencing cocktail consisted of 4 µl sequencing buffer, 0.25 µg template, 3.75 pmol primer, 1.0 µl of dNTP mix, 1.0 µl of Taq polymerase, 11.25 µl of distilled water and 12.5 pmol of the near-IR labeled ddNTP. Cycle sequencing was completed in a Flexigene thermal cycler (Techne, Minneapolis, MN) using the following conditions (34 cycles): (i) 95 °C for 30 s; (ii) 55 °C for 15 s; (iii) 70 °C for 60 s.

5.2.2. Solid-Phase Reversible Immobilization of DNA Sequencing Fragments onto Sheet PC

A PC sheet (25 x 25 mm, 1.0 mm thickness) (Goodfellow, Berwyn, PA) was exposed to UV radiation (254 nm at 0.5 mW/cm²) for 30 min through a photomask. The PC was then washed with isopropanol. The activated PC was incubated with the dye-terminator DNA sequencing fragment mix, which consisted of 10 µl of the sequencing cocktail, 10 µl of binding buffer (CleanSEQ, Agencourt, Beverley, MA), and 129 µl of 73% EtOH for 10 min. The PC was finally washed extensively with 73% EtOH. The
surface of the PC was subsequently imaged using a near-IR fluorescence scanner constructed in-house.\textsuperscript{57}

5.2.3. Immobilization of Radiolabeled DNA onto Sheet PC

An oligonucleotide (100 bp, random sequence) was radiolabeled at its 5’ terminus with \([\gamma-^{32}\text{P}]\) ATP (Promega, Madison, Wisconsin). The reaction cocktail consisted of 200 pmols of DNA, 5 \(\mu\text{L}\) of 10X Kinase buffer, 15 \(\mu\text{L}\) of \([\gamma-^{32}\text{P}]\) ATP (3,000 Ci/mmol), 2 \(\mu\text{L}\) of T4 Polynucleotide Kinase (10 U/\(\mu\text{L}\)), and 15 \(\mu\text{L}\) of ddH\textsubscript{2}O. The reaction mixture was incubated at 37 °C for 10 min. The action of the enzyme was halted by adding 2 \(\mu\text{L}\) of 0.5 M EDTA. Unincorporated radiolabeled dATP was removed using a MicroSpin\textsuperscript{TM} G-25 column (Amersham Pharmacia Biotech, Piscataway, NJ). The efficiency of \(^{32}\text{P}\) labeling was determined by an assay using a DNA binding cellulose filter. One \(\mu\text{L}\) of the \(^{32}\text{P}\) DNA was spotted onto the filter. The filter was then dried under a heat lamp and washed twice with 0.5 M sodium phosphate to remove unincorporated nucleotides. Submerging the filter paper into a scintillation counter (LS6000IC series counter, Beckman Instruments, Fullerton, CA) and measuring the activity of the paper along with the known activity of the \(^{32}\text{P}\)-label allowed calculation of the labeling efficiency, which was determined to be 47\%. A 0.5 cm\textsuperscript{2} PC sheet was activated by UV irradiation and incubated with \(^{32}\text{P}\) DNA as described above. The radioactivity was measured by submerging the sample into a vial containing scintillation cocktail and monitoring the radioactivity using the scintillation counter.

5.2.4. Microchip Fabrication

The topographical layout of the SPRI microchip is shown in Figure 5.2B. Microchips were fabricated using procedures previously reported.\textsuperscript{58} Briefly, the procedure
involved fabricating a metal molding die by LIGA, which consisted of raised microstructures electroplated from Ni on a stainless steel base plate. The device topology is depicted in Figure 5.2C. The immobilization bed was 500 µm in width, 50 µm in depth and 4.0 mm in length. This bed was filled with an ordered array of microposts (d = 10 µm; spacing = 10 µm). Replicates of the molding die were hot embossed into PC (Goodfellow, Berwyn, PA). The embossing system consisted of a PHI Precision Press model TS-21-H-C (4A)-5 (City of Industry, CA). A vacuum chamber was installed into this press to remove air (pressure <0.1 bar) so complete filling of the die could take place. The wafers inserted into the press were 133 mm in diameter and the maximum area that could be patterned was 100 mm. Before molding, all residual water present in the polymer had to be removed. Therefore, the PC wafers were baked in an oven at 80 °C for 8 hrs. The die was coated with a release agent, MoldWiz (Axel, Woodside, NY), to improve demolding. During embossing, the molding die was heated to 180 °C and pressed into the PC wafer with a force of 1000 lbs for 4 min. During this process, the die was heated to 185 °C. After 4 min, the press was opened and the polymer part removed and cooled. The PC wafer was maintained at 85 °C throughout the demolding process. Following embossing, the devices were rinsed with copious amounts of isopropyl alcohol to remove any residual release agent remaining. The final devices were exposed to UV radiation (only the immobilization bed) as described above and assembled by thermal annealing the molded piece to a PC cover plate (1.0 mm thickness) at 154 °C in a circulating air oven for 10 min. The microchannel possessed a total available surface area of 2.1 x 10^7 µm² and a net volume of 1.6 x 10^8 µm³ (160 nL).
For slab gel electrophoresis, which required a larger volume for manual manipulation and loading into one of the wells of the gel, a 30 µl (total volume) open channel was fabricated using mechanical milling of a sheet of PC. This channel contained the following dimensions; 1,000 µm width, 300 µm depth and 5 cm in length and was an open channel and thus, did not contain microposts. Following milling, the channel was exposed to UV radiation using the procedure described above and thermally sealed with a PC cover plate.

5.2.5. PC Microchannel Purification of DNA Fragments for Slab Gel Gel Electrophoresis

For slab gel electrophoresis, purification of dye-terminator sequencing fragments was accomplished within the 30 µl PC open channel. Four channels were used, one for each terminator (single color sequencing). Each channel was filled with the appropriate purification mix, which consisted of 10 µl of ddATP, ddCTP, ddGTP or ddTTP sequencing fragments, 10 µl of binding buffer (CleanSEQ, Agencourt) and 129 µl of 73% EtOH. The channel was washed twice with 73% EtOH, air-dried and subsequently filled with ddH₂O and allowed to incubate for approximately 5 min. The eluate was transferred into a centrifuge tube and concentrated to ~1.5 µl in an Eppendorf Concentrator (Brinkman Instruments Inc., Westbury, NY) and mixed with 1 µl of loading dye.

The slab gel sequencing was performed on a Li-COR model 4000 automated DNA sequencer (Lincoln, NE). The sieving matrix was an 8%T (w/v) cross-linked gel (FMC Bioproducts Long Ranger, Rockland, ME) that contained 7.0 M urea as the denaturant and 1x TBE (pH 8.0). Polymerization of 30 ml of the 8% gel was initiated by adding 200 µl of 10% (w/v) ammonium persulfate and 20 µl TEMED. A 1.2 µl volume
of each sample was loaded into the sample wells and the electrophoresis run at 1,650 V for 8 hrs. Data acquisition and base calling were performed using the LI-COR software.

In order to investigate the ability to reuse the PC purification device once photoactivated, a channel was subjected to 4 rounds of SPRI purifications with each round analyzed on the LI-COR DNA sequencer. For these experiments, G or C terminator tracts were analyzed. Following release of the captured sequencing fragments, the channels were washed with ~4 column volumes of ddH$_2$O, and then incubated with the appropriate sequencing cocktail. The capture microchannel was not re-activated between purifications.

5.2.6. PC Microchannel Purification of DNA Fragments for capillary Gel Electrophoresis

Figure 5.2A shows a schematic of the instrumental set-up with integration of the PC-SPRI microchannel to CGE. The inner diameter of the separation column was 75 µm with a total length of 75 cm, 50 cm to the detection window. The gel column was connected to the microchannel through a reservoir machined into the microchip. The separation matrix was POP 6 (Applied Biosystems, Foster, CA) with 7 M urea used as the denaturant. The DNA fragments migrating through the gel column were detected using a custom-built near-IR laser-induced fluorescence system, which has been described previously.$^{59}$

The microchannel was filled with the purification mix using a micro-syringe through the input reservoir. The purification mix consisted of 10 µl of ddGTP sequencing fragments, 10 µl of binding buffer (CleanSEQ, Agencourt), and 129 µl of 73% EtOH. After incubation (2 min), the microchannel was washed twice with 73% EtOH (~2 column volumes). The microchannel was then evacuated and filled with ddH$_2$O.
Figure 5.2. (A) Schematic of the integrated PC-SPRI-CGE system. The separation matrix used for CGE was POP 6. The inner diameter of this column was 75 µm with a total length of 75 cm, 50 cm to the detection window. The gel column was connected to the microchannel through an access hole machined into the microchip. (B) Schematic diagram of the microfluidic device topology. The microchannel used for SPRI was 500 µm in width, 50 µm in depth, and 4.0 mm in length. Reservoirs were formed from 500 µm id holes drilled through the chip. (C) Optical micrograph of PC-SPRI capture bed and its dimensions. Shown is the embossed piece in PC fabricated using the metal master.
Following release from the PC wall (1 min), the gel-filled separation column was attached directly to the microchannel through the outlet reservoir. The sequencing fragments were electrokinetically injected into the gel column by applying +15 KV to the anodic end of the gel column for 45 s with the cathodic end of the microchannel grounded (see Figure 1A). Following injection, the gel column was disconnected from the microchannel and inserted into a microfuge tube containing the electrophoresis run buffer. The field strength was set to 200 V/cm during electrophoresis and the running buffer was a 3700 buffer with EDTA (Applied Biosystems, Foster City, CA).

For comparison of results obtained using the microchip SPRI to off-chip SPRI with direct injection for CGE, the sample for off-chip SPRI was prepared in a similar fashion as that described for slab gel electrophoresis. However, following pre-concentration in the Brinkman centrovap to a volume of 1.5 µl, the sample was taken to a total volume of 5 µl using ddH₂O in a micro-centrifuge tube and then electrokinetically loaded into the capillary gel column.

5.3. Results and Discussion

In this work, the purification of dye-terminator sequencing products was carried out using SPRI. Instead of magnetic beads containing a carboxylate layer, PC microchannels containing photoactivated immobilization beds were used. To increase the available immobilization sites, the capture bed contained high aspect ratio microposts within the microchannel. In Figure 5.2C an optical microscopic image of the capture bed is shown. The microposts were 10 µm in diameter with a center-to-center spacing of 20 µm and a height of 50 µm producing an approximate five-fold increase in the surface-to-volume ratio compared to an open channel of the same dimensions. By exposing the PC
surface to UV radiation, a photo-oxidation reaction took place that resulted in the formation of surface carboxylate groups. Formation of surface carboxylates on PC were hypothesized to serve as an effective capture medium for DNA sequencing fragments to allow purification of the sequencing cocktail when using dye-terminator chemistry prior to electrophoresis. Due to the low affinity of dye-labeled nucleotides for these types of surfaces in the appropriate buffer system, the sequencing fragments could be washed to remove these reagents and subsequently eluted from the surface by flooding the microchannel with ddH$_2$O.

5.3.1. Photoactivation of PC and DNA Capture

In order to investigate the feasibility of using photo-activated PC as an immobilization medium, we took sheet PC exposed to UV radiation through a photomask and allowed fluorescently labeled DNA-sequencing fragments produced via cycle sequencing with a near-IR dye-labeled terminator to incubate with the activated PC. In this experiment, excess terminators were removed prior to incubation with the activated PC surface to make sure that any fluorescence observed resulted from only the sequencing fragments. As seen in Figure 5.3B, those areas exposed to the UV radiation demonstrated higher levels of fluorescence, indicating the presence of fluorescently-labeled DNA fragments preferentially to those areas. After the PC was washed using ddH$_2$O, its surface was reimaged (see Figure 5.3C). As can be seen, the resulting fluorescence was removed with the surface image was comparable to that obtained prior to DNA immobilization, indicating that the immobilization was reversible. It should be noted that following UV exposure of PC through the photomask, those areas subjected to the UV radiation showed slightly higher levels of near-IR fluorescence compared to those
areas not exposed to this radiation (see Figure 5.3A). Repeating this same experiment with dye-labeled terminators demonstrated negligible amounts of fluorescence on the surface (data not shown). Therefore using this photoactivated PC and the appropriate binding solvents, only the extension products displayed a high affinity for the surface and not dye-labeled ddNTPs. Interestingly, similar experiments with PMMA (poly(methylmethacrylate)) photoactivated in the same manner indicated no affinity of the dye-labeled DNAs to this photoactivated surface, even though it produces carboxylate groups as well.\textsuperscript{60}

One important consideration to this purification format was the amount of DNA that could be captured (\textit{ie}, surface concentration). We carried out an experiment focused on quantifying the amount of DNA captured by photoactivated PC using \textsuperscript{32}P-labeled DNA and the binding buffer mix. The amount of radioactivity generated from the PC sheet was measured by submerging the entire PC sheet into a vial containing a scintillation cocktail and monitoring the activity using a scintillation counter. Table 5.1 summarizes the results. Based upon these results, we determined that the surface concentration of DNA immobilized to the surface (not corrected for surface roughness) was 3.9 x 10\textsuperscript{-12} mol cm\textsuperscript{-2}. In an earlier report, we used a streptavidin-coated capillary to immobilize sequencing templates, prepared via PCR, containing a biotin handle and found a surface coverage of 1.0 x 10\textsuperscript{-12} mol cm\textsuperscript{-2}.\textsuperscript{61} In the present case, we see an approximate 3-fold increase in surface coverage most likely due to the elimination of streptavidin as the anchoring scaffold, allowing the surface concentration to be determined by the immobilized DNA and not the streptavidin molecule. Given the
Figure 5.3. (A) Photoactivated PC (through UV photomask) imaged with a near-IR microscope. (B) Fluorescence image of the slide in (A) following incubation with the dye-labeled M13 sequencing ladder. The fragments were produced using cycle sequencing and a near-IR dye-labeled terminator. Excess terminator was removed using off-chip SPRI cleanup prior to incubation of the labeled fragments with the photoactivated PC surface. (C) Fluorescence image of the slide in (B) following washing with distilled water. In both cases, the surfaces were imaged with a near-IR confocal microscope (780 nm excitation) operated in a scanning mode. The fluorescence was processed on a single photon avalanche diode transducer.
available surface area in the microchip immobilization bed (0.21 cm$^2$), the mass of DNA that could be immobilized was calculated to be 0.82 pmol.

**Table 5.1.** The amount of DNA immobilized to a photoactivated PC surface.

<table>
<thead>
<tr>
<th>PC surface area (not corrected for roughness)</th>
<th>0.5 cm$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific activity of $^{32}$P-labeled DNA</td>
<td>55,696 cpm/pmol</td>
</tr>
<tr>
<td>Scintillation intensity</td>
<td>128,968 cpm</td>
</tr>
<tr>
<td>Moles of immobilized DNA</td>
<td>3.9 pmol/cm$^2$</td>
</tr>
</tbody>
</table>

### 5.3.2. PC-SPRI Coupled to Slab Gel Electrophoresis

To demonstrate the ability of the PC-SPRI format for purifying sequencing fragments prior to electrophoresis, the large volume (30 µl) PC channel was used with the purified fragments loaded onto a slab gel automated DNA sequencer. In this case, single base tracts were purified in each of four channels and then loaded into separate lanes of the sequencer because only a single color dye terminator was used. In Figure 5.4A is shown the slab gel electropherogram of these purified sequencing fragments. As can be seen, a large and intense band in the electropherogram from unincorporated dye-terminator, which typically appears ~20 bases into the electrophoretic tract, is absent indicating its effective removal using the PC SPRI cleanup protocol. A trace of the sequencing data as well as the called bases is shown in Figure 5.4 B. Using the base caller for this machine, we were able to read 620 bases with the number of ambiguities from this trace determined to be 7 (calling accuracy = 98.9%). In a similar fashion, we generated sequencing data of these dye-terminator products using commercial magnetic beads for purification. The results obtained (data not shown) indicated a 629 base read with the number of ambiguities equal to 8. Therefore, the PC channel purification format could produce high quality sequencing data by effectively removing the excess dye-
Figure 5.4. (A) Slab gel electrophoresis of sequencing fragments generated from a single stranded M13mp18 DNA using a M13(-40) forward primer. A single color near-IR fluorescent dye-labeled terminator was used with each terminator run in a separate lane of the slab gel. The sequencing fragments were purified using PC-SPRI prior to gel loading. The sieving gel was 8% (W/V) polyacrylamide with 1.2 µl of sample loaded into each well. The electrophoresis was performed using a LI-COR 4000 automated sequencer at 1650 V for 8 hours. (B) Sequence trace and called bases from the LI-COR 4000 automated DNA sequencer.
terminators with results comparable to conventional magnetic bead cleanup formats. In comparison to traditional ethanol precipitation protocols for dye-terminator chemistry, the SPRI method using magnetic beads or the PC microchannel increased the average read length by ~20 bases.

Our next set of experiments was focused on investigating the reusability of the photoactivated PC microchannel. In Figure 5.5 is shown the gel image of G and C tracts, which were purified using the same PC channel. Between each purification round, the PC

![Figure 5.5. Gel image of G and C tracts subjected to four purification rounds on the same photoactivated PC channel. See Figure 3 for experimental details.](image)
channel was washed with copious amount of ddH$_2$O to remove any DNA fragments potentially remaining from a previous purification. As can be seen from this data, after four purification runs the signal from the extension products is clearly observable and that from the dye-terminator absent. To eliminate the possibility of sample carryover from a previous purification round, G (runs 1, 2) and C (runs 3, 4) fragments were purified using the same channel. The data showed that the sequencing fragments from a previous run were completely removed, because bands appearing in the G-tract did not appear in the subsequent C-tracts.

5.3.3. SPRI and Capillary Gel Electrophoresis

SPRI was initially used for isolation of PCR products in the presence of polyethylene glycol (PEG) and NaCl.$^{44,45}$ Unfortunately, high NaCl concentrations are known to inhibit electrokinetic injection of DNAs in CGE, producing a significant variation in peak height.$^{30}$ In addition, PEG contamination reduces the sequencing trace quality and requires several washes prior to gel loading.$^{62}$ Seventy percent ethanol is widely used to desalt and precipitate DNA. However, excess dye terminators from the sequencing reactions also bind to the carboxylated solid surface using this buffer.$^{41}$ Flock and co-workers described DNA precipitation through changes in the solution dielectric constant.$^{63}$ Tetraethylene glycol (TEG) has desirable properties as an additive, which includes its low dielectric constant, high polarity, low viscosity, neutral charge, and a solution density greater than water. Elkin and co-workers optimized the binding buffer for magnetic bead purification of DNA fragments specifically for CGE using TEG and ethanol.$^{41}$ The optimal combination of TEG and ethanol was found to be 57.4% ethanol with 5% TEG. In this work, we used a combination of 57.4% ethanol with 5% TEG for
selective binding of the sequencing fragments to the photoactivated PC immobilization bed prior to CGE.

Salas Solan et al. have extensively characterized electrokinetic injection into capillaries and found that injection from water gives better peak resolution and a higher loading level of DNAs compared to injection from typical formamide/EDTA solutions.\textsuperscript{30} In addition, the presence of other ionic species in the injection solution can significantly decrease the amount of DNA introduced into the capillary column. Since the ion concentration is low in ddH\textsubscript{2}O, the DNA is preferentially loaded during electrokinetic injection. High loading and the thin injection plugs resulting due to conductivity differences between the injection plug and the sieving matrix improve read length. Therefore, to improve data quality, we used only ddH\textsubscript{2}O to elute the DNA fragments from the PC microchannel.

With SPRI, the binding efficiency of DNA to the solid surface is a function of the DNA size, with the short DNA fragments yielding lower binding efficiency compared to the longer fragments.\textsuperscript{43} This is beneficial for DNA sequencing in capillary gel electrophoresis, since the amount of sample injected into the column depends on the field strength, time of injection and the mobility of the sample components in the column sieving matrix. The DNAs migrate through the gel matrix at a rate dependent upon the size of the fragments. It is therefore common to observe biasing effects, in which the shorter DNAs are loaded more effectively than longer DNAs. Therefore, one typically observes smaller fluorescence signals for larger DNAs. Because of the low binding efficiency of short DNA fragments in SPRI, the biasing effect induced by electrokinetic
injection is counteracted by SPRI, potentially producing a more uniform peak height in the sequencing tract.

In Figure 5.6 is shown the capillary electropherogram for near-IR dye G-terminated sequencing fragments purified in the PC microchannel with subsequent separation using CGE. This result indicated that excess dye-terminators were eliminated. The increase in surface to volume ratio in the microchannels with microposts resulted in the increase in available immobilization sites and therefore, the DNA fragments eluted from the microchannel could be directly loaded into the gel column without the need for a pre-concentration step as was done for the slab gel example. From this electropherogram, we estimated an effective read length of ~450 bases, which was determined by peak counting and multiplying by 4 to a point in the electropherogram where resolution permitted single-base identification (R > 0.7).

Calculation of the plate number for the peak at 1,623 s yielded a value of 8.2 × 10^4, compared to 4.51 × 10^6 plates calculated for this same peak using direct injection (V_{inj} = 5 KV; t_{inj} = 5 s, injection time and voltage adjusted to give comparable signal intensities). The total zone variance calculated for the microchip purification was 3.1 × 10^{-2} cm^2. Assuming plug injection, the variance due to the finite injection volume for the microchip purification was determined to be 2.6 × 10^{-2} cm^2, nearly 84% of the total variance. Therefore, the extended injection time used in the case of the microchip purification was a major contribution to the reduced efficiency. There are two design modifications that can be envisioned for future PC-SPRI microchips that would improve plate numbers: (1) Decrease micropost diameter and inter-post spacing. If the post spacing and diameters were reduced, the effective mass loading level of DNA would
Figure 5.6. Capillary gel electropherogram of a near-IR dye G-terminated sequencing ladder purified in the posted PC microchannel using SPRI. The fluorescence of the labeling dye was excited with 5 mW of laser power at 780 nm. The column length, from injection to detection, was 50 cm with a total column length of 75 cm. The separation matrix was POP 6 gel. The electrophoresis was carried out using a field strength of 200 V/cm. In this Figure, the electrophoresis was run for 2400 s following sample injection prior to commencing data collection. Only 5000 s of data is shown.
increase as well as its concentration once released from the surface. (2) Reduce the volume of the interconnect between the microchip and capillary gel column. In the present configuration, a small hole was mechanically drilled into the chip substrate to serve as the outlet reservoir and also as the interface to the capillary gel column. The total volume of this reservoir was estimated to be 12.5 µl, which reduced the pre-concentrating effects exhibited by the 160 nL immobilization bed. Fabrication of low dead volume interconnects machined into the microchip containing the immobilization bed could dramatically reduce dilution effects associated with the interconnection and therefore, reduce electrokinetic injection times to improve electrophoretic separation efficiency.64

5.4. Conclusions

We have demonstrated the use of a novel PC-SPRI format for the purification of dye terminator DNA sequencing fragments and integrated this process to both slab and capillary gel electrophoresis. The method produced high quality sequencing results when using dye-terminator chemistry by effectively removing excess dye-terminator and other soluble components in the sequencing cocktail. The protocol is inherently simple, requiring no special chemical treatment of the sequencing reagents, primers, or templates and, does not require the use of a magnetic field or magnetic beads. In addition, the PC microchannel can be scaled to smaller volumes than that used herein (160 nL), which could potentially produce a significant reduction in sequencing costs associated with the use of consumables. Another attractive feature of this PC microchannel format is the ability to perform multiple purifications from a single chamber without requiring reactivation of the surface. In comparison to traditional ethanol precipitation, the method
requires no centrifugation and is thus more conducive to automation in microfluidic systems.

While we were able to demonstrate the integration of this microchip purification platform to CGE, the observed plate numbers compared to direct injection with off-chip SPRI purification were found to be significantly reduced. This was shown to be due partly to the longer injection times required to obtain similar peak heights for both methods. To improve signal intensity using the microchip format without sacrificing separation efficiency, the concentration of the released fragments needs to be increased. This could be accomplished by increasing the micropost density within the microchannel (ie, smaller diameter posts with smaller inter-post spacings) or redesigning the interconnect.

The attractiveness of this cleanup protocol could be realized when integrated to a microchip electrophoresis device. Sample cleanup, injection and separation could then be performed on a single planar device in a fully automated fashion, since the cleanup protocol only requires addition of the appropriate buffers or solvents. The potential benefits of such a device results from the fact that the immobilization bed is fabricated during embossing/molding of the microfluidic network and activation of the surface appropriate for SPRI requires only exposure to UV light. In addition, the activated surface can be used for multiple rounds of purifications. Work in our laboratory is currently focused on developing such an integrated microfluidic device.

5.5 References


5. Dovichi, N. J. "Human Mutation" 1993, 2, 82-84.


60. Patel, A.; Vaidya, B.; McCarley, R.; Soper, S. A. *(manuscript in preparation)* **2003**.


Chapter 6

Conclusions and Future Work

6.1. Conclusions

Miniaturization, integration and automation are all well-recognized trends in the development of various protocols for high-speed and low-cost DNA sequencing. This work focused on the development of methodologies for sample preparation and cleanup on a nanoliter scale in capillary gel electrophoresis DNA sequencing.

In Chapter 1, the background literature and history of fluorescence-based DNA sequencing were outlined. The theory of electrophoresis was discussed as it developed from the traditional slab gel electrophoresis to capillary gel electrophoresis. The principles of fluorescence were outlined with focus on visible and near-IR fluorescence. Finally, several different detection methods and sequencing formats were reviewed.

In Chapter 2, solid-phase micro-reactors were prepared in glass capillaries for DNA sequencing applications using slab gel electrophoresis, which consisted of a fused silica capillary (i.d. = 100 µm; length = 15 cm) that contained a covalently bound biotin molecule. Upon the addition of streptavidin to the capillary, an anchoring site was produced for the tethering of biotinylated DNA sequencing templates to the wall of the capillary. Using a four-lane, single dye primer sequencing run, a read length was found to be 589 bases. The ability to use dye-terminator chemistry was also investigated by using a near-IR dye-labeled terminator.

In Chapter 3, a miniaturized, solid-phase nanoreactor (volume ~62 nL) was developed to prepare Sanger DNA-sequencing ladders, which was directly interfaced to a
capillary gel electrophoresis system. The capillary reactor was placed inside a thermocycler to control the temperature during chain extension and was directly connected to the gel column via zero dead volume fused-silica connectors. The complementary DNA fragments generated (C-track only) in the reactor were denatured using heat and directly injected onto the gel-filled capillary for size separation with detection accomplished using near-IR laser-induced fluorescence. Extension and single-base separation resolution of the C-track, which was directly injected onto the gel column, was estimated to be >450 bases from the primer annealing site with plate numbers ranging from $1 \times 10^6$ to $2 \times 10^6$.

In Chapter 4, DNA sequencing with dye-labeled terminators in conjunction with slab gel and capillary gel electrophoresis was investigated. A set of terminators were labeled with near-IR heavy-atom modified dyes. Sample preparation methods for this dye terminator set were optimized to yield intense electrophoretic bands. Using the slab gel machine, we were able to read 605 bases with the number of ambiguities from this trace determined to be 9. To acquire lifetime values in real dye-terminator DNA sequencing using CGE separation, single tracts of four dye-terminators were separately electrophoresed in a capillary with POP6 as the separation medium. The lifetime data were obtained by a pixel-by-pixel analysis of the electropherograms. A two dye-labeled terminator read was called with lifetime identification methods.

Solid-Phase Reversible Immobilization (SPRI) chemistry is based on DNA binding to a surface with carboxylate groups. By exposing the PC surface to UV radiation, a photo-oxidation reaction takes place that results in the formation of surface carboxylate groups. In Chapter 5, a SPRI-based PC microchannel containing microposts
was investigated for the purification of dye-terminator DNA sequencing fragments. Unpurified near-IR dye terminator sequencing fragments were precipitated onto the surface of the microchannel, washed and eluted in water. The method was reproducible, leads to long reads and high sequencing accuracy. A read length was obtained (620 with number of ambiguities of 7). High-speed, integrated sequencing was accomplished by directly connecting the microchannel to capillary gel electrophoresis.

6.2. Future Work

The nanoreactor and the SPRI-based PC microchip will be integrated to a multi-channel modular microfluidic system to carry out various processing steps associated with DNA sequencing. The integrated device will consist of modular units that will PCR amplify template DNA molecules, purify the PCR amplicons, perform cycle sequencing on the PCR product, purify the sequencing products using SPRI technology, electrokinetically load and preconcentrate the purified products in a microfabricated multichannel electrophoresis device, and perform high speed electrophoretic sorting of the DNA ladders with detection and base identification via time-resolved, near-IR fluorescence. In its final version, each chip will consist of 16 independent channels that will read sequencing data from PCR generated products. The throughput of the device will be augmented by running a two-color, four lifetime format (8 spectroscopically unique probes), that will allow reading sequencing data from two unique samples run simultaneously in a single gel run. In addition, the device will be based on microfluidics, in which nL volumes of sample and reagents will be used in a single sequencing run. The DNA sample will be shuttled through each device and between devices via its electrophoretic mobility. The schematic diagram of this process is shown in Figure 6.1.
Figure 6.1. Integrated modular microfluidic system for the automated processing of sequencing templates.

6.3. References


4. Soper S. A. *R01 application for development of a multi-channel modular microfluidic system to carry out various processing steps associated with DNA sequencing.*
Vita

Yichuan Xu was born in Wangjiang, Anhui, People’s Republic of China, on December 13, 1962. He has three sisters and one brother. He married Yonghong Shang in 1990. They have a daughter, Mengze Xu and a son, Kenneth Xu. He attended Taichi High School from 1976 – 1979.

In the fall of 1979, he entered Fudan University in Shanghai, China. He received a bachelor of science degree in chemistry in the Spring of 1983 and a master of science degree in chemistry in the Spring of 1986. In the Fall of 1986, he worked as an instructor in the Department of Chemical Engineering at Hefei University of Technology in Hefei, China. In the Spring of 1996, he entered the graduate program in the Department of Chemistry at Louisiana State University in Baton Rouge, Louisiana. In the Spring of 2000, he entered the master program in the Department of Computer Science at Louisiana State University in Baton Rouge, Louisiana. He is currently a candidate for the degree of Doctor of Philosophy in analytical chemistry which will be conferred at the May 2003 Commencement.