

UNIVERSITY OF ALBERTA

**EVALUATION OF MULTIPLE CAPILLARY SYSTEM FOR DNA
SEQUENCING**

BY

JUYING YAN



A thesis submitted to the faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Master of Science

DEPARTMENT OF CHEMISTRY

EDMONTON, ALBERTA

Fall, 1996



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Voire référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence évocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-18337-8

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: JUYING YAN

TITLE OF THESIS: EVALUATION OF MULTIPLE CAPILLARY SYSTEM
FOR DNA SEQUENCING

DEGREE: MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED: 1996

Permission is hereby granted to the UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Yan, Juying

(Student's signature)

Permanent Address:

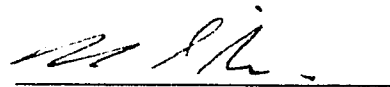
35731 Bettencourt St. Apt #234

Newark, CA 94560, USA

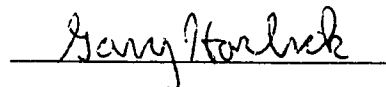
DATE: June 26, 1996

UNIVERSITY OF ALBERTA
FACULTY OF GRADUATE STUDIES AND RESEARCH

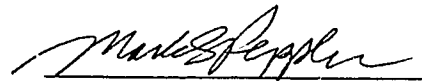
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled EVALUATION OF MULTIPLE CAPILLARY SYSTEM FOR DNA SEQUENCING submitted by JUYING YAN in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE



Dr. N. J. Dovichi, Supervisor



Dr. G. Horlick



Dr. M. Peppler

DATE: 14-Aug-96

Abstract

A five capillary electrophoresis system has been used for DNA sequencing study in this thesis. Over ten months intense work demonstrated that this system is robust, versatile and generates high throughput for DNA sequencing. Improvement in data acquisition and data processing are implemented to make the system more efficient and easy to use.

Optimum separation conditions (including electric field, capillary length, gel composition, sample injection amount and time, sample purity, etc.) have been studied. Different protocols of sample preparation were tested and large numbers samples were run through the machine. A sequence rate up to 1100 bases/hr can be achieved.

The electric field behavior of DNA sequencing fragments and reptation model are also discussed, and value of $\beta=1$ obtained within experimental error in contrast to the classic model in which $\beta=2$ is suggested. It is shown that the onset of biased reptation is inversely proportional to the electric field.

An automatic base-calling program was tested and used for the DNA sequencing data processing.

Acknowledgments

I would like to express my most sincere thanks to Dr. Norm. J. Dovichi for his continuing encouragement, invaluable advice, support and guidance in the research project.

Thanks to Dr. J. Z. Zhang, R. Jiang, Dr. X. R. Liu, Dr. X. B. Feng, Dr. H. Ren and Pieter Roos, Joan Hou, and members of our research group. for their generous assistance, helpful suggestions and great cooperation.

Table of Contents

Chapter 1	1
Introduction	1
1.1 Introduction to DNA and DNA Sequencing techniques.....	1
1.2 DNA sample cloning and preparation for DNA sequencing by gel electrophoresis	5
1.3 Capillary Gel Separation.....	11
1.4 Models of DNA separation in gel electrophoresis	13
1.4.1 Ogsten Sieving Model.....	14
1.4.2 Biased reptation model.....	14
1.5 Scope of this project.....	16
Chapter 2	17
Instrumentation and experiment	17
2.1 Multiple capillary electrophoresis system	19
2.2 Data acquisition subsystem.....	23
2.3 Detection Limit	27
2.3.1 Detection limit obtained for Ar ⁺ blue laser	28
2.3.2 Detection limit obtained for HeNe green laser	29
2.4 Five channel cross-talk test.....	30
2.5 Separation matrix	30
Chapter 3	35
Study of separation conditions	35
3.1 Effect of polymer matrix.....	36
3.1.1 Effect of urea concentration	36
3.1.2 Effect of Polymerization temperature.....	38

3.1.3 <i>Used polymer matrix</i>	38
3.1.4 <i>Effect of total acrylamide concentration</i>	47
3.2 Effect of Electric field.....	52
3.3 Effect of Capillary length.....	61
3.4 Sample injection related parameters	66
3.4.1 <i>Minimum sample volume requirement</i>	66
3.4.2 <i>Sample loading effect on the separation</i>	68
3.5 The reuse of the separation matrix.....	73
3.6 Effect of sample preparation conditions	82
3.6.1 <i>Effect of ddNTP:dNTP ratio and with or without Mn⁺²</i>	84
3.7. Summary	88
Chapter 4	89
Sequence data processing and base calling	89
4.1 Channel separation and data editing	90
4.2. Spike peak detection and removal	91
4.3 Bandpass digital filtering	92
4.4. Base calling	96
4.5 Summary	98
Chapter 5	100
Summary and Future work	100
5.1 Evaluation of the system	100
5.2 Future work	102
References	104

Table of Figures

Figure 1.1 Structure of DNA	2
Figure 1.2 Structure of four bases.....	3
Figure 1.3 Structure of ddNTP and dNTP	8
Figure 1.4 DNA sample preparation for gel electrophoresis separation with fluorescence detection.....	9
Figure 1.5 Sketches of the emission spectra of four labeling dyes (Fam, Joe, Tamra and Rox) with four bandpass filters (dotted lines) illustrating the relative transmission efficiency of a dye on the four filters	10
Figure 2.1 Schematic diagram of a typical capillary electrophoresis system with post-column detector.....	18
Figure 2.2 Overview of the five capillary system.....	21
Figure 2.3 The multiple capillary sheath flow cuvette (from reference 51)	22
Figure 2.5 Data-acquisition subsystem.....	26
Figure 2.6 Electropherograms of (a) fluorescein (1.56×10^{-12} M, excited by Ar ⁺ laser, 488 nm) and (b) Rhodamine 6G (9.4×10^{-13} M excited by He-Ne laser, 543.5 nm)	30
Figure 2.7 Electropherograms of the five capillaries with sample injected into Ch1 and Ch3 and no samples are injected into Ch0, Ch2, Ch4.....	32
Figure 3.1 Effect of different urea concentration on the separation of the reaction mixture of an A-terminated M13mp18 sample.....	39
Figure 3.1.1 Effect of urea concentration. Zoomed view at about 100 bases	40
Figure 3.1.2 Effect of urea concentration. Zoomed view at 100--200 bases	41
Figure 3.1.3 Effect of urea concentration. Zoomed view at >300 bases.....	42
Figure 3.2 Comparison of separation with gels polymerized at different temperature.....	43

Figure 3.2.1 Effect of polymerization temperature on separation efficiency N_p (a) and on resolution R (b).....	44
Figure 3.3 Comparison of separation for degassed and non-degassed polymer matrices.....	46
Figure 3.4 Effect of gel composition on the separation of four color labeled DNA fragments.....	49
Figure 3.4.1 Effect of gel composition. (a) Plot of migration time vs. fragment length; (b) Plot of mobility vs fragment length.....	50
Figure 3.4.1 Effect of gel composition. (c) Plot of resolution R vs. fragment length for different gel compositions, (d) Plot of $\ln(\text{mobility})$ vs %T for different fragment length.....	51
Figure 3.5 Separation of single base terminated sequencing fragments at electric fields ranging from 100 to 300 V/cm.....	55
Figure 3.5.1 Zoomed view of Fig. 3.5 at base pairs around 320.....	56
Figure 3.5.2 Effect of electric field. (a) Log mobility vs. fragment length. (b) Mobility vs. $1/N$	57
Figure 3.5.2 Effect of electric field. (c) log-log plot of N^* versus electric field E ; and (d) $1/N^*$ vs. Electric field E	58
Figure 3.5.3 Separation efficiency vs. fragment length for different electric field.....	60
Figure 3.6 Separation of single base terminated sequencing fragments with capillary lengths ranging from 35 to 61 cm.....	62
Figure 3.6.1 Zoomed views of Fig 3.6 at about 320 base pairs.....	63
Figure 3.6.2 Zoomed views of Fig 3.6 at about 560 base pairs.....	64
Figure 3.6.3 Effect of capillary length. (a) Plots of Separation efficiency vs. Fragment length, and (b) Plot of resolution vs. fragment length.....	65
Figure 3.7 Effect of sample concentration on separation with a given injection condition.....	70

Figure 3.7.1 Effect of injection time on (a) separation efficiency N_p and (b) resolution R for different fragment length	72
Figure 3.8 The normal running current and reverse polarity current	77
Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18 (a) First time run.....	78
Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18 (a) First time run.....	79
Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18 (b) Second time run after 30 min. Polarity reversal	80
Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18 (b) Second time run after 30 min. Polarity reversal	81
Figure 3.9 Comparison of two samples with or without Mn^{+2} . Zoomed view. Sample S1, has no Mn^{+2} ; Sample S2, with Mn^{+2}	86
Fig 3.10 Comparison of two samples with different dNTP:ddNTP ratios, 100/1 and 200/1.....	87
Figure 4.1 A single color electropherogram, its Fourier transform, and the frequency domain filtering function.....	93
Figure 4.2 A portion of a four color electropherogram (A and C) and the processed data (B and D). (A), (B) the beginning of the electropherogram; (C), (D) the end of the electropherogram.	95
Figure 4.3 Relative fluorescesene intensities of the four bases on the four bandpass filters. The bases are called according to their relative intensities after normalization and filtering (a). The raw data is plotted in (b).	97
Figure 4.4 A plot of base calling accuracy vs. fragment size	99

List of Tables

Table 3.1 Minimum sample volume for adequate injection (100 V/cm).....	67
Table 3.2 Separation conditions used for DNA sequencing by the multiple capillary CZE system.	74
Table 3.3 Sample preparation conditions for Mn ⁺² effect and ddNTP:dNTP ratio effect.	85
Table.5.1 Total number of sample run.....	101

Glossary of Abbreviations

CZE:	Capillary Zone Electrophoresis
CGE:	Capillary Gel Electrophoresis
FFT:	Fast Fourier Transform
DAC:	Digital - to - Analog Converter
ADC:	Analog - to -Digital Converter
F/V:	Frequency - to - Voltage converter
DNA:	Deoxyribonucleic Acid
Grin Lens:	Gradient Refractive Index Lens
PCR:	Polymerase Chain Reaction
dNTP:	Deoxynucleoside Triphosphates
ddNTP:	Dideoxynucleoside Triphosphates
BRF:	Biased Reptation with Fluctuations
Vis:	Virtual Instruments
TBE:	Tris-Boric-EDTA

Chapter 1

Introduction

1.1 Introduction to DNA and DNA Sequencing techniques

DNA is an abbreviation for deoxyribonucleic acid. The biological function of DNA is for information storage and transmission. In nature, most DNA molecules exist in the double stranded form as a ribbon-like, long chain. The backbone consist of alternating sugars and phosphate groups. One of four bases A (Adenine), G (Guanine), T (Thymine), C (Cytosine) is covalently bound to each sugar, which in turn is covalently bound to a phosphate group. The two strands are associated with each other through hydrogen bonds (See Figure 1.1). A is always paired with T, and C is always paired with G. The sequence of these four bases along the sugar-phosphate backbone encodes the genetic information. The structures of the four bases are shown in Figure 1.2.

A DNA sequence is the order of the 4 nucleotides A, G, T, C on a particular strand of DNA. DNA sequence is important in the areas of genetic, biochemical and medical research, in that the basic information for biological structure and function is encoded in sequences of the DNA molecules of cells [1]. DNA determines the structure of proteins (e.g., each block of three nucleotide residues corresponds to a separate amino acid) [2]. By revealing the similarities of homologous genes, it provides insights into the possible regulation and functions of those genes as well as into their evolutionary history. Also, sequence information leading to an understanding of disease states related to genetic variation should have an enormous impact not only on biological research but also on the practice of medicine [3]. Many of the remarkable advances in biology and medicine in

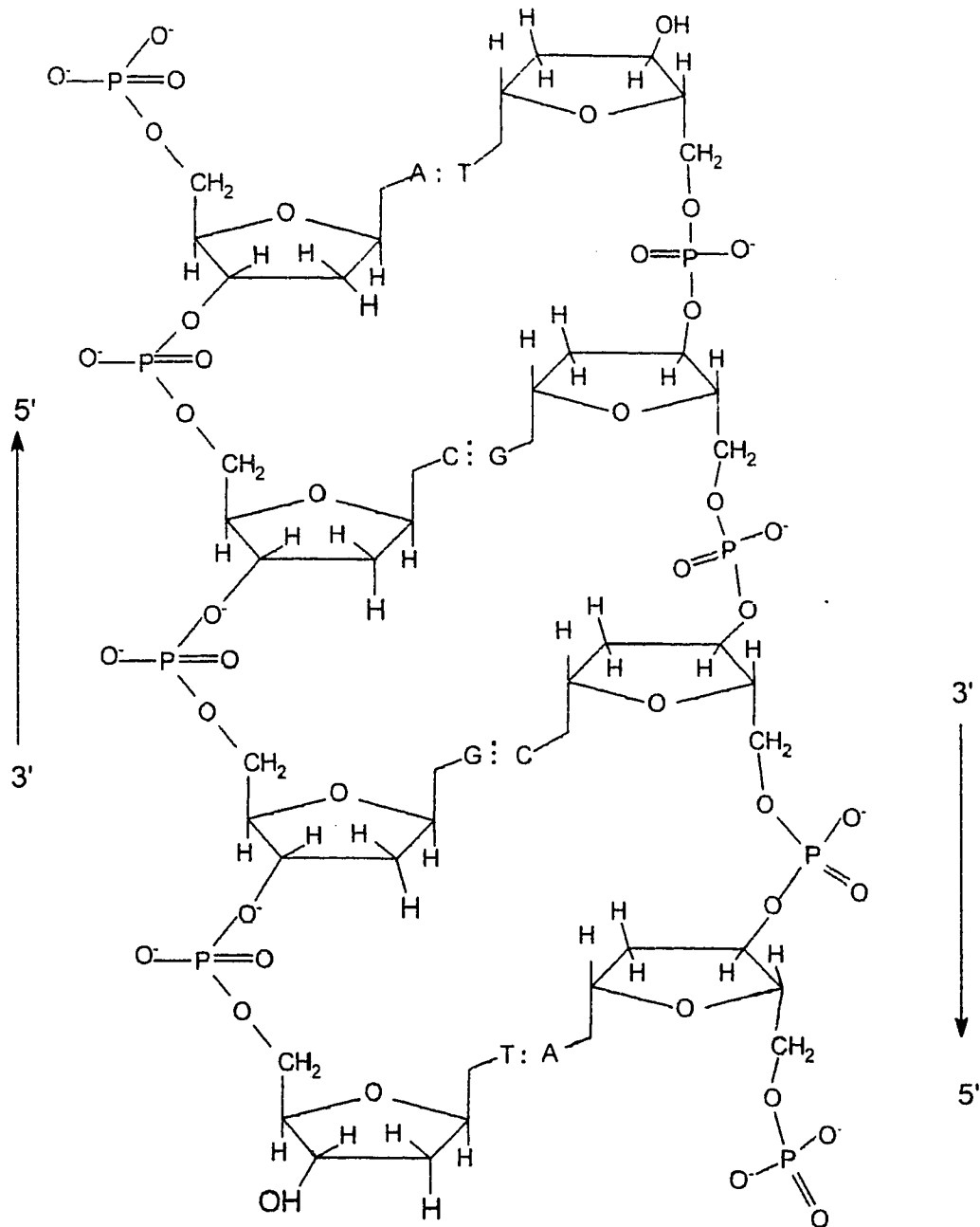


Figure 1.1 Structure of DNA

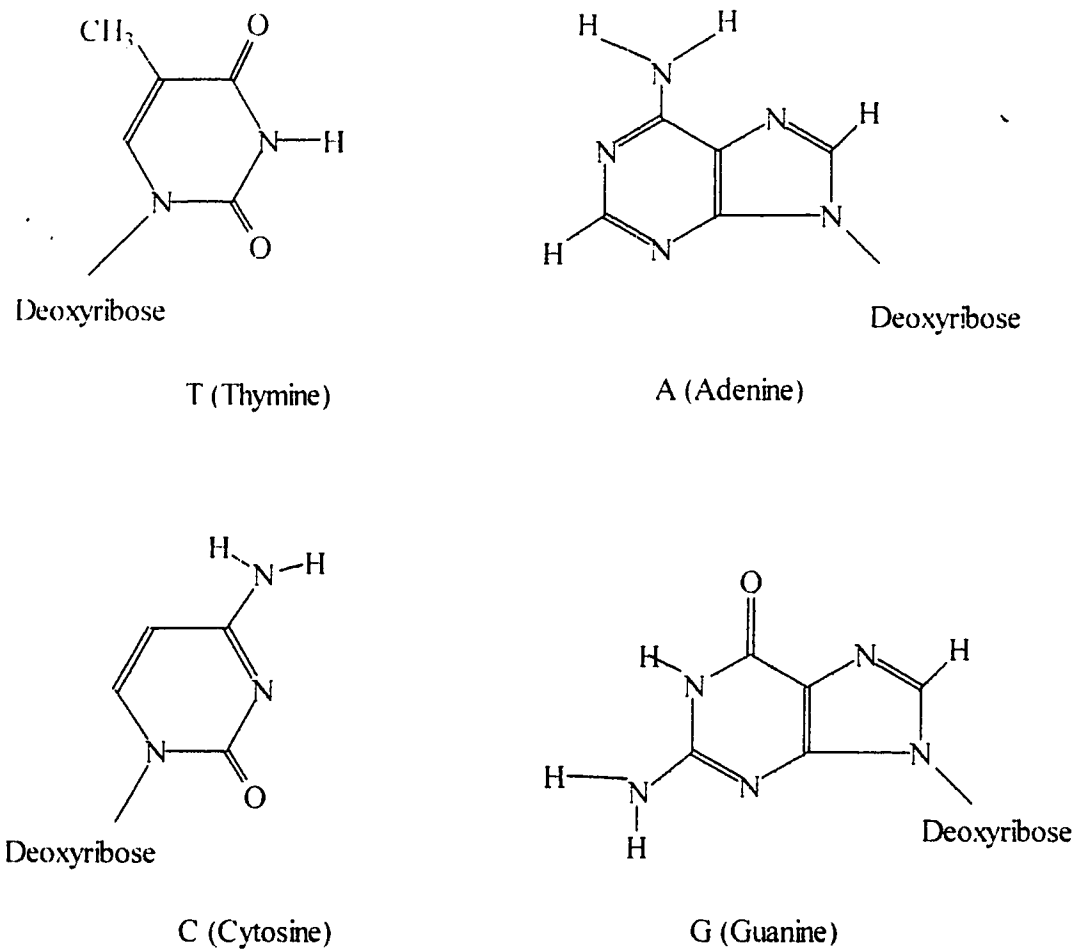


Figure 1.2 Structure of four bases

recent years have come about because we can read and manipulate that information.

The advent of large scale DNA sequencing projects (e.g., the Human Genome Project) has fostered an interest in high-speed, high-throughput DNA sequencing, which drives the continued development of sequencing technologies. The goal of the Human Genome Project is to sequence the human genome within 15 years (by the year 2006). This information is expected to contribute significantly to medical sciences and molecular biology. This effort involves the determination of the sequence from the roughly 3×10^9 base pairs of DNA, which are organized into 23 pairs of chromosomes. With present slab gel technology, 100 of the most advanced automated slab gel systems operating for 60 years would be required to achieve the goals of the Project [3 - 8].

Clearly, methodologies for much higher DNA sequence throughput are necessary. Various technologies (including conventional slab gel electrophoresis, capillary gel electrophoresis, atomic force microscopy, scanning tunneling microscopy, and time-of-flight mass spectrometry, etc.) are currently under evaluation.

Atomic force microscopy [9] or scanning tunneling microscopy [10] would be used to generate a high resolution image of individual DNA molecules. The goal is to identify individual bases in single stranded DNA. The sequence is determined by imaging along the length of the strand. The best resolution for DNA fragments has been between 2 and 5-nm. The dimension of the tip of the scanning probe limits resolution. An order of magnitude improvement in resolution is required to obtain useful sequencing information. Another challenge for DNA sequencing by scanning microprobe is the immobilization of single stranded DNA.

Extraordinary sequencing rates are projected based on the very rapid separation of DNA fragments by mass spectrometry. However, before mass spectrometry can be used for sequencing, three issues need to be addressed. First, the sequencing sample must be

introduced to the gas phase without damage. Second, the sequencing fragments must be separated based on mass. Last, the massive DNA fragments must be detected efficiently. Separation and detection of the sequencing fragments remain the area where the greatest research is required. The ion yields of current mass spectrometry for DNA sequencing are low. Sensitivity for larger fragment is half that of the smaller fragments. In order to achieve practical sequencing accuracy by mass spectrometry, the technique needs to be improved by ten times in mass range, over ten times in resolution, and several orders of magnitude in sensitivity [11, 12].

Gel electrophoresis is by far the most successful technique for DNA sequencing. Capillary gel electrophoresis provides the potential of high sensitivity and high speed analysis of DNA sequencing fragments. To increase the DNA sequence throughput, a multiple capillary system has been developed in this group and will be evaluated in this work for routine DNA sequencing.

1.2 DNA sample cloning and preparation for DNA sequencing by gel electrophoresis

The process of DNA sequencing involves several steps that include DNA sample preparation, separation and detection of the DNA fragments, and data processing and base-calling [8]. Because chromosomes are far too large to be sequenced directly, the first steps in the sequencing process involve breaking the DNA into successively smaller pieces. These smaller pieces are then subjected to the sequencing reaction. DNA cloning and sample preparation are the necessary prelude to any DNA sequencing efforts.

Preparation of DNA for sequencing takes advantage of several molecular biology tools. These tools include restriction enzymes to cut DNA, cloning to introduce pieces of DNA into organisms for propagation, and specific vectors, which are organisms that are genetically engineered to be cloned easily.

Restriction nucleases are one kind of enzymes that can recognize certain sequences and cut DNA near the recognition site. They are used to cleave the DNA molecule at a defined location.

Gene cloning makes use of recombinant DNA techniques [2], that is, inserting a fragment of DNA carrying a gene into a cloning vector. The vector acts as a vehicle that transports the gene into a host cell. A vector (e.g. M13mp18, a most commonly used phage) that contains an insert is called a sequencing template. It is this template that is subject to Sanger's [13] or Maxim and Gilbert's sequencing reaction [14]. Subsequent propagation of the recombinant in a host organism can be achieved. Gene cloning is used to multiply or amplify the DNA concentration.

An alternative method for amplifying the DNA sample without cloning is PCR (Polymerase Chain Reaction) [15, 16]. The PCR is a relatively new technique involving enzymatic amplification of nucleic acid sequences via repeated cycles of denaturation, oligonucleotide annealing, and DNA polymerase extension. It allows virtually any nucleic acid sequence to be readily generated in vitro in relatively great abundance.

The most common sequencing method today is based on Sanger's dideoxy reaction to prepare sequencing fragments. The prepared DNA fragment sample is then sequenced by polyacrylamide gel electrophoresis. Sanger's chain termination reaction generates populations of labeled DNA oligonucleotides (fragments) that begin from a fixed common 5' origin and terminate randomly at a certain residue (base-specific 3' termini). The DNA synthesized by this method is complementary to the template molecule DNA.

This method requires:

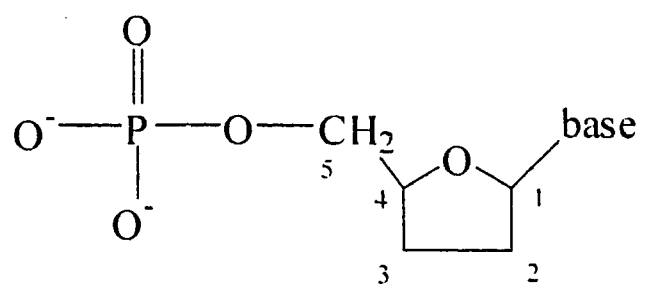
- a) A template, which is the sought-for DNA sample for the replication.

- b) A primer, which is a synthetic oligonucleotide complementary to a specific sequence on the template strand. It defines the starting point for the chain extension reaction. Primer molecules are tagged at the 5' end with a fluorescent dye.
- c) Deoxynucleoside triphosphates (dNTPs) A, G, T and C, which are the four basic building blocks for DNA molecules.
- d) DNA polymerase, which is an enzyme that catalyzes the chain extension reaction.
- e) Dideoxynucleoside triphosphates (ddNTPs), which differ from conventional dNTPs in that they lack a hydroxyl residue at the 3' position of the deoxyribose. They can be incorporated by DNA polymerases into a growing DNA chain through their 5' triphosphate groups.

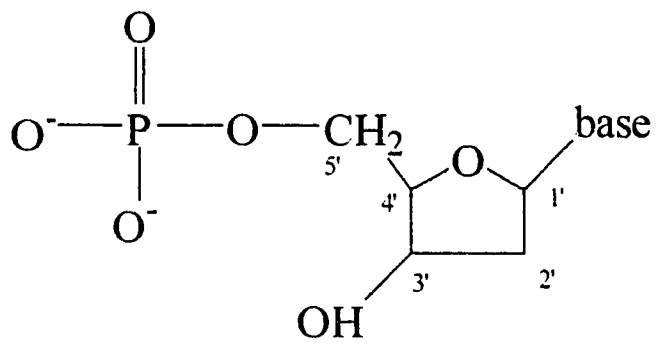
However, the absence of a 3'-hydroxyl residue prevents formation of a phosphodiester bond with the succeeding dNTP. Further extension of the growing DNA chain is therefore impossible. The structure of ddNTP and dNTP is shown in Fig.1.3.

The reaction is usually carried out in four separate vials as illustrated in Fig. 1.4. Each vial has template, fluorescently labeled primer, four different types of deoxynucleoside triphosphates, enzyme, and one type of terminator. The difference is that each vial uses a different fluorescence dye labeled primer and a different dideoxynucleoside triphosphates. The fluorescence labels used for the experiment are fluorescein (FAM), modified fluorescein (JOE), Tetramethyl rhodamine (TAMRA), and Texas red (ROX). Figure 1.5 is the sketches of the emission spectra of four labeling dyes (Fam, Joe, Tamra and Rox) illustrating the relative transmission efficiency of a dye on the four filters.

After the reaction, the products are pooled in one tube. The four different dyes allow



ddNTP (Terminator)



dNTP

Figure 1.3 Structure of ddNTP and dNTP

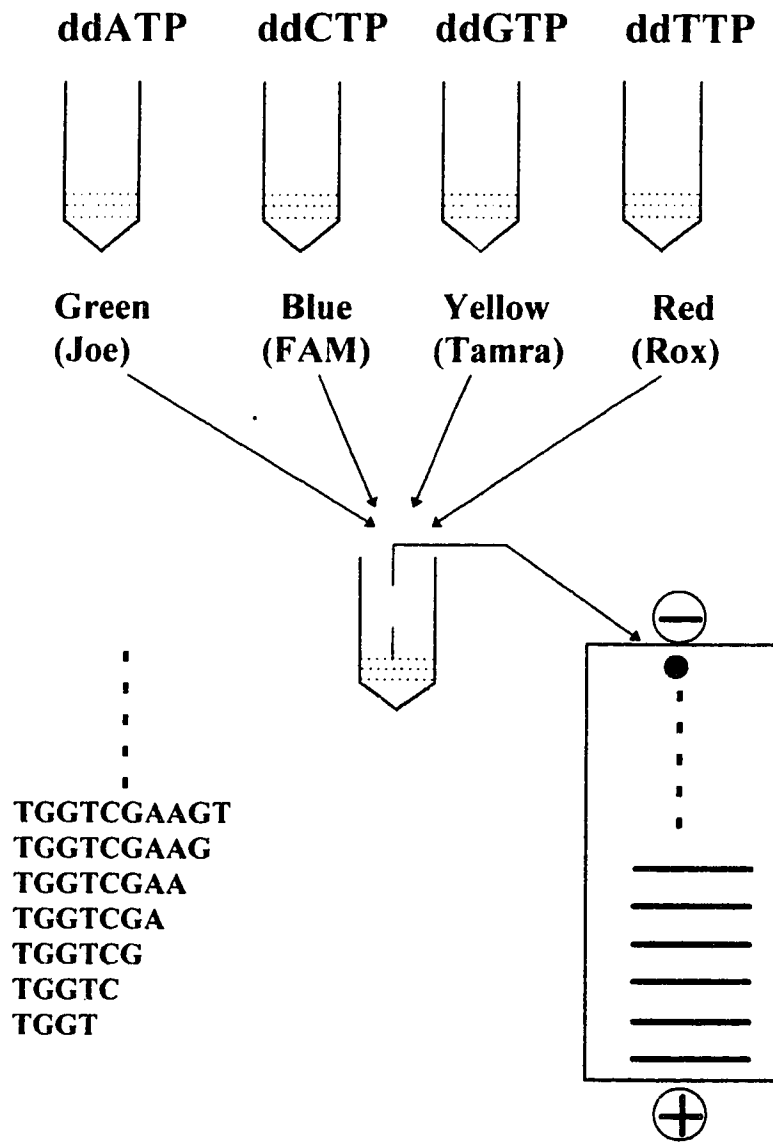


Figure 1.4 DNA sample preparation for gel electrophoresis separation with fluorescence detection

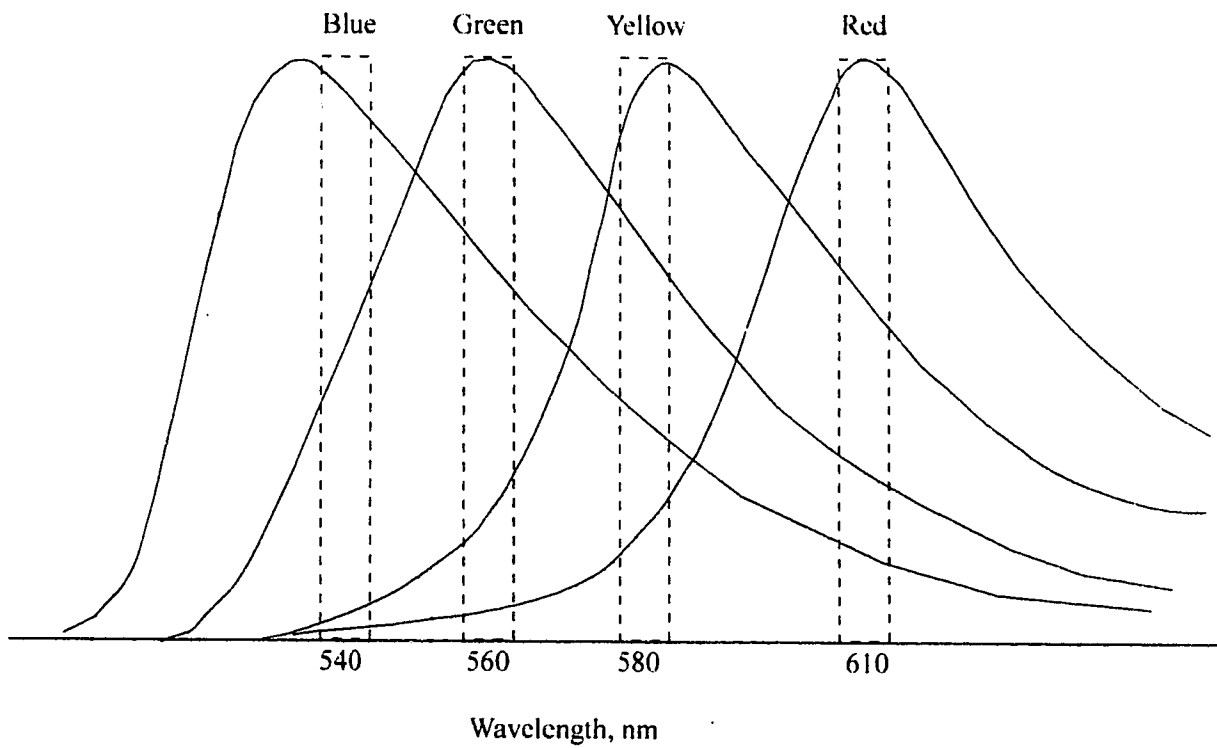


Figure 1.5 Sketches of the emission spectra of four labling dyes (Fam, Joe, Tamra and Rox) with four bandpass filters (dotted lines) illustrating the relative transmission efficiency of a dye on the four filters.

all four termination reactions to be separated in one column and identified by their emission spectrum. Each dye represents the chain termination for a specific base.

1.3 Capillary Gel Separation

Although different methods have been used for DNA sequencing, so far the most widely used DNA sequencing method is gel electrophoresis, which includes both conventional slab gel electrophoresis and capillary gel electrophoresis. Slab gel electrophoresis is simple and inexpensive. However it is also a low speed separation technique [17]. In contrast, capillary gel electrophoresis (CGE) provides advantages in terms of speed, small sample loading, efficiency and automation potential. Capillary gel electrophoresis can easily achieve high sequencing speed, because the large surface-to-volume ratio of the capillary facilitates the use of high electric fields for fast separation and efficient heat dissipation. In addition, the high resistance of the capillary requires low current and generates low heat. Coupled with laser induced fluorescence detection, CGE allows separation and detection of small amounts of DNA sequencing samples (nL of sample is injected).

For DNA molecules, each phosphate moiety carries a negative charge and it will migrate in an electric field. With the cathode on the injection side and the anode on the detection side, the negatively charged DNA molecules migrate toward the anode. However, the separation of different sized DNA fragments can not be achieved successfully by free zone (no gel) electrophoresis. The use of gel in DNA sequencing by electrophoresis is critical. Separation by free zone (no gels) electrophoresis is based on the difference of charge-to-mass ratio of the analytes. In free zone electrophoresis, all DNA molecules will migrate with roughly the same velocity because the charge on the molecule increases as the size of the molecules increases [17]. However, in gels, the DNA molecules are separated on the basis of size while the electric field is the driving force to

push the molecules through the gel. The gel acts as a sieving medium. Gel electrophoresis will resolve DNA fragments differing only by one base in size. CGE was first developed by Hierten for the separation and molecular weight determination of peptides and proteins [18].

Separation of charged polymer species is the main domain of capillary gel electrophoresis [19]. Capillary gel electrophoresis has been applied to the separation of DNA fragments using either ultraviolet (UV) absorbance [20-22] or fluorescence detection [23-28]. Swerdlow and Gesteland reported the separation of a single dideoxynucleotide reaction mixture with a single spectral channel laser-induced fluorescence detector [24]. Drossman and co-workers in Smith's laboratory also reported the separation of a single reaction mixture in gel filled capillaries [25]. A subsequent report from Smith's laboratory described the use of capillary gel electrophoresis for the separation of the reaction products of the four-spectral-channel, single-lane-sequencing system [26]. Resolution was a factor of 1.5 to 2 times superior to slab-gel data. Gesteland's laboratory has reported the use of a charge-coupled-device for fluorescence detection in DNA sequencing by CGE [27]. Our group has reported several designs for fluorescence detection in CGE for DNA sequencing [29-31].

In capillary gel electrophoresis, two types of separation media are presently used, i.e., non-cross-linked and cross-linked polymers [32]. Those of the first type, involving relatively low-viscosity polymer networks, such as those consisting of linear polyacrylamide or alkylcellulose, are often referred to as "replaceable". The second type of polymer is a relatively high-viscosity crosslinked gel that is "fixed" inside the capillary. Although there is a radical difference in the gel structure, the separation generated by the two type of gels seems similar. However the non-cross-linked gel is much easier to handle and replace.

The polymer is covalently bound to the capillary wall through use of γ -methacryloxy-propyltrimethoxysilane (MAPT), a bifunctional agent. One functional group of this reagent is intended to react with the fused-silica surface while a double bond is available for copolymerization with acrylamide/bisacrylamide during the formation of the gel. The use of a coated capillary in conjunction with the polymer network results in negligible electroosmotic flow. Polymerization of the polyacrylamide gel is usually initiated by ammonium persulfate and catalyzed by N, N, N', N' - tetramethylethylenediamine (TEMED) [33, 34].

High-efficiency, fast separations of DNA molecules can be achieved by the use of linear polyacrylamide capillary columns [35-38]. Linear polyacrylamide has relatively low viscosity. In principle, it may be pumped from the capillary after use [37]. By refilling the capillary after every separation, it would be possible to eliminate realignment of the optical systems in automated DNA sequencers [38].

1.4 Models of DNA separation in gel electrophoresis

One way to decrease the time required for electrophoresis is to increase the electric field driving force. High electric fields, however, also lead to a reduction in the spacing between bands with a concomitant loss of resolution. In the context of DNA sequencing, the magnitude of the mobility difference between adjacent bands plays a dominant role in the resolution of the bands on a gel, and the electrophoretic mobility of DNA is found to be a function of the applied electric field.

A number of theories have been developed to model the migration behavior of DNA in electrophoresis [19, 39-47]. These models can provide insight into the separation mechanism to understand how the matrix influence the separation, and provide useful

guides for the optimization of the separation. These models tend to fall into two classes based on the size of the fragment, Ogsten sieving model and biased reptation model.

1.4.1 Ogsten Sieving Model

In this model, the gel is treated as a molecular sieve formed by randomly distributed fibers, which form randomly distributed pore size. Small fragments are usually assumed to exist as a random coil with mobility governed by Ogsten sieving, wherein the mobility of a fragment is proportional to the fraction of pores in the gel that are larger than the fragment.

$$\mu = \mu_0 e^{-kC\lambda N} \quad (1.1)$$

where μ_0 is the free solution mobility, C is the gel concentration, λ is related to the radius of gyration of the DNA fragments, k is a constant of proportionality, and N is the fragment length in bases. If DNA fragments migrate according to the Ogsten model, a plot of fragment length versus log mobility will be linear.

1.4.2. Biased reptation model

The Ogsten model is observed to fail for larger fragments. Instead, the fragment mobility is described by the biased reptation model, in which the fragments are assumed to thread their way through the gel, much as a snake moves through grass [40, 43-47]. The term reptation is derived from the presumed alternating periods of stretching and relaxation that occur as the molecule moves through the small pores [19]. The mobility of the DNA is predicted to be inversely related to fragment length.

$$\mu \approx \chi \left[\frac{1}{N} + \left(\alpha \frac{E}{T} \right)^\beta \right] = \chi \left[\frac{1}{N} + \frac{1}{N^*} \right] \quad (1.2)$$

Where χ and α are constants, T is the absolute temperature, and β is an exponent that defines the relationship between electric field and the onset of reptation. E is electric field. N is fragment length in bases, N^* is the fragment length for which the onset of biased reptation becomes noticeable. Which is when pieces of DNA are aligned by the electric field. These aligned fragments move through the gel at a rate that is independent of their size. This movement leads to an increase of mobility with the increasing field. The first term in the equation is length dependent and electric field independent, whereas the second is length independent and electric field dependent. These terms are the first two terms in an expansion based on fragment length; the model will not hold for very long fragments.

The classic biased reptation model predicts that $\beta = 2$ [19, 39-41]; the onset of biased reptation is predicted to decrease with the square of electric field. A new model, called biased reptation with fluctuations (BRF), predicts instead that $\beta = 1$; the onset of biased reptation is inversely proportional to electric field [45-47]. The value of β is quite important in practical applications. A high value of β will severely limit the sequence read length at high electric fields.

There have been studies of the value of β for double stranded DNA in agarose and in hydroxypropylcellulose. There have, however, been few studies of the value of this parameter for the separation of single stranded sequencing fragments. Since single stranded DNA is much less stiff than double stranded DNA, single stranded DNA is expected to have different behavior than double stranded DNA. In chapter 3, I will discuss, in detail, the electric field behavior of DNA sequencing fragments at high electric fields for non-crosslinked polymers. Results of the value for β obtained from our experiments will also be discussed.

1.5 Scope of this project

Capillary gel electrophoresis has demonstrated great advantages for DNA sequencing. However there is room for improvement in order to make the CGE system a routine DNA sequencing tool. The following aspects will be discussed in this work.

1. Instrumentation. The five capillary CGE system needs to be improved to make it reliable and easy to use for large scale DNA sequencing analysis. This aspect of my thesis mainly includes the automatic data acquisition and processing and will be discussed in Chapter 2 and Chapter 4.
2. Sample preparation. The quality of the template DNA, the type of DNA polymerase, the ratio of dNTP to ddNTP, the labeling reaction, and the experimental conditions directly affect the quality of the prepared DNA sample, which in turn determines the final sequencing read-length and accuracy. These parameters were studied and discussed in Chapter 3.
3. Gel preparation. The quality of gel directly influences the separation efficiency. Parameters affecting the gel quality were studied and discussed in Chapter 3.
4. Sample injection conditions. Sample injection may also affect precision and accuracy as well as separation performance. An over-loaded sample may jeopardize the resolution, while the under-loaded sample may result in insufficient sensitivity.

In this thesis, I will focus mainly on the multicapillary CGE system and the parameter studies which affect the separation efficiency. Large amounts of sequencing data are generated. Automatic data processing and base calling will also be addressed.

Chapter 2

Instrumentation and experiment

Capillary electrophoresis has had a long history, if one considers electrophoresis conducted in columns as its ancestor. In the 1930's, Tiselius [48] inaugurated the modern science of electrophoresis with an elegant series of experiments in a tube apparatus.

The instrumentation used for capillary electrophoresis is relatively simple. Fig. 2.1 is the schematic diagram of a typical capillary electrophoresis system with a post-column detector. The main parts of the system are described as follows.

The high voltage supply (Spellman CZE 1000 R, Model: CZE 1000 PN 30, 0-30 KV Plainview, NY. USA) provides a high voltage driving force for electrophoresis through a platinum electrode that is inserted into the buffer vial. With sample at cathode side and with detection at anode side (Anion mode), samples are injected electrokinetically into the capillary by immersing the cathodic end of the column in the sample vial and applying a voltage. The negatively charged DNA molecules migrate toward the anode. For safety, the electrode connected to the high-potential side of the power supply is surrounded by a Plexiglass box equipped with a safety interlock. The electric circuit is completed by connecting the sheath flow cuvette holder to ground through a 100 k Ω resistor. The voltage drop across the resistor is used to monitor the current change during the process of electrophoresis.

The capillary outlet is inserted into a sheath flow cuvette that serves as a post-column fluorescence detection chamber [49, 50]. In contrast to on-column detection, the sheath flow cuvette provides low background and high sensitivity. On-column detection is simple, but laser light is scattered at the curved surface of the capillary. This scattered

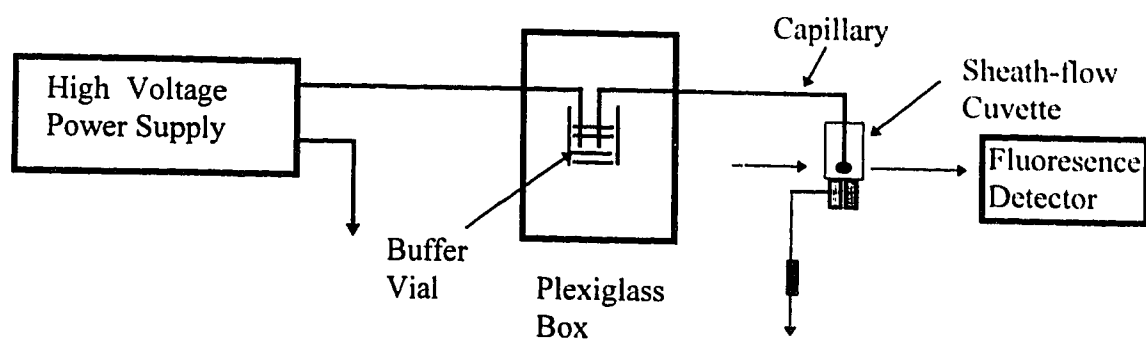


Figure 2.1 Schematic diagram of a typical capillary electrophoresis system with post-column detector.

light will limit the sensitivity of detection. The light scattering problem can be virtually eliminated by using a flat cuvette window in post-column detection. In the sheath flow cuvette, the eluting sample flows as a narrow stream through the center of the cuvette, while buffer flows around the capillary through the cuvette. The sheath stream is generated by gravity-driven siphon flow. No light scatters from the sample-sheath flow interface because the sheath flow buffer composition is made the same as the running buffer and there is no refractive index boundary between them.

2.1 Multiple capillary electrophoresis system

The five capillary CGE system was first designed and setup by Dr. Jiangzhong Zhang. More detailed information can be found in his PhD thesis [51]. Some modifications are implemented in this work and will be discussed later in this Chapter.

Figure 2.2 is the overview of the five capillary CGE system. The major components of the system are described as follows. Because the sequencing products are labeled with four different fluorescent dyes, two lasers are used to excite these dyes (see Fig. 2.2). One is an argon ion blue laser (Cyonics, 488 nm) and the other is the helium-neon green laser (Melles Griot, 543.5 nm). The 488 nm blue laser generates a strong Raman band at ~ 580 nm from the aqueous buffer. This Raman band overlaps with fluorescent emission excited by the green laser. To minimize Raman scattering effects, a sector wheel is used to alternately block each laser beam. The laser beams are then focused onto the outlet of the capillary array, which is held inside the sheath flow cuvette.

The sheath flow cuvette is the key to the fluorescent detection. A modified sheath flow cuvette (150 μm deep, 750 μm wide in cross-section, 2-cm long, and 1-mm thick windows) is used in this system to house five capillaries [51]. Figure 2.3 is the schematic

diagram of the multiple capillary sheath flow cuvette. The narrow wall of the cuvette is designed to be 50 μm wider at the top than at the bottom.

As the linear array of capillaries is inserted into the cuvette, they are squeezed together and equally spaced. The sheath flow draws the sample streams from each capillary into the flow chamber. The sample streams should be aligned like the teeth of a comb in the flow chamber. Since the refractive indices of the sheath and sample streams are identical, a single laser beam is used to illuminate all the sample streams. Because of the minute absorbance of the highly dilute analyte, there is no significant attenuation of the laser beam in traversing the array of sample streams. Furthermore, the high optical quality of the flow chamber minimizes light scatter, which produces a low background signal and excellent detection limit.

Fluorescence is collected at 90 degrees by a 20 \times , 0.50 numerical aperture microscope objective. The image produced by the collection optics is an array of 1-mm diameter spots with 3-mm spacing (center-to-center). Avalanche photodiodes are used as the photon detectors. These detectors have an active area of 0.2-mm diameter and are provided in 6-mm wide cans. This aspect ratio is not compatible with the fluorescence image. Instead, an array of fiber optics is used to couple the fluorescent image into the photodiodes. However, the fiber optics have 100- μm core, which is much smaller than the 1-mm diameter image of the sample stream produced by objectives. An array of GRIN lens (Gradient Refractive index lens, which are able to focus a collimated beam into a small spot), with 1.8-mm diameter (NSG part number FCM-0F-100-063), are used to couple fluorescence into the fiber optics in order to maximize the collection efficiency.

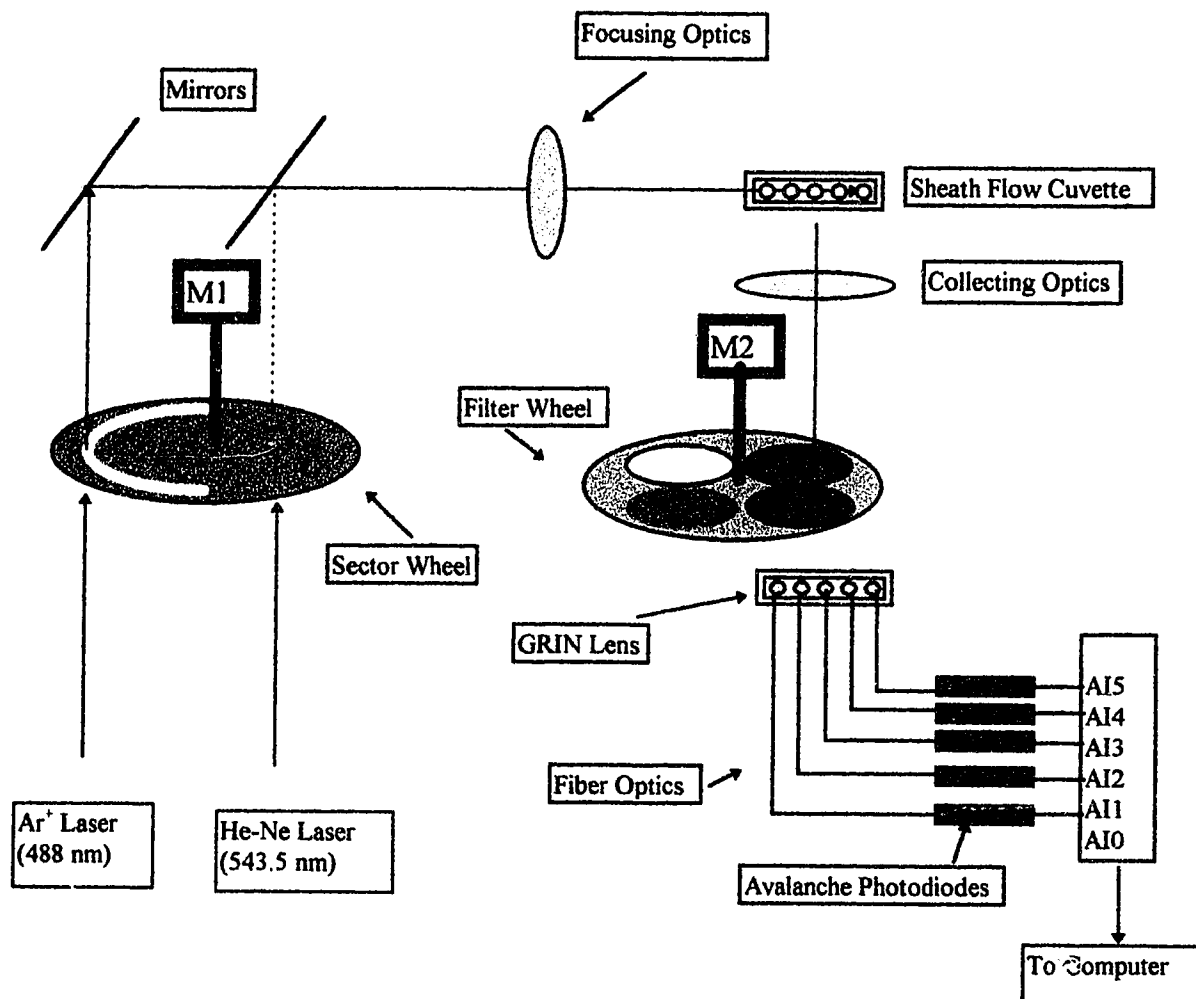


Figure 2.2 Overview of the five capillary system

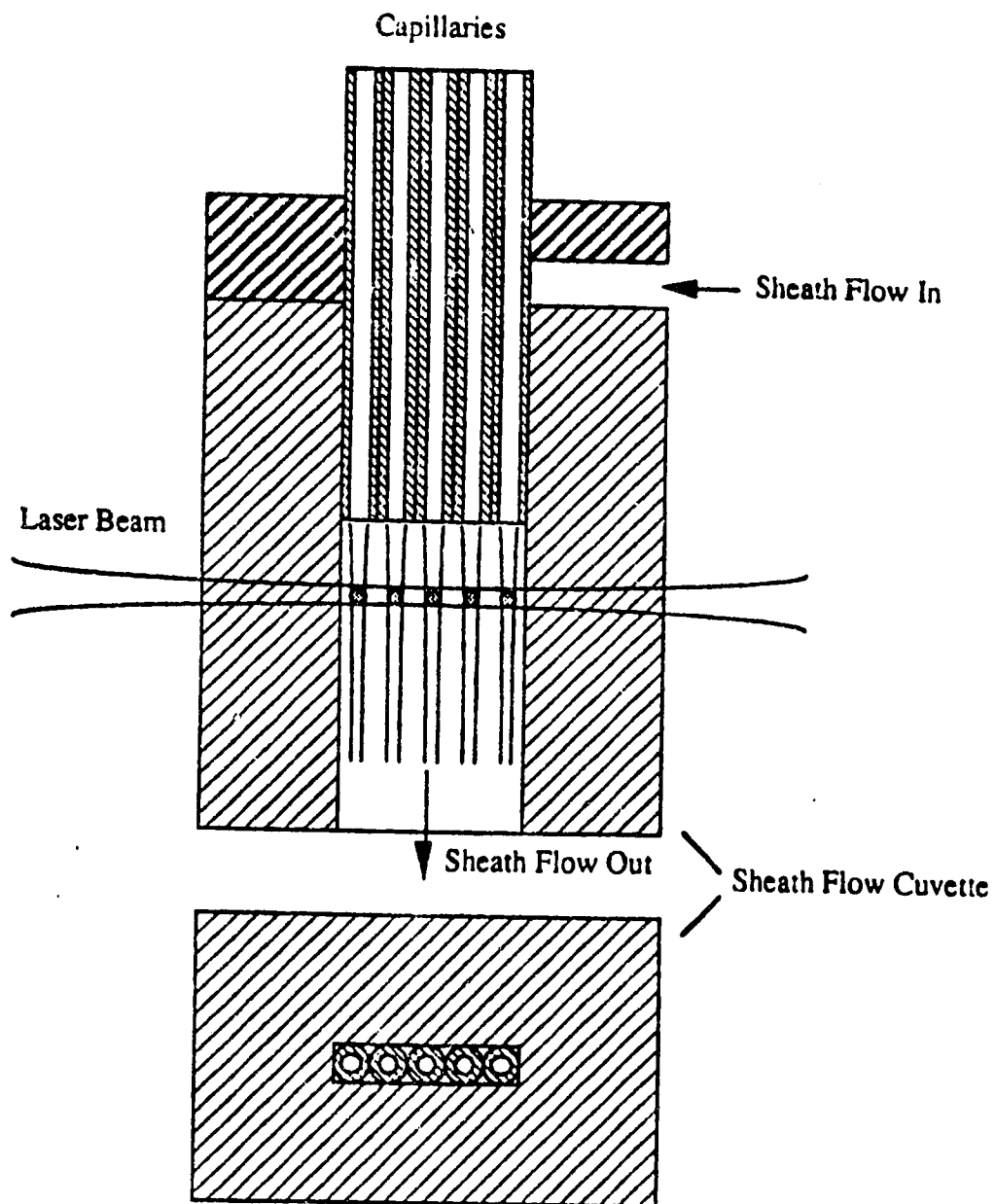


Figure 2.3 The multiple capillary sheath flow cuvette (from reference 51)

Between the GRIN lens and the objective is a rotating filter wheel, which is synchronized with the sector wheel through use of stepper motors with a common controller. The filter wheel has four band-pass filters mounted on it (540 nm, DF10; 560 nm, DF10; 580 nm, DF10; and 610 nm, DF10; Omega, VT, USA), which are used to isolate the emission spectra of the four dyes. The fluorescence images are collected by the GRIN lens and transmitted to the avalanche photodiodes through a set of fiber optics. The fluorescence signals are then detected by avalanche photodiodes (EGG, Canada SPCM), which work in the Geiger mode and provide photon counting output. The signal from these avalanche photodiodes is a train of pulses. The rate of the pulses is proportional to the intensity of the fluorescence. A frequency-to-voltage converter (AD 650 chip for V/F or F/V converter) is then used to convert the pulse rate signal to an analog signal. The voltage signals are digitized by a data acquisition board (National Instruments NB-MIO-16X, with 16 analog input, two DAC channels and 16 bit digital I/O) plugged in a Macintosh Quadra 700 computer.

It should be pointed out that in DNA sequence determination using four dye systems, the selection of excitation and emission wavelengths are not at their spectral maxima because of the limited choice of laser lines and required spectral resolution [51].

In this system, the collection optics were fixed in space. The optical alignment was achieved by flowing a fluorescein solution through the capillaries. The position of both the sheath flow cuvette and the laser beam were adjusted until a maximum fluorescence signal was reached from each photodiode.

2.2 Data acquisition subsystem

When the instrument was first developed, the data acquisition was not synchronized with the filter wheel. Instead, the data acquisition speed was set much faster than the filter

wheel rotation speed to ensure no data points were missed for each fluorescence wavelength. More data were acquired than necessary and the data processing was complicated by the irregular adding or missing of data points for certain filters. The system has been modified by the use of a photo-interrupter, which is added to the sector wheel to sense the position of the four filters (See Figures 2.4 and 2.5). The photo-interrupter is a gallium arsenide infrared emitting diode facing a silicon NPN photo-transistor in a model plastic housing. A slot in the housing between the emitter and the detector provides the means for mechanically interrupting the infrared beam. Figure 2.4 shows the schematic diagram of the photon interrupter and its out put.

The output of the photo-interrupter is used to trigger data acquisition. One data point is recorded per filter. In this way, the data acquisition is synchronized with the filter position. The data obtained is a four color interleaved array, and the data processing is straight forward. The software used for controlling and data acquisition is written in LabView 2.0 (National Instruments). The virtual instrument (VIs) library of LabView consists of the graphic icons of different functions. By connecting different VIs (icons) with proper lines (input/output), some data acquisition and processing steps, such as ADC/DAC, digital input/output, and data manipulation and displaying, can easily be achieved with minimum effort and error.

The filter wheel is set to 2 revolutions per second (2 rps) and, therefore, the data acquisition rate is 8 Hz per channel (2 revolutions per second \times 4 filters/revolution), which is a data rate of 2 Hz per color for each capillary. This data acquisition speed is appropriate for our application, because the narrowest peak signals found with our typical operation conditions is about 3 to 5 seconds at FWHM (Full Width at Half Maximum).

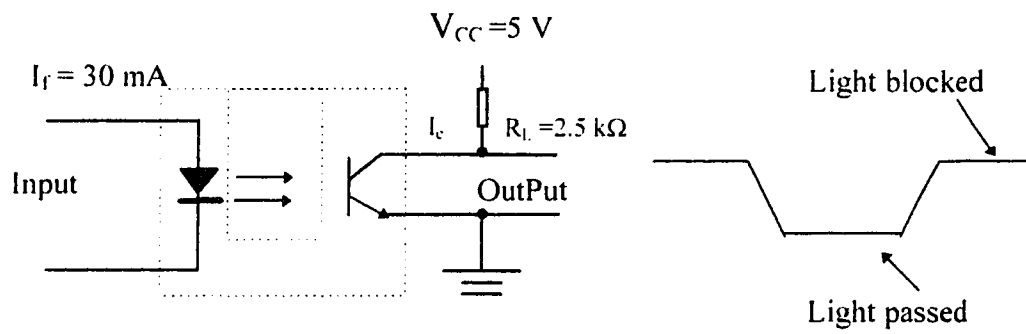


Figure 2.4 Photo-interrupter and its output

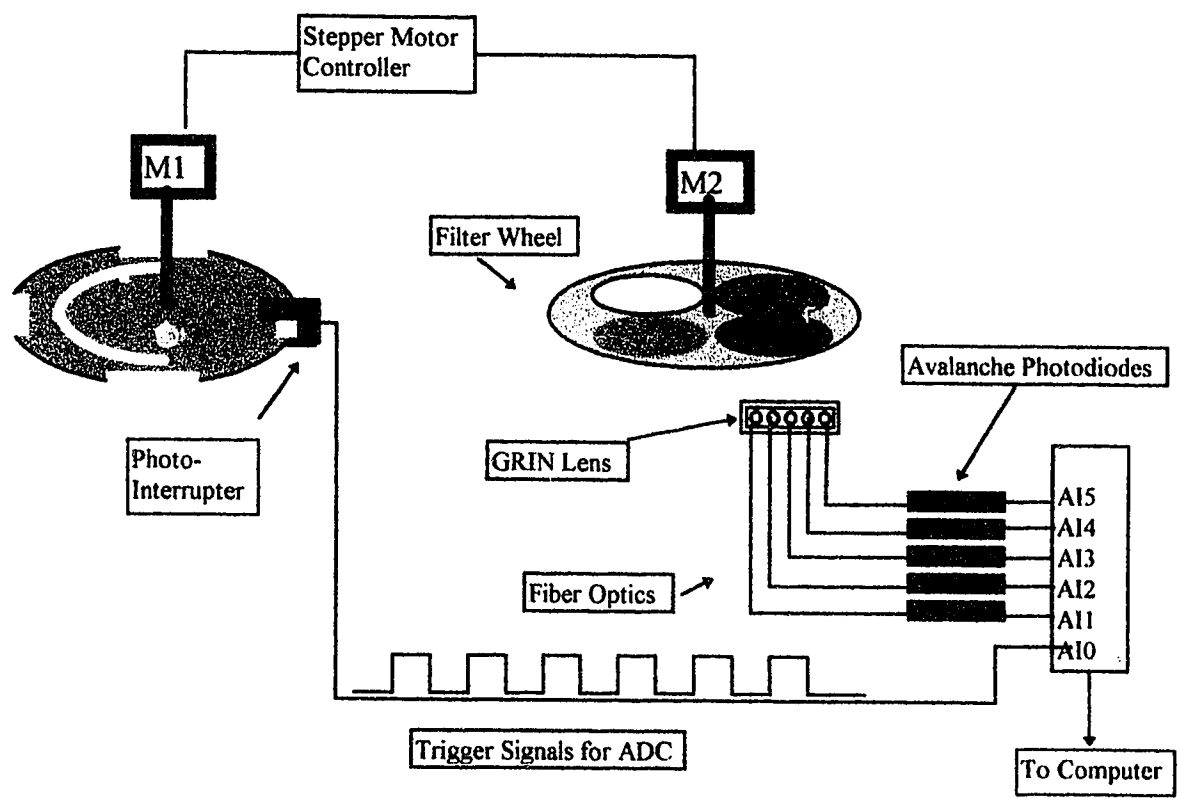


Figure 2.5 Data-acquisition subsystem

2.3 Detection Limit

To ensure that the local electric field is not perturbed by the ionic strength of analyte, it is necessary that the ionic strength of the analyte be less than 1% of the ionic strength of the buffer [52]. Although the buffers used in gel electrophoresis have concentrations in the high-millimolar range, they are poorly ionized and the total ion concentration of the buffer is about 10 mM. Therefore, the total ionic strength of the sample in the electrophoresis gel must be less than 100 μM to minimize electric field artifacts. In conventional sequencing reactions, at least half the analyte molecules are unlabeled template, since only one sequencing fragment is produced per template. The template, which is about 7500 bases long, has one negative charge per base and most of the ionic strength of the sample is associated with template. If the average sequencing fragment is 300 bases in length, only 4% of the ionic strength of the sample is associated with the analyte; the ionic strength associated with the DNA sequencing fragments is 4 μM . However, since each analyte molecule has, on the average, 300 bases, the average concentration of fluorescently labeled analyte is about 1×10^{-8} M. Clearly, very sensitive detection is required for DNA sequencing.

Laser-induced fluorescence is well suited for detection in capillary electrophoresis. The spatial coherence of the laser allows very small volumes to be excited with high intensity. When coupled with efficient collection and detection, minute amounts of analyte may be detected. Detection limits in fluorescence measurements are usually determined by shot-noise in the background signal. Parker [53] listed four sources of background signal in fluorescence: fluorescence from solvent impurities, fluorescence from cuvette windows, Raman and Rayleigh scatter from solvent, and light scatter at the cuvette-sample interface. Detector dark current is a fifth source of background. To

produce excellent detection limits, one must minimize each of these sources of background signal.

The detection limit of the system is first determined with free zone electrophoresis by injecting appropriate fluorescence solutions. The measurements are made in replicates for both lasers.

2.3.1 Detection limit obtained for the blue laser

A fluorescein stock solution was prepared by dissolving 0.00026 g fluorescein (high purity standard from Molecular Probes, OR, USA) in 1.00 mL of ethanol (HPLC grade) by use of an ultrasonic bath. Then, the solution was serially diluted using 10 mM borate buffer to a final concentration of 1.56×10^{-12} M. The solution was injected at 1000 V for 5 second into a capillary with a length of 37.1 cm (i.d. = 50 μ m, fused-silica capillary, Polymicro Technologies, Phoenix, AZ., USA). After injection, the sample vial was replaced with a vial containing the 10 mM borate running buffer, and an electric field of 300 V/cm was applied for performing the electrophoresis.

The detection limits are calculated as follows [54]. The volume of sample injected (Vol_{inj}) is given by

$$Vol_{inj} = (Vol_{cap} \times t_{inj} \times V_{inj}) / (t_m \times V_{CE}) \quad (2.1)$$

Where, t_{inj} is the injection time, Vol_{cap} is the volume of the capillary, V_{inj} is the injection voltage, t_m is the migration time of the analyte, and V_{CE} is the running voltage

The amount of sample injected is given by

$$Q_{inj} = Vol_{inj} \times C \quad (2.2)$$

where C is the analyte concentration.

The limit of detection is given by

$$\text{L.O.D} = 3\sigma (\text{Amount of sample injected}) / h \quad (2.3)$$

where σ is the standard deviation of the background signal, or

$$\text{L.O.D} = 0.65(h_2)(\text{Amount of sample injected}) / h \quad (2.4)$$

Where h is the peak height of the signal peak and h_2 the peak-to-peak noise of the baseline. Usually the peak-to-peak noise variation of the baseline (h_2) of the electropherogram is about 5σ and, hence, 3σ is approximately equal to $0.6 h_2$. Therefore the detection limit can be easily estimated from the electropherogram of a single injection of the analyte with concentration about 10 times that of the detection limit [55]. The measured detection limit with the Ar^+ laser excitation is 160 ± 44 molecules of fluorescein injected onto the capillary. An electropherogram used to estimate the detection limit for Ar^+ laser is shown in Fig. 2.6 (a).

As mentioned in Section 2.1, one of the modifications of the system is that a F/V converter is added to convert the pulsed signal into analog signal in order to use the buffered analog input VIs in LabView. The results show that the detection limit obtained with F/V converter is comparable to those obtained with direct pulse counting method (130 ± 30 molecules) [51]. The F/V converter did not add discernible noises.

2.3.2 Detection limit obtained for HeNe green laser

Rhodamine 6G was used here as the test solution to determine the detection limit for the HeNe green laser. The stock solution of Rhodamine 6G was prepared in 10 mM borate buffer. The final concentration of the solution was 9.4×10^{-13} M. The sample was injected at 1000 V for 5 second. The capillary length was 39.5 cm, and the i.d. was 50 μm . The electrophoresis was performed under the electric field of 300 V/cm. The

measured detection limit is 180 ± 40 rhodamine 6G molecules. An electropherogram used to estimate the detection limit of the HeNe laser is shown in Figure 2.6 (b).

2.4 Five channel cross-talk test

The interference between the five channels was examined to determine the possibility of cross-talk between different channels. In this experiment, a sample was injected into ch1 and ch3 while ch0, ch2 and ch4 had no samples. The electrophoresis was run with typical parameters (see Table 3.2). The results are shown in Fig. 2.7. Ch1 and ch3 gave normal sequence electropherograms, while other channels were free of signal. Some tiny spikes were seen; and they are not due to the other channel's signal but they are instead caused by particles or small bubbles in the buffer. The spikes are much sharper than normal signal peaks and they can be easily recognized and removed in the data processing step.

The experiment results proved that there is no interference between these five channels.

2.5 Separation matrix

The separation matrix used for the experiment is linear polyacrylamide. The polymer matrix plays an important role in gel capillary electrophoresis. The property and quality of the polymer directly determine the separation efficiency. Great attention and care should be paid in the polymer preparation. The polymer preparation procedure that I used is as follows.

A stock solution of $10 \times$ TBE (Tris-Boric-EDTA: Tris 0.89 M; boric acid 0.89 M; EDTA 26 mM) buffer solution was prepared from 5.40g Tris (Tris(hydroxymethyl) amino methane, F. W. 121.14; Ultra pure, ICN Biomedicals, Inc), 2.75g boric acid

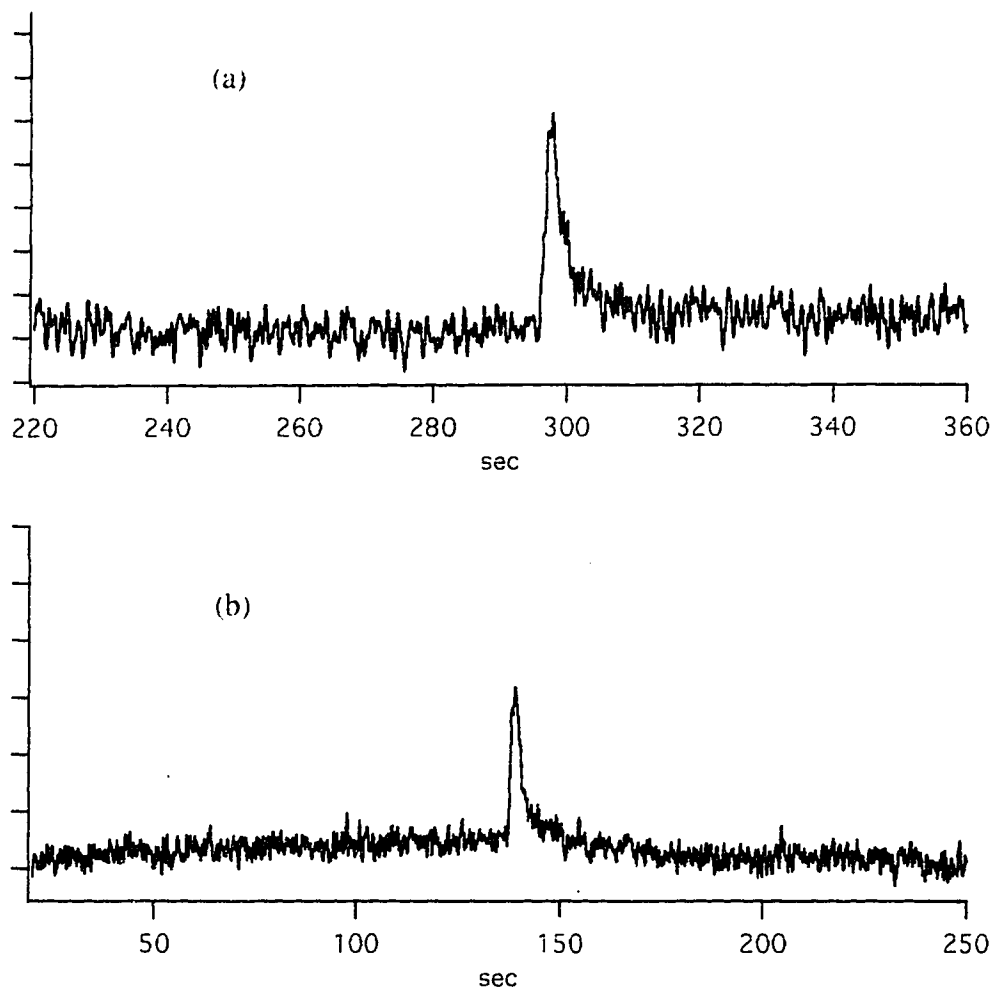


Figure 2.6 Electropherograms of (a) fluorescein (1.56×10^{-12} M, excited by Ar ion blue laser, 488nm) and (b) Rhodamine 6G (9.4×10^{-13} M, excited by He-Ne green laser, 543.5 nm).

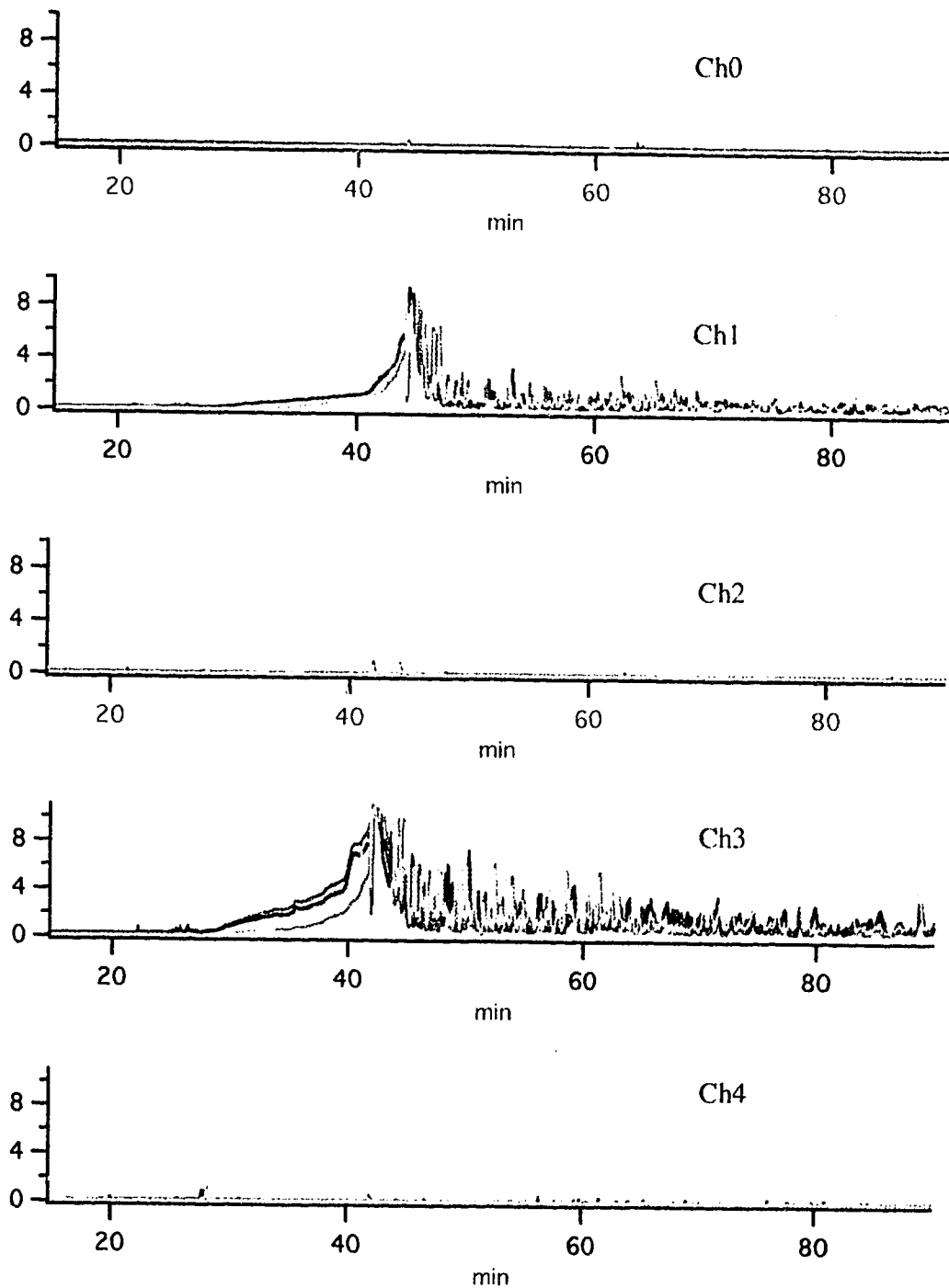


Figure 2.7 Electropherograms of the five capillaries with sample injected into ch1 and ch3 and no samples are injected into ch0, ch2, ch4

(H_3BO_3 , F.W. = 61.38, BDH Inc.), and 0.48 g $\text{EDTA}\cdot 2\text{Na}\cdot 2\text{H}_2\text{O}$ (Ethylene-diaminetetraacetic acid disodium salt, $\text{C}_{10}\text{H}_{14}\text{N}_2\text{O}_8\text{Na}_2\cdot 2\text{H}_2\text{O}$, F.W. = 372.24, BDH Inc.). The solids were dissolved in water, diluted to a total volume of 50 mL with water, and filtered through a 0.22 μm filter. The 5 \times TBE and 1 \times TBE (89 mM Tris, 89 mM boric acid, 2.6 mM EDTA) buffers were obtained by dilution of 10 \times TBE with deionized water.

A 40% acrylamide stock solution was prepared by dissolving 20 g of acrylamide (>99.9% purity, F.W. = 71.08; Bio-RAID, electrophoresis purity reagent) in water to make a final volume of 50 mL. An aliquot of 3.1 mL was taken from the 40% acrylamide stock solution and mixed with 10.5 g of urea (Ultra pure; ICN Biomedicals, Inc. H_2NCONH_2 , F.W. = 60.06) and added to 5 mL of 5 \times TBE buffer (Tris-Boric-EDTA). A few milliliters of water was added and the solution was shaken occasionally to dissolve the urea. Then water was added to make a final volume of 25 mL. The solution was then passed through a 0.22 μm filter under a vacuum provided by a water aspirator. The solution was stored in a refrigerator of 4°C. Water used here was from Omni Solv. BDH. Inc., and filtered through a 0.2 μm filter.

Polymerization solutions were degassed with a water aspirator for 20 min. prior to addition of the initiator and catalyst. For each 5 mL polymerization solution, 2 μL TEMED (N,N,N',N'-Tetramethylethylenediamine, Electrophoresis Grade, F.W. = 116.21, purity: > 99%, Ultra Pure; Life Technologies. Inc.), and 20 μL 10 % AP (Ammonium persulfate, Electrophoresis Grade, F.W. = 228.20, Ultra Pure, Life Technologies. Inc.) were added to initiate the polymerization.

To increase the stability of the polymer mixture in the capillary, the capillaries are silanized by flowing a 1% silanizing solution through them before preparing the matrix. The silanization solution was prepared by mixing 10 μL γ -methacryloxypropyltrimethoxysilane (Sigma, St. Louis, MO, USA) with 1 mL (5% water

+ 95% ethanol) or with 0.5 mL water and 0.5 mL acetic acid. After approximately 30 minutes, all the capillaries were filled simultaneously with the same acrylamide monomer mixture by using a vacuum system to draw the polymer solution through the capillaries. This procedure leads to similar gel composition and retention time. Capillaries were left overnight for complete polymerization before use.

Before doing a separation, the capillaries were prerun (pre-electrophoresed) with 1× TBE buffer for about an hour at 100 to 150 V/cm. This pre-electrophoresis run helped to remove the residual catalysts and monomers.

Chapter 3

Study of separation conditions

To obtain optimum separation conditions for DNA sequencing, some basic experiments have been carried out with this multiple capillary system. One advantage of the multiple capillary system over a single capillary system is the convenience of parallel comparison. Capillaries of different gel compositions can be tested at the same time to eliminate systematic variation in separation voltage and temperature. Capillaries of the same gel composition can be run in replicate to monitor measurement fluctuations and to minimize the effects of day-to-day variation in gel composition.

By analogy to chromatography, the separation efficiency of an electrophoretic system can be expressed in terms of theoretical plates. The number of theoretical plates (N_p) [55] is estimated as

$$N_p = 5.54 (T_m / W_{1/2})^2 \quad (3.1)$$

where T_m is the migration time and $W_{1/2}$ is the full-width at half-height of the peak.

Resolution is defined as the ratio of peak spacing, ΔT , to average peak width, W , of two adjacent peaks with one base difference.

$$R = \Delta T / W \quad (3.2)$$

A resolution of 0.5 is adequate for DNA sequence identification. This level of resolution marks the limit for convenient sequence determination. The resolution become poorer for the larger fragments and so does the accuracy of sequence determination [38].

In the following sections, different parameters that will affect the separation efficiency, resolution, read length, and separation speed will be studied and discussed.

3.1 Effect of polymer matrix

High separation efficiency is a characteristic feature of capillary gel electrophoresis. The separation of charged polymers that have low diffusion coefficients in gel filled capillary columns permits as high as 3×10^7 theoretical plates m^{-1} to be achieved [21]. Since the early sixties polyacrylamide has been the most widely used electrophoretic medium. A number of its advantages, such as optical clarity, high hydrophilicity, etc., led to its enormous popularity.

The gel polymerization reaction is a free-radical reaction, affected by temperature (exothermic reaction), oxygen (quencher), and the amount of initiator and catalyst present. Gel composition, gel polymerization conditions are critical to the gel properties and separation efficiency.

There are studies on the effect of gel polymerization conditions on the formation of cross-linked polyacrylamide [56, 57]. Fewer studies have been done for the non-crosslink gels. In this section, we will discuss the results of gel polymerization condition effect on separation of DNA sequencing fragments.

3.1.1 Effect of urea concentration

Peak overlap usually occurs late in the sequence as resolution degrades. However, peak overlap is occasionally observed as an artifact relatively early in the DNA sequence. This overlap is associated with formation of secondary structure within the DNA and is called a compression. For example, a compression is often noted near base 65 in the M13mp18 sequence. The sequence in this region reads ...GGTACC. A hairpin structure

can form, where GGT is paired with ACC. This compact structure migrates at a faster rate compared to migration in the absence of secondary structure. As a result, the fragments that terminate near this sequence will have smaller peak spacing than fragments migrating in normal regions. The electropherogram is said to suffer from a compression at this point in the sequence; sequencing accuracy suffers in a compression because resolution is degraded [8].

To eliminate secondary structure, the DNA sequencing sample is heated at 95°C for 2 min. before injection to help DNA molecules denature (denaturation is the loss of secondary structure, which is the disruption of the hydrogen bond that create the double-helix structure). In a capillary gel separation column, the denaturing capacity of the gel usually can be increased by the addition of urea. Urea acts as a chemical denaturant, which increases the solubility of hydrophobic compounds in water as well as hindering hydrogen bond formation in the aqueous phase. Urea can disrupt hydrogen bonds, reduce intramolecular base pairing of nucleic acids, and break up secondary structure [58].

Fig.3.1 demonstrates the effect of different urea concentrations on the separation of the reaction mixture of an A-terminated M13mp18 sample. In general, the separation patterns of the electropherogram were quite different for non-urea and urea-containing polymer matrices. Fig. 3.1 shows that the polymer matrices that contain urea produce more uniform electropherograms (both in peak height and peak spacing) and are able to separate much larger fragments compared with the matrices that do not contain urea. The peaks generated in the 0 M urea matrix tend to be piling up, which is the results of co-migration of different bases. A higher urea concentration gave better separation for adjacent peaks. The three peaks at base 100 are well resolved with the urea containing matrices while the 0 M urea matrix only shows one peak, this is shown in Fig. 3.1.1. The two peaks near base 130 are separated with the 5 M and 7 M urea gels (see Fig. 3.1.2). The urea containing matrices are especially helpful for the separation of larger fragments

(see Fig. 3.1.3). In contrast, the 0 M urea gel did not separate fragments larger than about 360 bases. It is clear that the addition of urea in the polymer matrix does help to release the compression within DNA to certain extent and improve the resolution.

However, urea alone in the gel cannot resolve some compressions that occur in GC rich regions. Increased temperature could help to eliminate compression. In addition to operating the capillary at high temperature, other denaturing reagents, such as formamide, can be added to the sequencing gel [55].

Ruiz-Martinez et al.[59] reported that better separation efficiency was achieved by using the modified denaturing agent of the running buffer from the typical 7 M urea to 3.5 M urea plus 30% (v/v) formamide. Their aim was to reduce the viscosity of the medium and to enhance the denaturing ability with the addition of formamide.

3.1.2 Effect of Polymerization temperature

The polymerization temperature effect for the linear polyacrylamide is studied and the results are shown in Fig. 3.2. Figure 3.2.1a and Figure 3.2.1b are the plots of theoretical plate number and resolution vs. base size for separation matrix prepared at room temperature and 0°C, respectively. Superior resolution and plate number are observed with noncrosslinked polyacrylamide prepared at room temperature. Yet, more studies need to be done for other temperature settings. The polymerization reaction is carried out thereafter at room temperature unless otherwise indicated.

3.1.3 Degassed polymer matrix

To study the influence of oxygen during the polymerization step, electropherograms were generated with polyacrylamide that had been degassed and had not been degassed

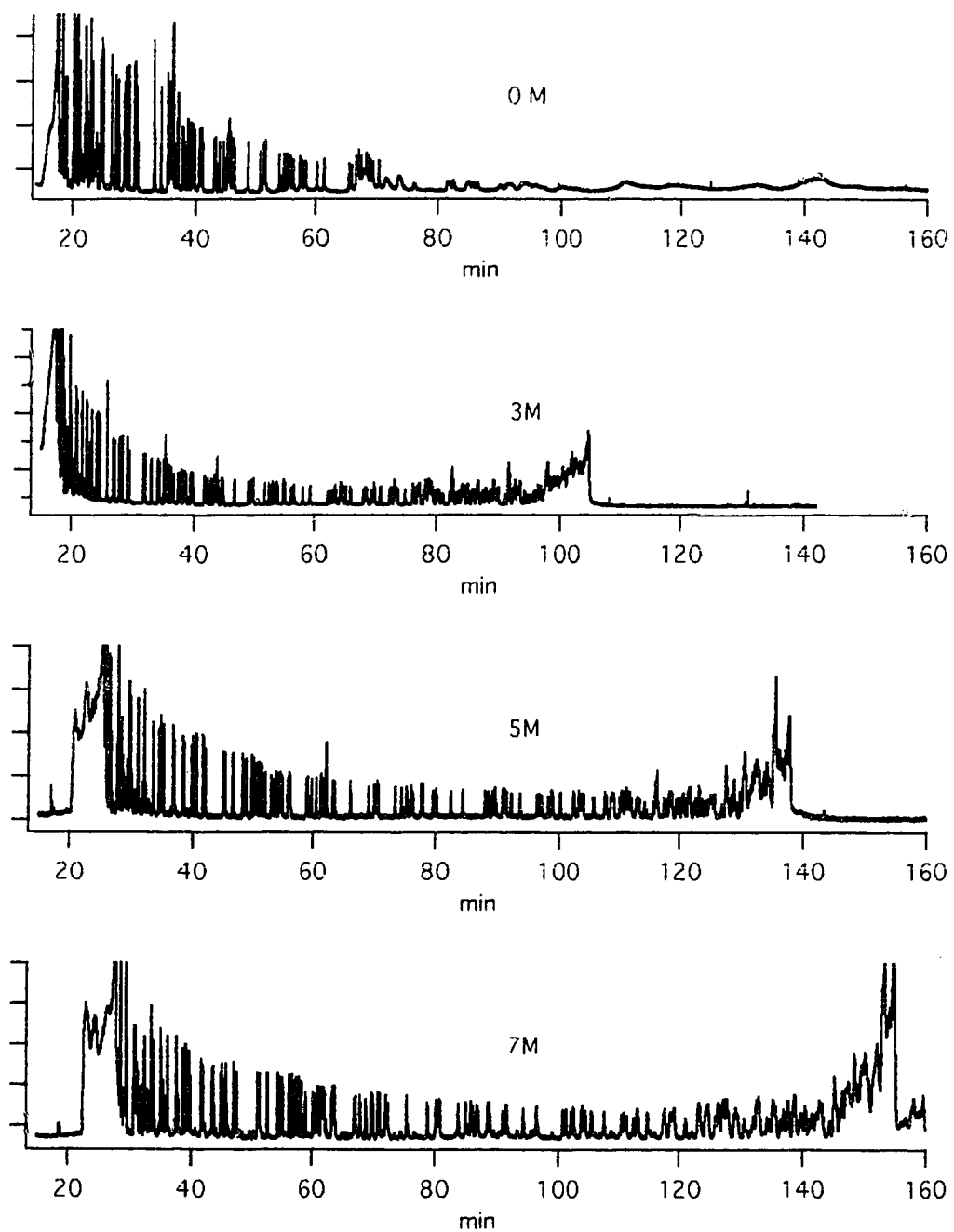


Figure 3.1 Effect of different urea concentration on the separation of the reaction mixture of an A-terminated M13mp18 sample

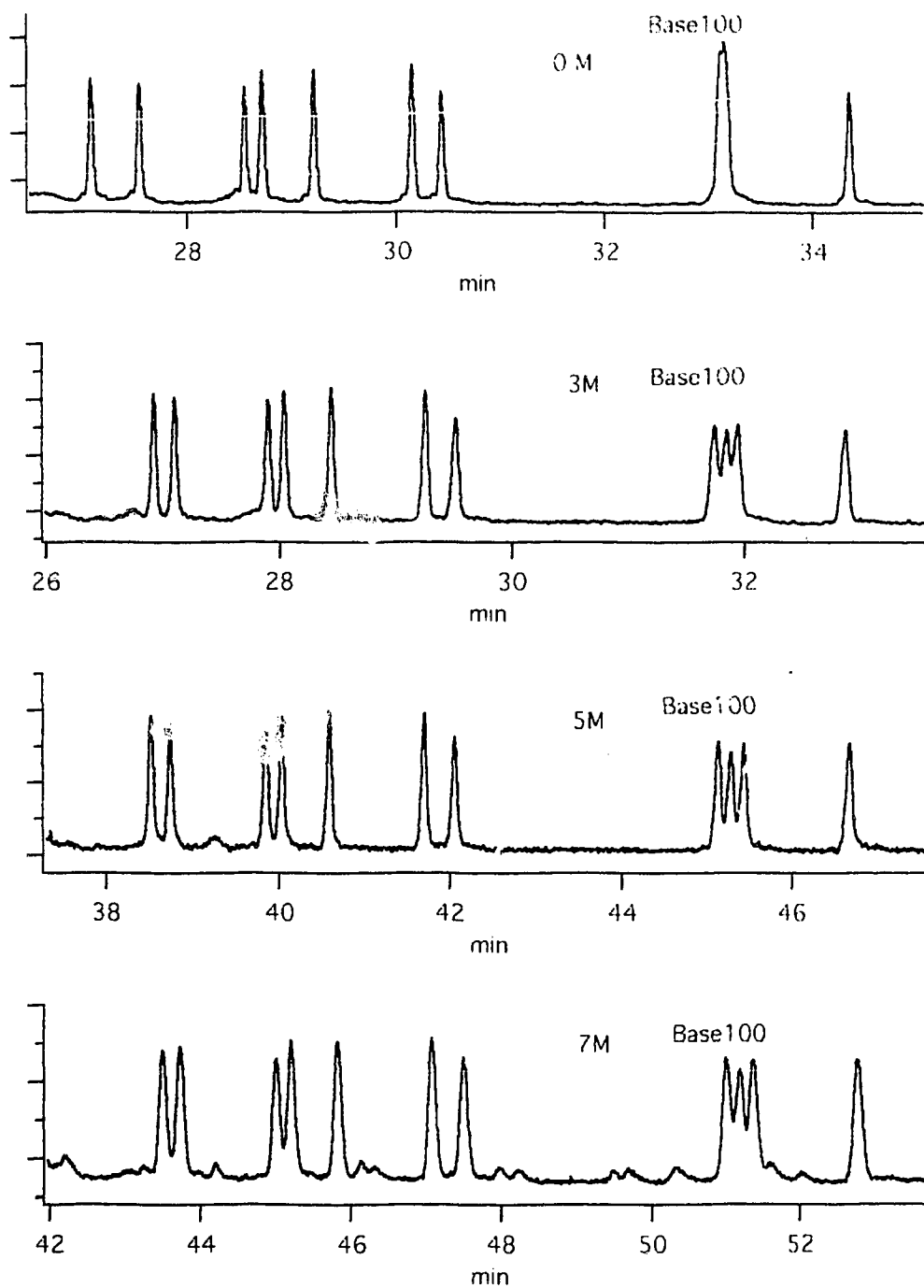


Figure 3.1.1 Effect of urea concentration . Zoomed view at about 100 bases

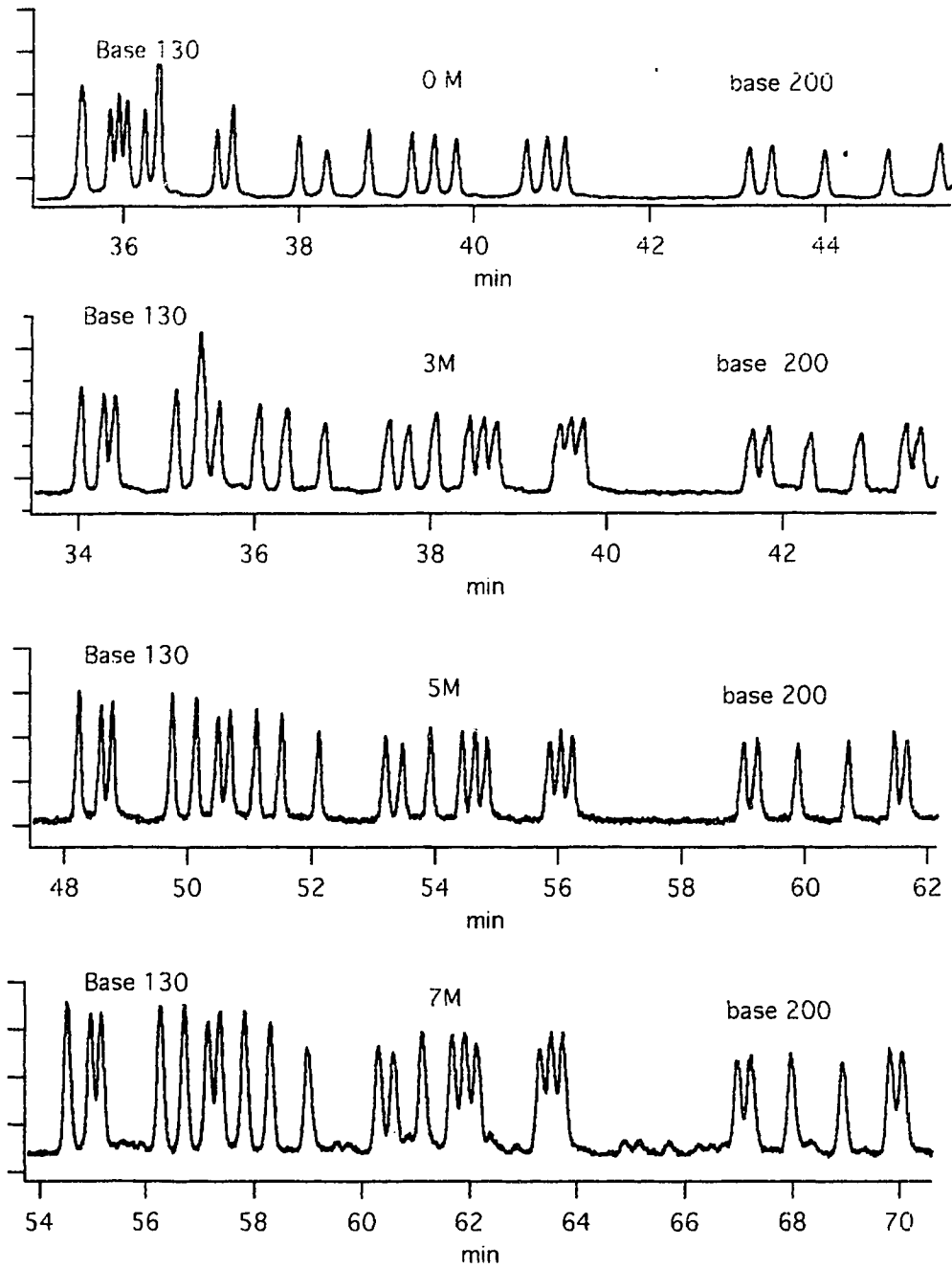


Figure 3.1.2 Effect of urea concentration. Zoomed view at 100 - 200 bases

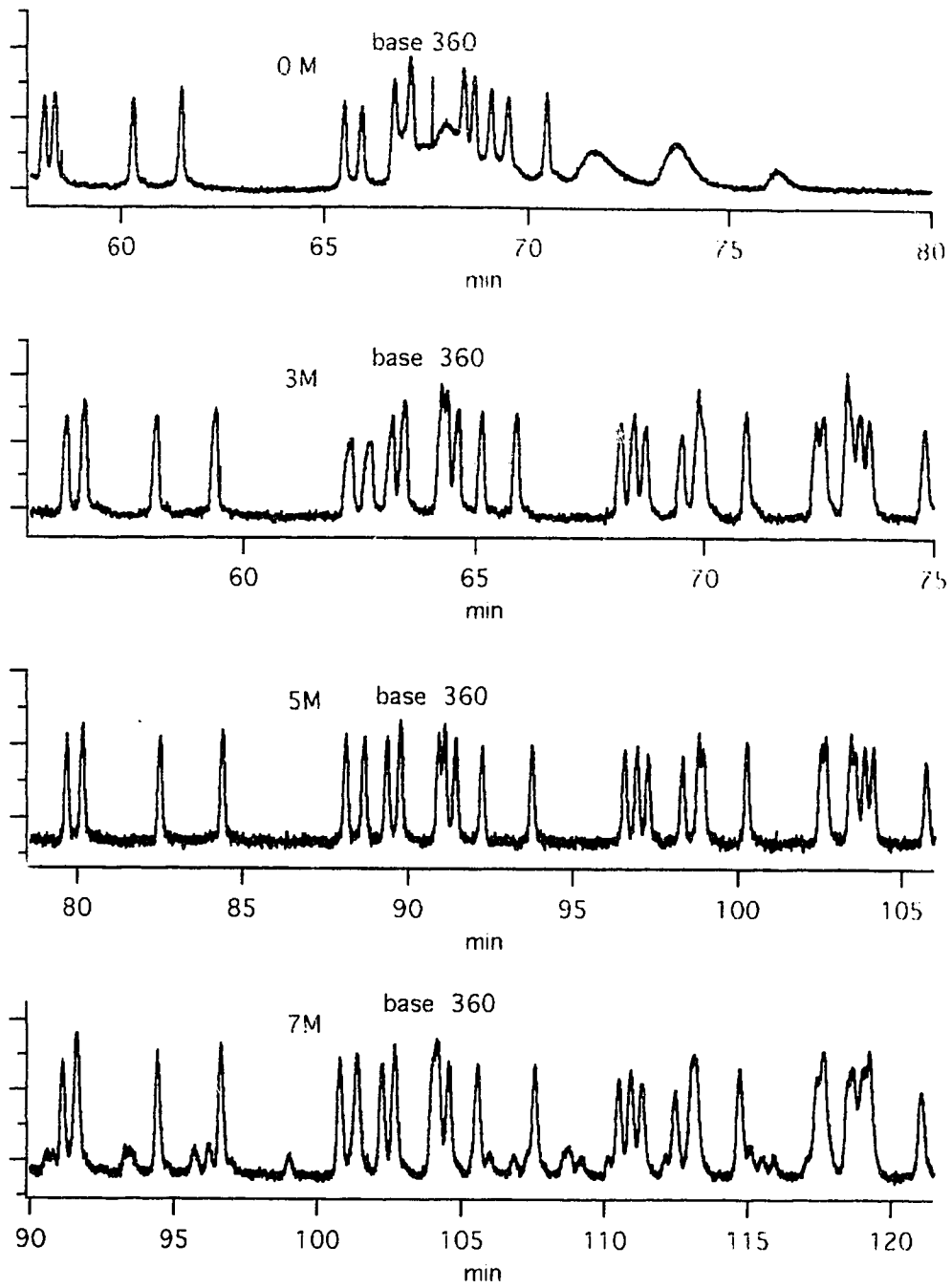


Figure 3.1.3 Effect of urea concentration. Zoomed view at >300 bases

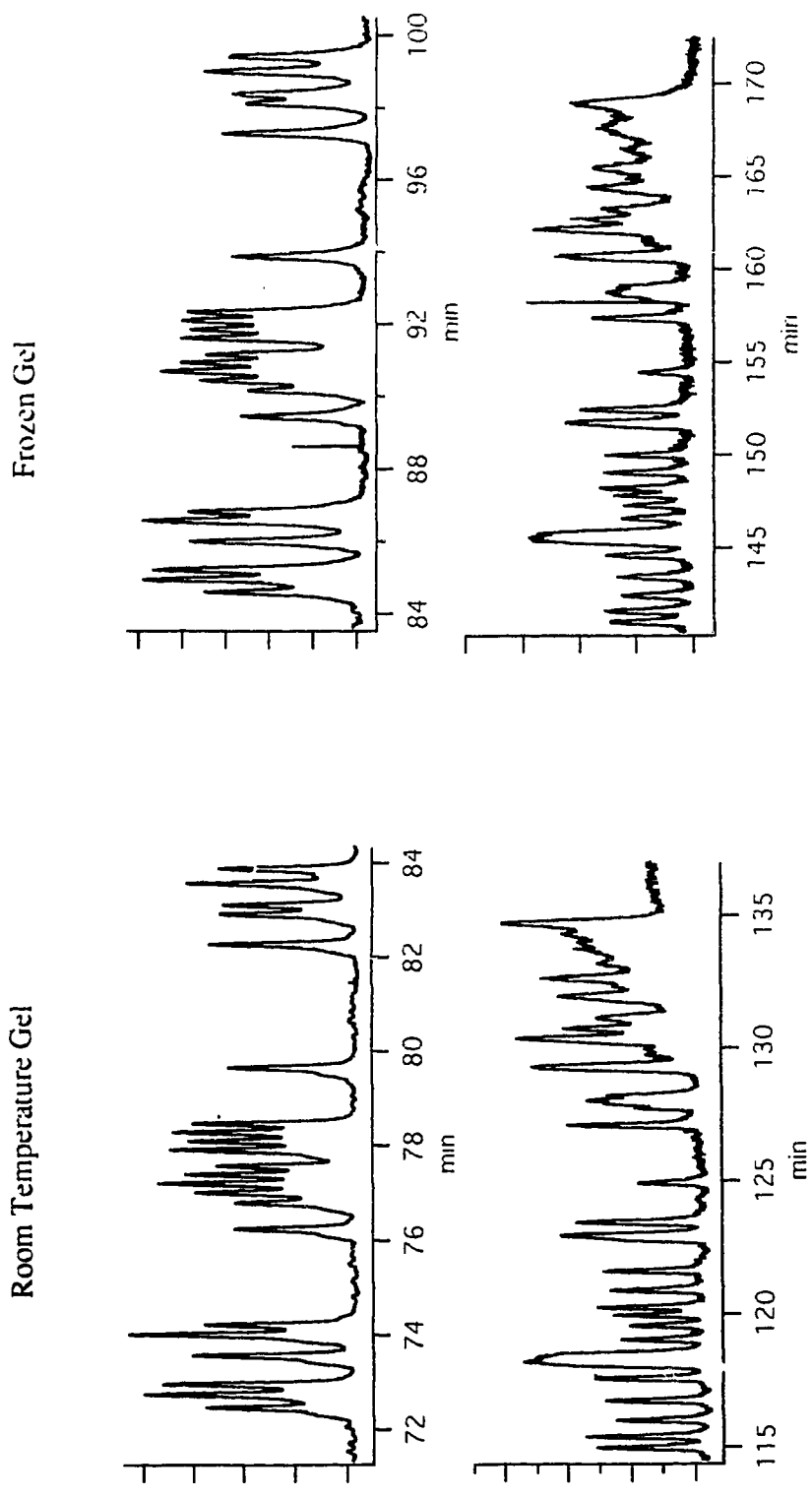


Figure 3.2 Comparison of separation with gels polymerized at different temperature

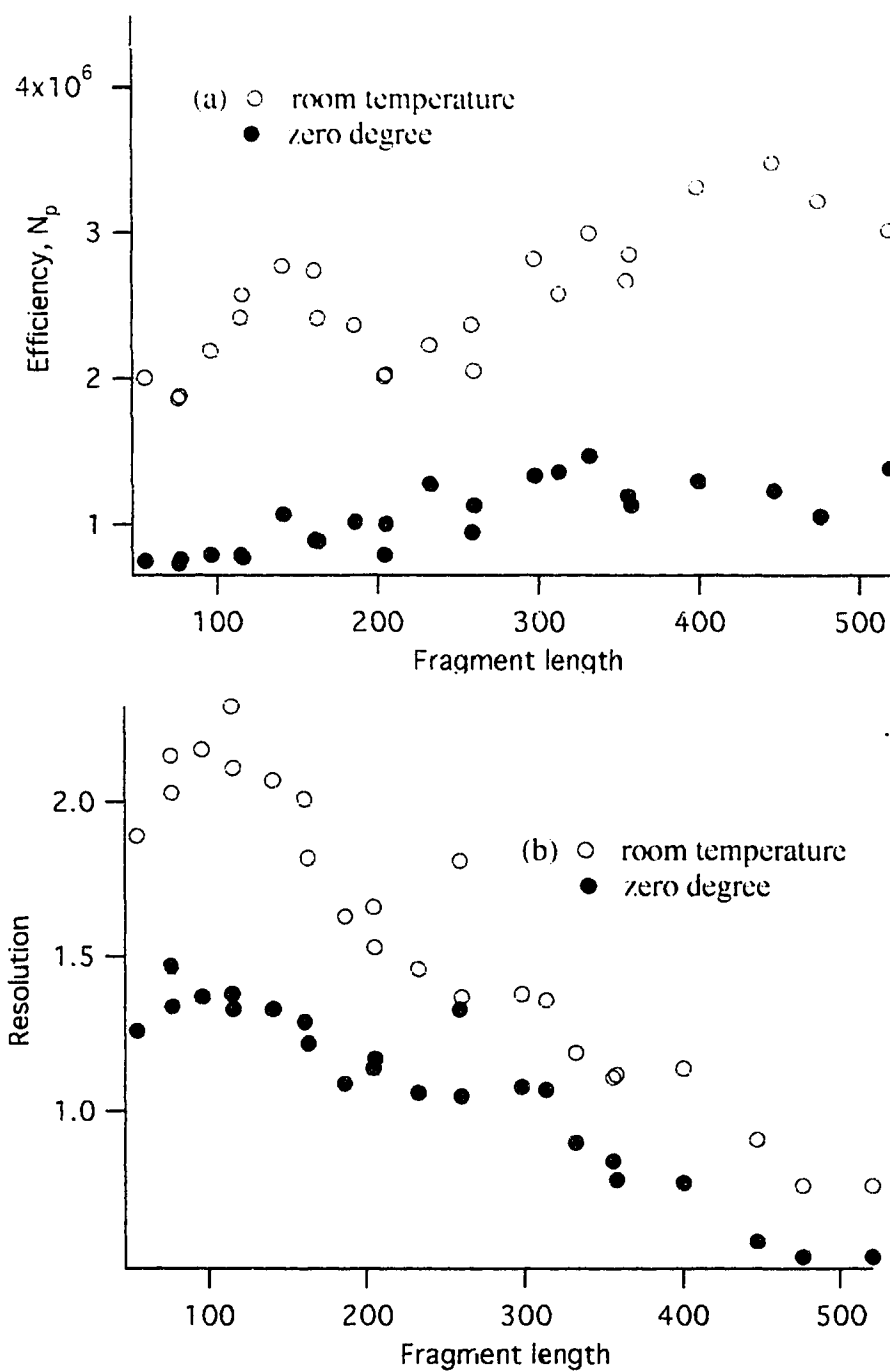


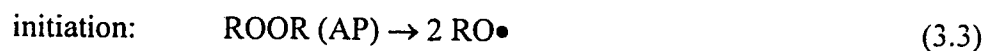
Figure 3.2.1 Effect of polymerization temperature on separation efficiency N_p (a) and on resolution R (b)

before polymerization. Figure. 3.3 compares the results obtained with gels that had been degassed before polymerization with non-degassed gels. Fig 3.3 (a) and (b) are the results obtained with non-degassed gels; (c) and (d) obtained with degassed gels.

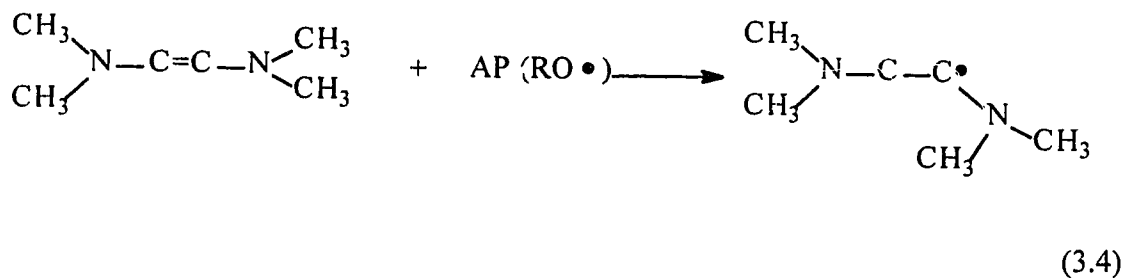
It is clear that the gels prepared with minimum oxygen provide much better separation efficiency and resolution. Obviously, oxygen removal is an important and essential step in the polymerization reaction. The reason for this is that the polymerization reaction is a free-radical reaction and the resulting polymer is formed by a chain-propagation process.

- presence of oxygen causes premature termination of the polymer chain, resulting in shorter polymer chains with poorer sieving properties.

The first step of the polymerization reaction is the activation of TEMED by ammonium persulfate (AP), which leaves the TEMED molecule with a reactive, unpaired electron.



Activation of TEMED:



The TEMED will then react with acrylamide, which is in turn activated. As the polyacrylamide chain grows, the active site shifts to the free end. Bisacrylamide, which consists of two acrylamide units joined through their $-\text{CONH}_2$ groups, can be incorporated into two growing chains. Hence the presence of bisacrylamide leads to the

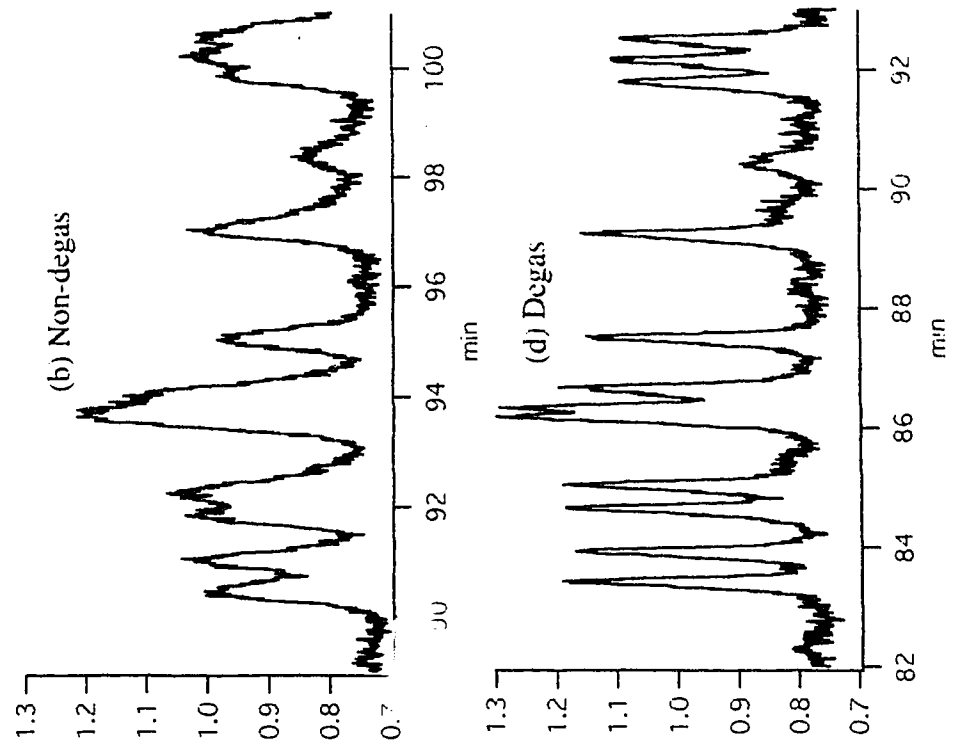
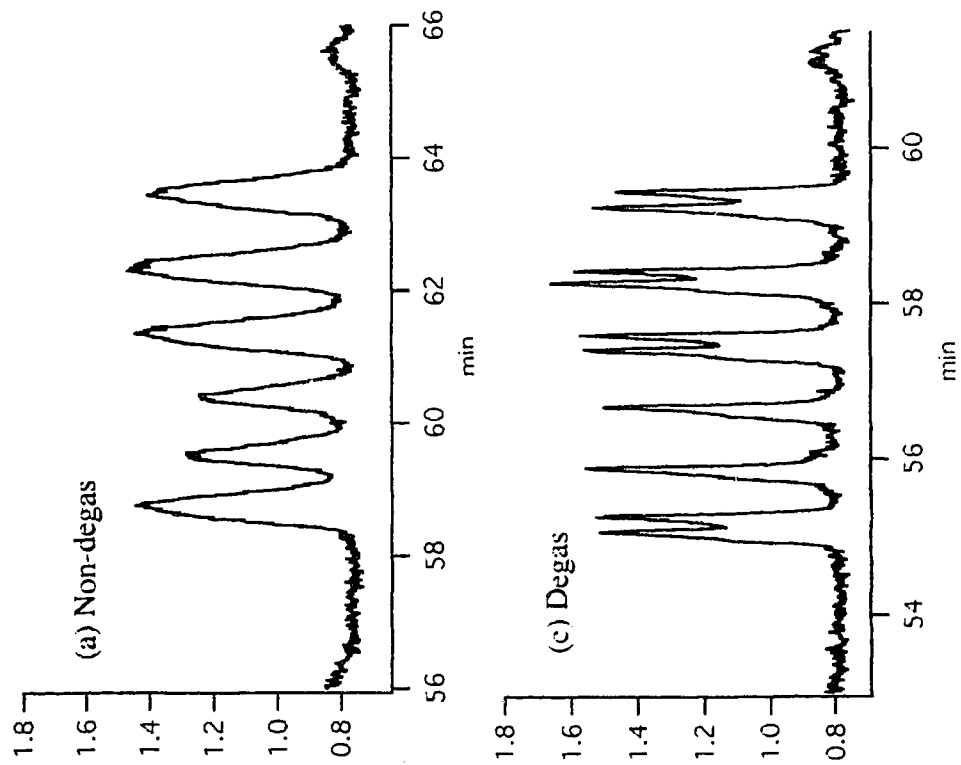


Figure 3.3 Comparison of separation for degassed and non-degassed polymer matrices

formation of cross-links between chains. With an abundance of cross-links the polymer of a gel has a topologically complex configuration, with loops, branches and interconnections. If the solution included only acrylamide monomers, the chain would always be straight, or unbranched [60].

The termination steps for polymerization are typical free-radical termination steps. The presence of oxygen (quencher) will quench the polymerization. The higher efficiency and resolution obtained with degassed gel (see Fig. 3.3) is probably due to the longer chains formed in polymerization reaction with the removal of the quencher.

3.1.4 Effect of total acrylamide concentration

Total acrylamide concentration is one of the most important factors affecting the gel performance and, in turn, affecting the separation efficiency. The weight percentage of monomer plus cross-linker (i.e. bisacrylamide) in the polymer is denoted %T. The mole fraction of cross-linker to monomer is denoted as % C. 0% C means that the gel is a non-crosslinked gel, also called a linear gel.

The effect of total acrylamide concentration in the absence of cross-linker was studied by using this five capillary instrument. Five different polymer solutions (4%T, 4.5%T, 5%T, 6%T and 7%T) were sucked into the five capillaries simultaneously by vacuum. Then, these different condition gel capillaries were run at the same time to make the results more comparable, since the experiments were done at almost identical conditions except the total acrylamide concentration. The experimental conditions are as follows: Sample: P40, running voltage: 150 V/cm; Capillary length $L = 38.5$ cm.

Fig. 3.4 shows that increasing the total acrylamide concentration drastically slowed the sequencing rate. The primer elution time also increases slightly as the total monomer concentration increases.

Fig. 3.4.1(a) presents a comparison of the migration time versus DNA fragment length for the different gel compositions. It is obvious that higher %T gels require longer separation time. With higher %T gels (7%) the onset of reptation occurs for shorter fragments and is more obvious, as indicated by the decreased slope of the curve for long fragments.

Fig. 3.4.1(b) is a plot of the mobility versus the fragment length. The mobility plot flattens-out for the longest fragments, which implies that the longest fragments are difficult to separate. This asymptotic limit of mobility seems most severe for the highest concentration gels. Higher concentration gels have smaller pore sizes, and biased reptation occurs for smaller fragment length.

The gel behavior can be described by a Ferguson plot, which is a plot of log mobility vs. %T. Fig. 3.4.1(d) shows this relationship for fragment with the different lengths: 24, 95, 182, 292, 405, 506 bases. However, a linear relationship seems to only apply for smaller DNA fragments. The Ferguson plots become more curved for larger DNA sequencing fragments at higher %T. This result is in good agreement with the analysis performed by Tietz and Chrambach [61]. This curvature indicates that larger DNA fragments undergo biased reptation.

The most important criteria for sequencing DNA samples are resolution, speed, and read length. In order to select optimum total acrylamide concentration for the DNA sequencing, a plot of resolution vs. fragment length is obtained based on the electropherogram data of Fig. 3.4 and the result is shown in Fig. 3.4.1(c). It is interesting to note that for the different region of the electropherogram, the total acrylamide concentration effect on the separation efficiency are quite different. The results are

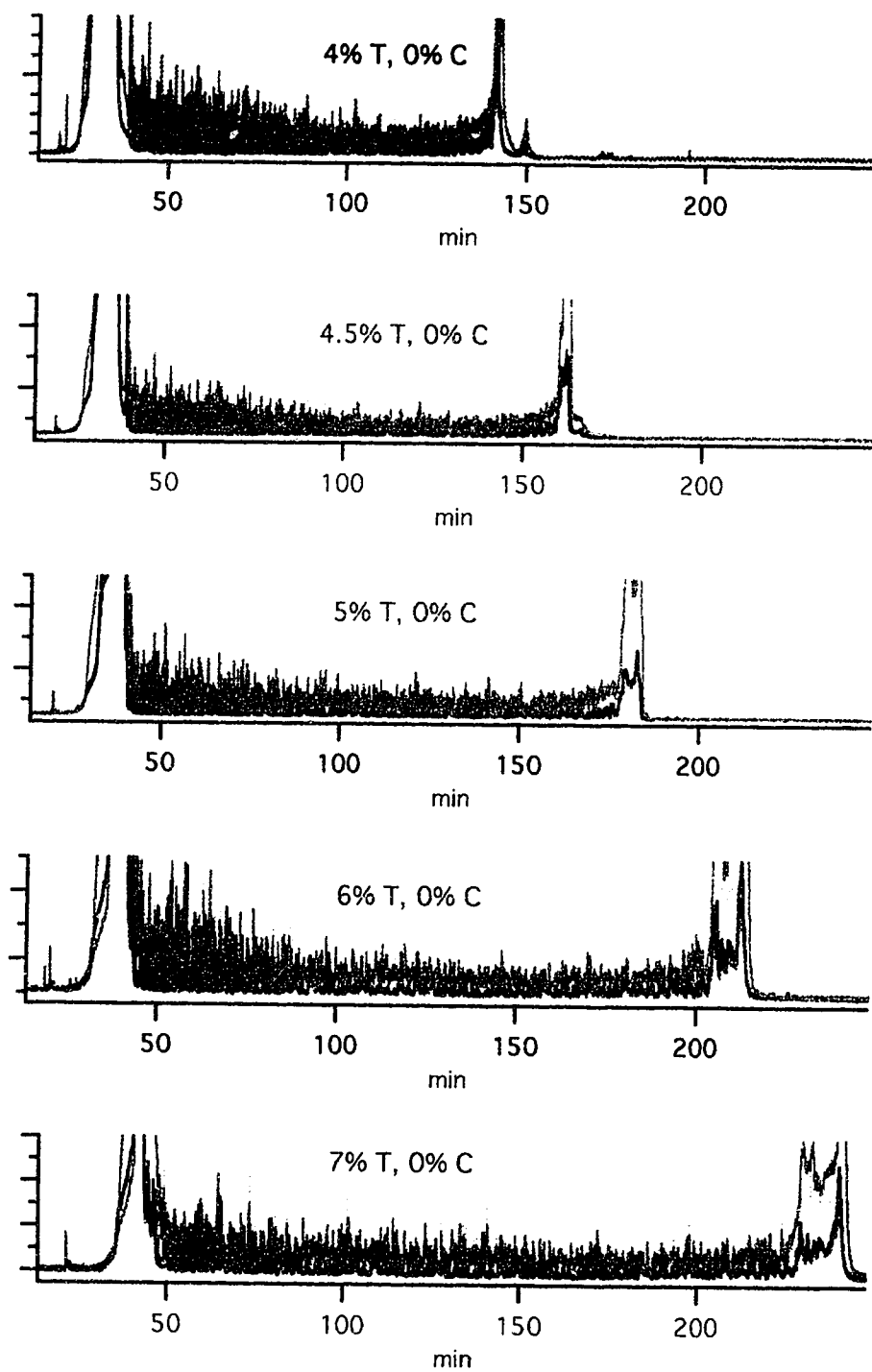


Figure 3.4 Effect of gel composition on the separation of four color labeled DNA fragments

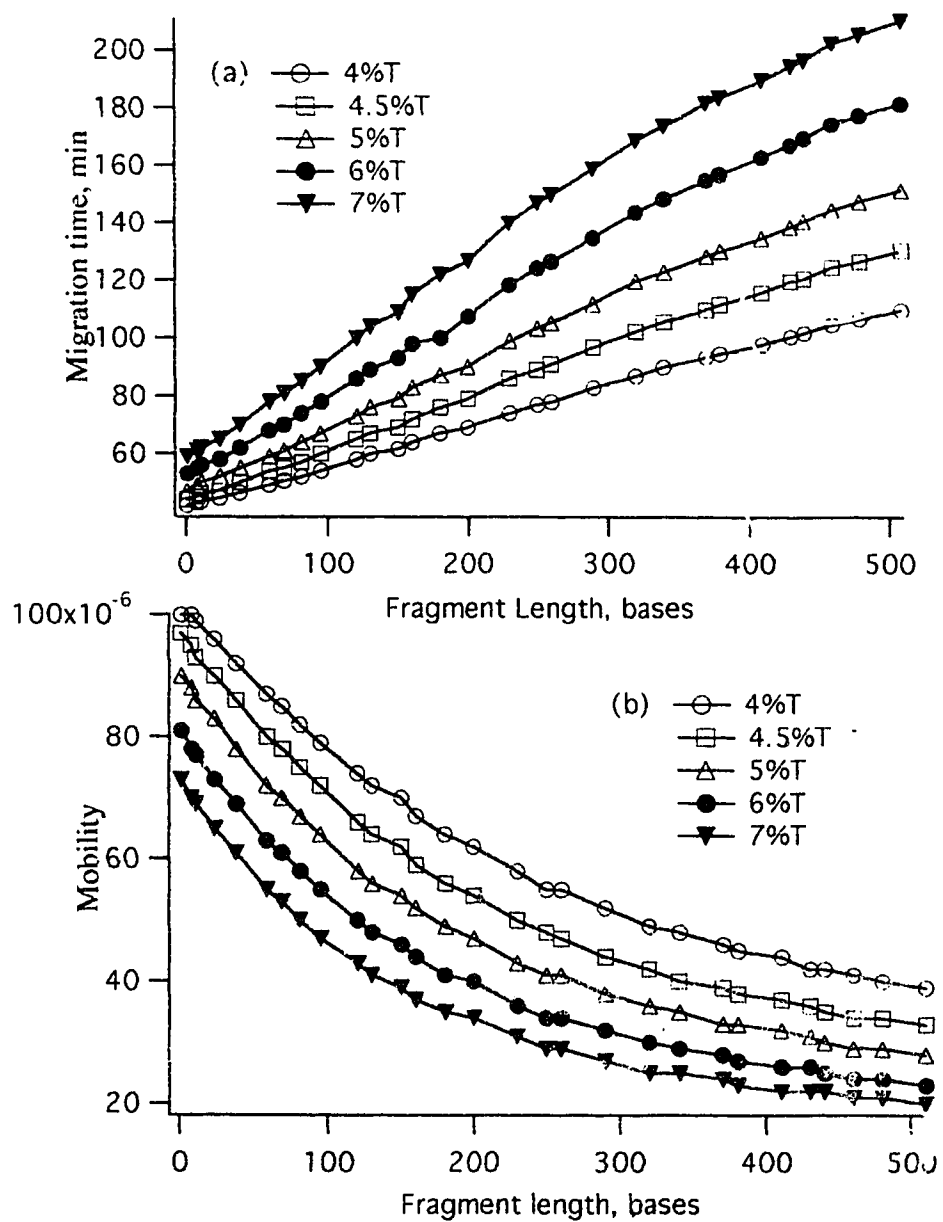


Figure 3.4.1 Effect of gel composition. (a) Plot of migration time vs. fragment length; (b) Plot of mobility vs. fragment length.

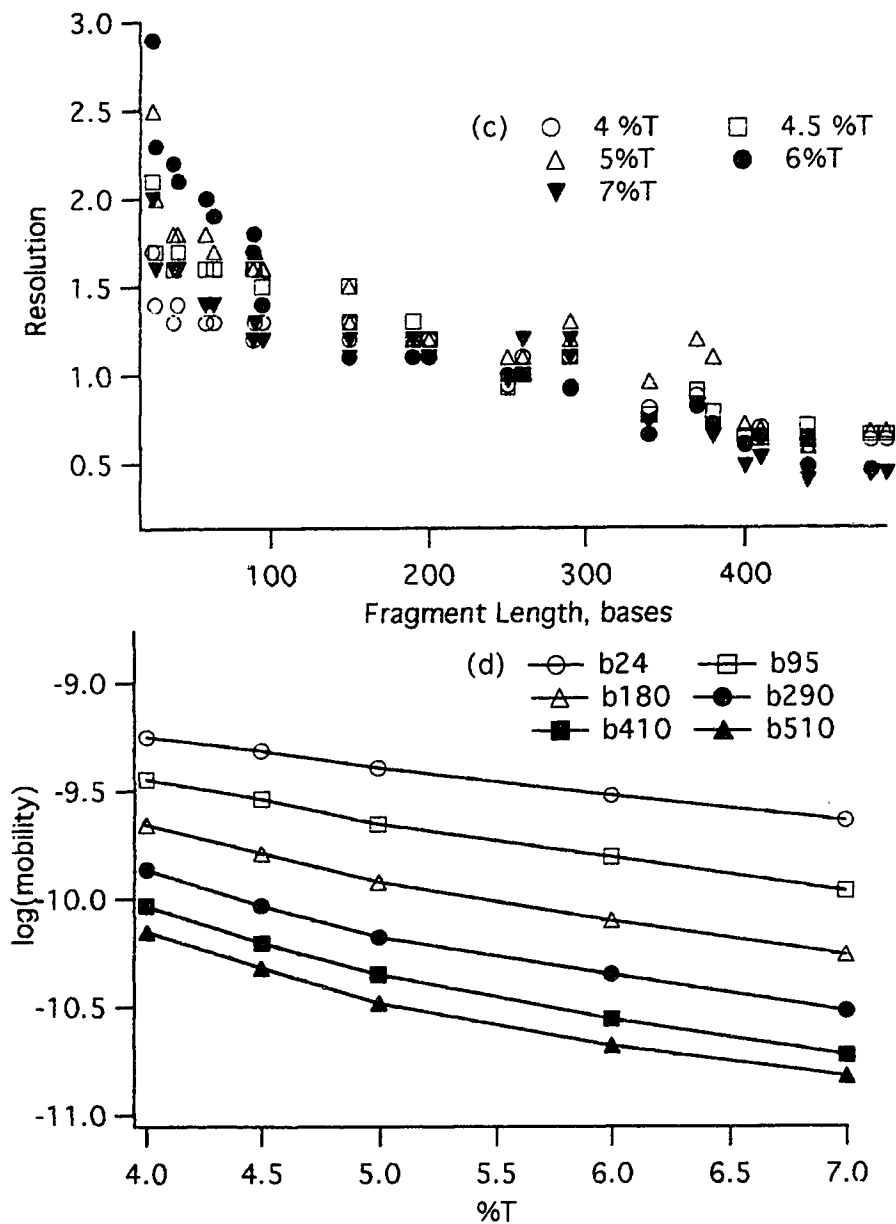


Figure 3.4.1 Effect of gel composition. (c) Plot of resolution R vs. fragment length for different gel composition; (d) Plot of $\log(\text{mobility})$ vs. %T for different fragment length.

summarized below:

- 1) Low monomer concentration gels provide better resolution for large fragments. Low %T polyacrylamide linear gels generate more sequence information than the higher %T polyacrylamide. From Fig. 3.4.1(c) we can see that the 4% T, 4.5% T, and 5% T gels still can maintain reasonable resolution after 400 bases whereas the 6% and 7% gels provide poorer resolution for those same fragments.
- 2) Higher concentration gels are better for separation of smaller fragments.
- 3) The 5% T gel provides the optimum separation for all the regions.
- 4) In general, the resolution decreases with the fragment length. However, the resolution change with the fragment length is relatively small for bases 200-300 , Fig. 3.4.1(c).

3.2 Effect of Electric field

Electric field strength has a profound effect on the separation efficiency and separation time for DNA sequencing by CGE. Different electric field strengths were studied in order to find the appropriate electric field for the separation of a mixture of DNA molecules with different chain lengths in a given polymer matrix.

With higher electric field strengths, DNA molecules can be separated in a relatively short time, speeding up the separation process. The high surface-to-volume ratio of the capillary allows efficient removal of Joule heat from the capillary, which allows use of high electric fields for rapid DNA separations. The highest electric field reported in the literature for DNA separations is 800 Vcm^{-1} , in which ~ 250 bases of sequence was determined in about seven minutes [62]. Unfortunately, it appears difficult to separate large fragments at high electric fields; the large fragments tend to co-migrate. At low field strength, good separation can be achieved even for large fragments [63]. This

separation of the larger fragments is attained at a cost of longer separation time. The increase in sequence information at lower electric field is associated with the delayed onset of biased reptation [63]. Diffusional band broadening due to the long separation time may affect the resolution of the smaller fragments.

As mentioned in chapter 1, a number of theories have been developed to model the migration behavior of DNA in electrophoresis [19, 39-47]. Among these models, the biased reptation model, Equation 1.2, can describe the mobility of DNA fragments at high electric field. The value of β in Equation 1.2 is quite important in practical applications. A high value of β will severely limit the sequence read length at high electric fields.

In this section, we discuss the electric field behavior of DNA sequencing fragments for non-crosslinked polymers, especially at high electric fields. Five different electric field strengths were used for this experiment: 100 V/cm, 150 V/cm, 200 V/cm, 250 V/cm and 300 V/cm. The sample used to do these experiments was a T-terminated M13mp18 sample with a ROX labeled primer. The capillary length was 35.4 cm and the sample was injected at 150 V/cm for 60 sec. The results show that the electric field strength has a significant effect on the separation of DNA fragments.

Fig. 3.5 presents separations of single base terminated sequencing fragments at electric fields ranging from 100 to 300 V cm⁻¹. While short fragments are base-line resolved, larger fragments tend to co-migrate. There is a large, narrow peak that terminates the electropherogram. This peak appears to be due to the asymptotic migration of the larger fragments in the sample. The terminal peak occurs earlier at higher electric fields; the peak ~~also~~ appears to grow in amplitude at higher fields as more of the large fragments are included within it, since the sieving matrix can no longer separate some of the aligned molecules [35]. The separation speed increases and the read length decreases at higher

electric fields. There is a noticeable loss of read length at 300 Vcm^{-1} compared with data obtained at lower electric fields. For example, note the loss of resolution for the pentet T / quartet T at bases 316-320 and 321-325 in Fig 3.5.1. This loss of resolution at higher field strength undoubtedly limited the read-length of the earliest sequencing data published using non-crosslinked polyacrylamide [6, 59, 62].

Recall that Ogsten model, $\mu = \mu_0 e^{-k_c \lambda N}$, predicts the plot of log mobility ($\log \mu$) vs fragment length (N) being linear. Fig. 3.5.2(a) presents a plot of log mobility versus fragment length, N . The plot deviates from linearity for larger fragment lengths at all electric fields. It appears that Ogsten behavior is observed for only very short fragments with this concentration of noncrosslinked polymer. The mobility at 300 Vcm^{-1} begins to approach a saturating value for the largest fragments.

The biased reptation model, $\mu = \chi (1/N + (\alpha E/T)^\beta) = \chi (1/N + 1/N^*)$, can interpret the behavior of larger fragment better. The mobility data generated at electric fields ranging from 100 to 300 Vcm^{-1} were analyzed to determine N^* , the onset of biased reptation with orientation; mobility was plotted against inverse fragment length ($1/N$). The plots were linear for intermediate length fragments (see Fig. 3.5.2(b)).

The value of N^* was estimated from the ratio of the slope to intercept of the linear fit. Fig. 3.5.2 (c) presents a log-log plot of N^* versus electric field. The straight line is a weighted least squares fit of a line to the data; the slope of the line is -1.27 ± 0.02 .

The slope of the log-log plot is equal to β , the exponent in the biased reptation equation. A value of -1 for β implies that the onset of biased reptation is nearly inversely proportional to electric field. Fig. 3.5.2 (d) presents a plot of $1/N^*$ versus the electric field. The plot is linear with a near-zero intercept. Biased reptation with orientation should not be important for practical DNA sequencing at 100 Vcm^{-1} . Ogsten sieving is not

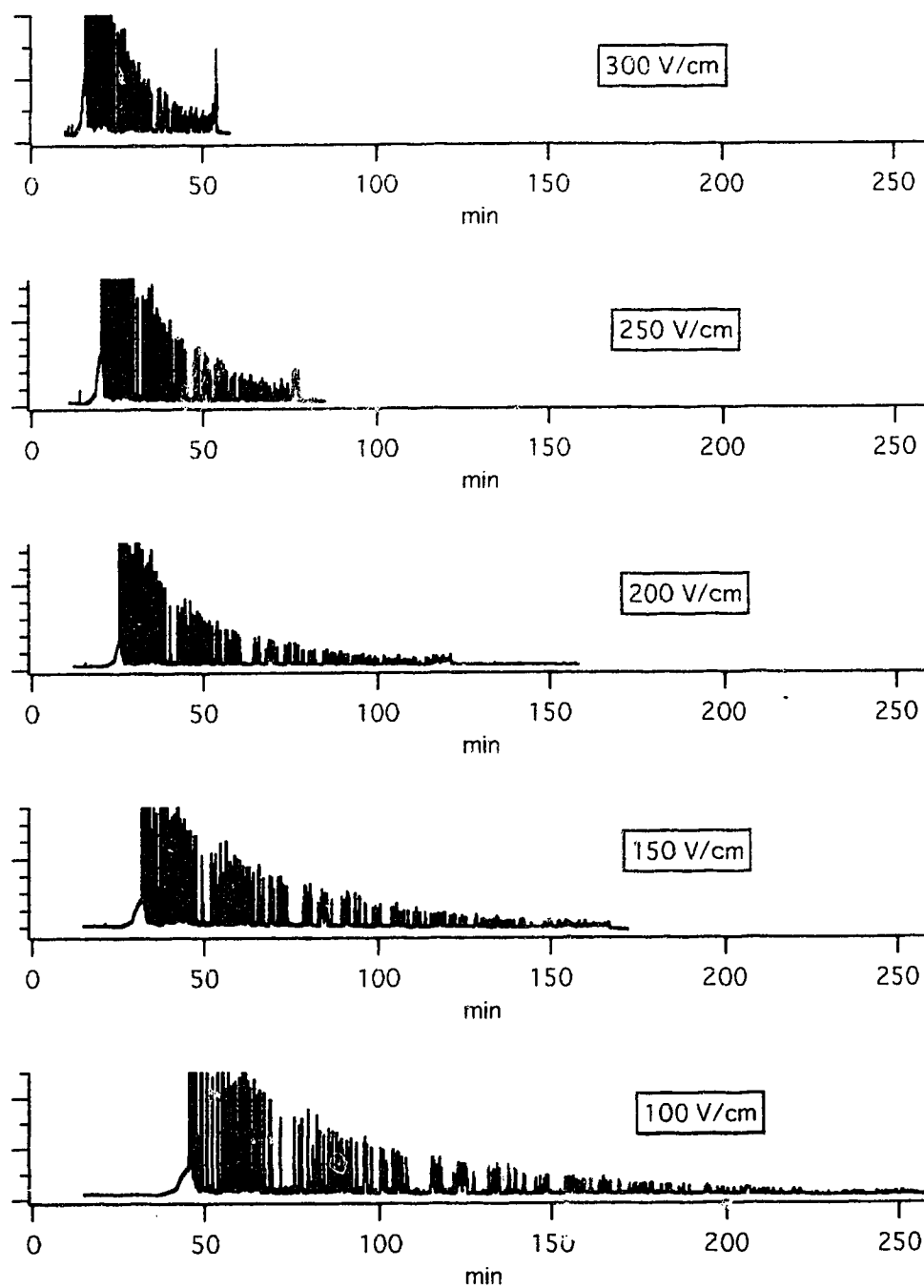


Figure 3.5 Separation of single base terminated sequencing fragments at electric fields ranging from 100 to 300 V/cm

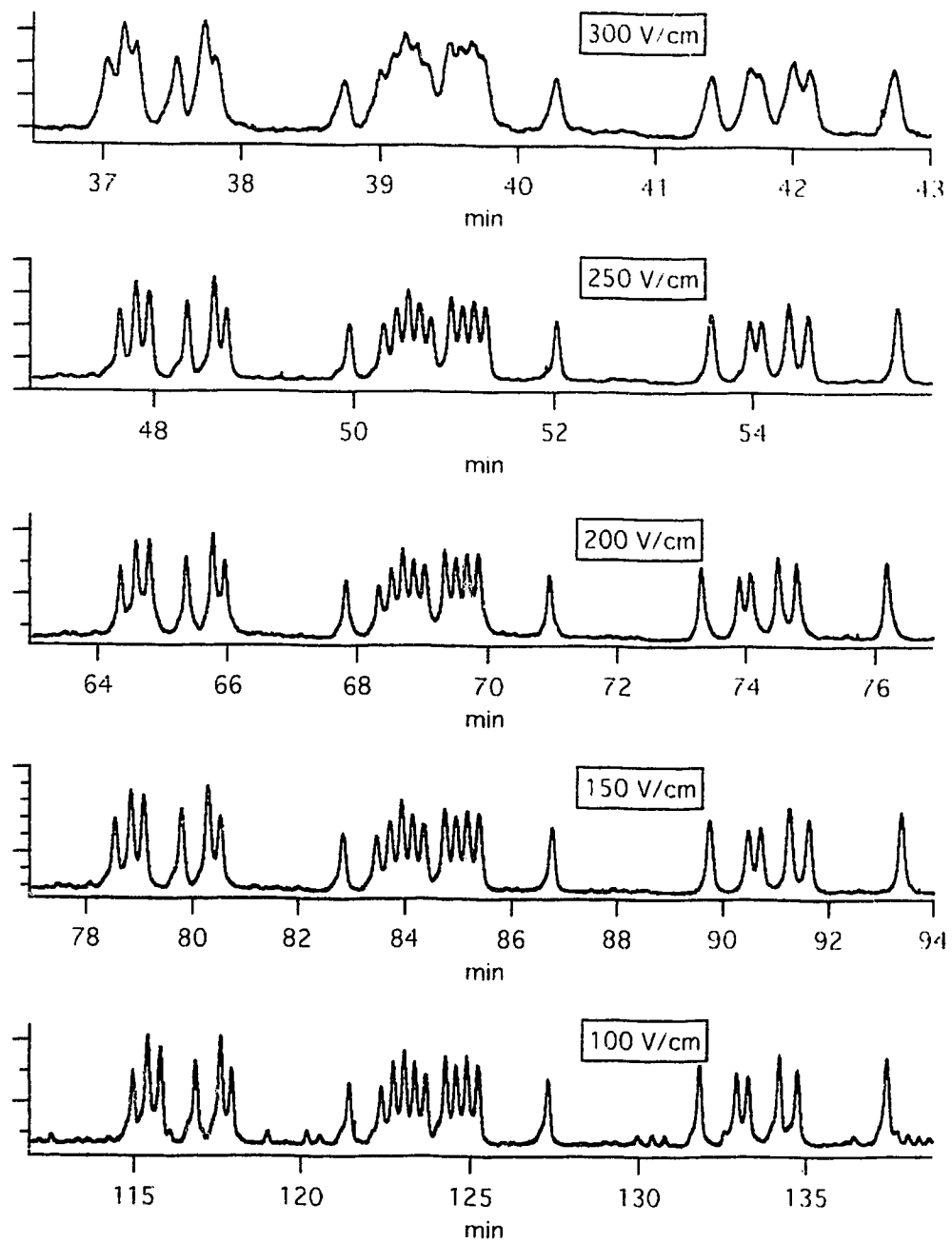


Figure 3.5.1 Zoomed view of fig.3.5 at base pairs around 320.

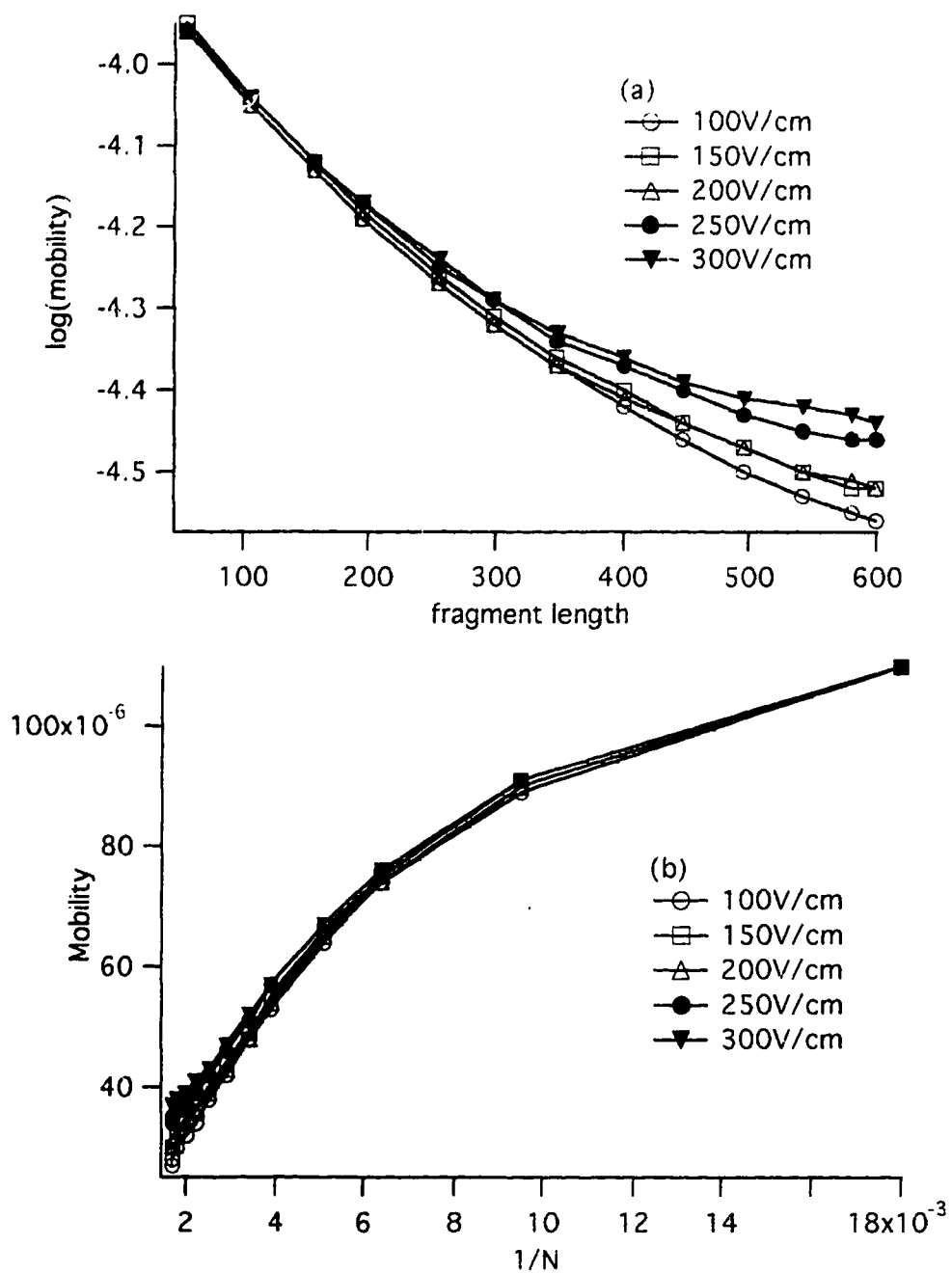


Figure 3.5.2 Effect of electric field. (a) log mobility vs. fragment length
(b) Mobility vs. $1/N$.

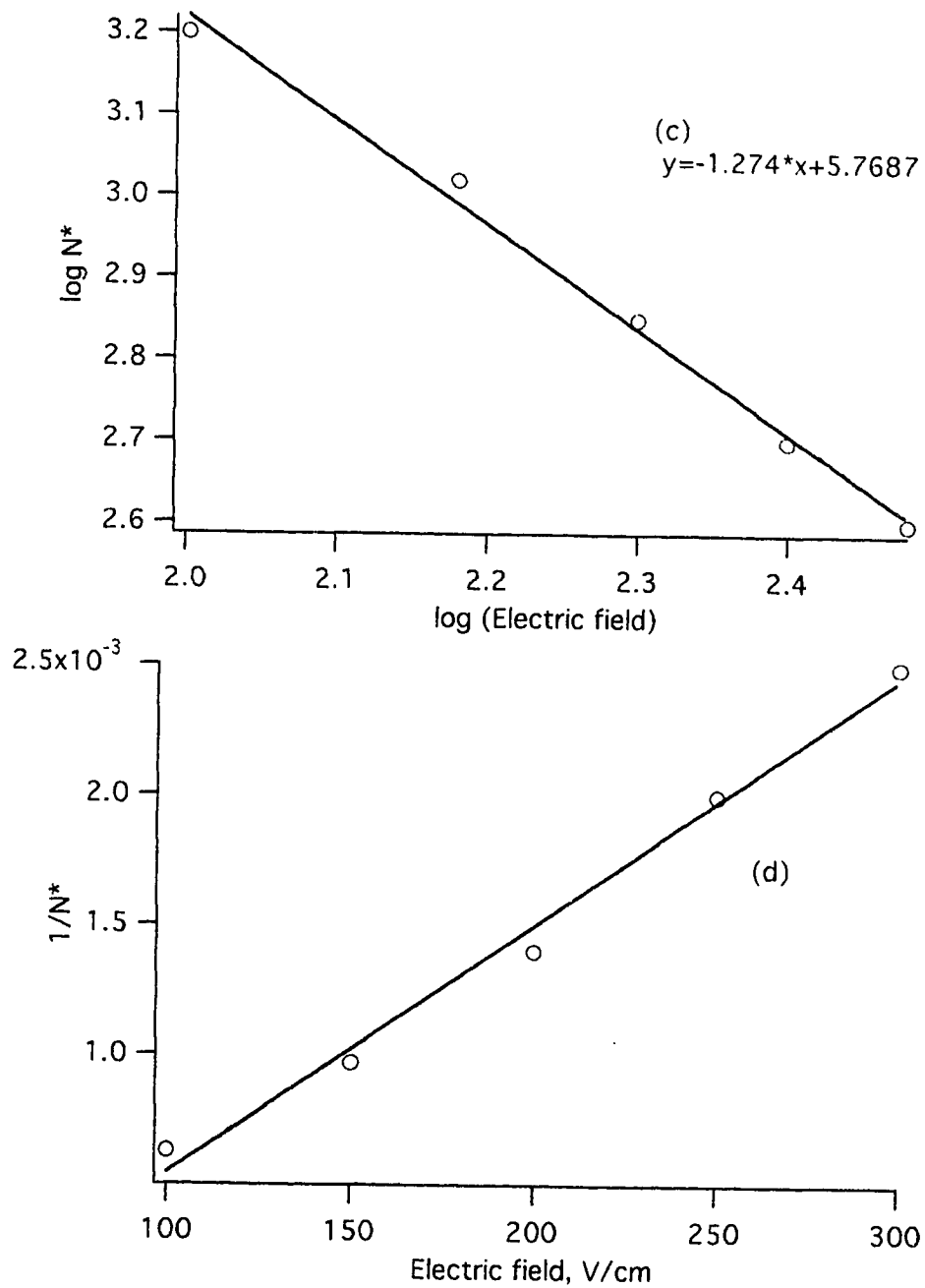


Figure 3.5.2 Effect of electric field. (c) log-log plot of N^* vs. electric field E ; (d) $1/N^*$ vs. electric field E

important for fragments larger than ~ 100 bases, even at the lowest electric fields employed for the separation [64].

The DNA chain becomes more oriented with increasing field strength, particularly for the larger size molecules, because the field biases the direction of the leading end of the molecule, which leads to an increase in mobility. In other words, by application of a high electric field, the longer chain length DNA molecules might be partially or completely stretched and aligned with the field. Thus the electrophoretic mobility of these big molecules becomes size-independent and field-dependent, causing poor separation at high field strengths.

Fig. 3.5.3 shows that separation efficiency (plate number, N_p) changes with the applied electric field for different fragment length N . Large fragments have much higher efficiency at low electric fields. The efficiencies of smaller fragments are less affected by the electric field. The effect of the electric field on the resolution shows a similar trend. It seems that 200 V/cm is an optimum value for all the fragment length in terms of efficiency and separation speed. The electric field can be changed during a single sequencing run to achieve optimum efficiency and separation speed. For example 250 V/cm - 300 V/cm can be used for the early stage of the sequencing (< 200 bp); then followed by 200 V/cm for separation of intermediate size sequencing fragments; and finally 150 V/cm can be used for sequencing large fragments.

It is interesting to see from Fig. 3.5.3 that the separation efficiency drops very quickly, especially for larger fragments, when the electric field is higher than 250 V/cm. The reason is not yet clear. One of the explanations is that Joule heating [65] generates a temperature gradient (a parabolic temperature profile in the capillary) that limits the maximum electric field strength. At high electric field, thermal gradient across the

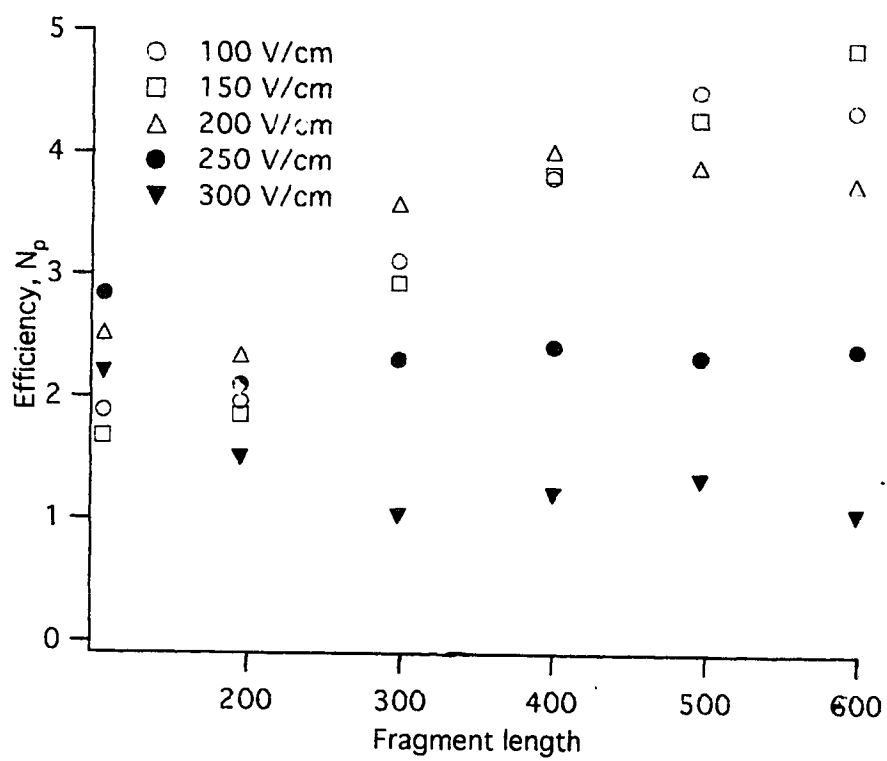


Figure 3.5.3 Separation efficiency vs. fragment length for different electric field

capillary axis leads to decreased efficiency [63]. If heating is significant, then, the efficiency change with the electric field for different fragment length should exhibit a similar pattern. However, in Fig. 3.5.3, the efficiency of a small fragment, such as 105 bp, increases with the electric field, then decreases slightly above 250V/cm. I think the reptation theory, as discussed in the beginning of this section, can explain the phenomenon better. At higher electric fields, co-migration due to the reptation occurs earlier in the separation, which limits the use of very high electric field. We should expect to see that the high electric field would have less effect on the efficiency for low concentration gels if the reptation is the limiting factor. However, more experiments need to be done in the future to confirm this assumption.

3.3 Effect of Capillary length

Another important parameter is the dependence of resolution upon column length [66]. In this part of the thesis, the effect of capillary length on separation is studied. Fig. 3.6. Fig. 3.6.1, and Figure 3.6.2 present the separation of single base terminated sequencing fragments with capillary lengths ranging from 35 to 60 cm and the expanded views of separation at certain base pair regions. The migration time is related to the length of the separation column. Not surprisingly, longer capillaries generate better separations (the plate number N_p is proportional to the capillary length, L , according to the equation: $N_p = \mu V / (2D) = \mu EL / (2D)$) but at the expense of longer running time. Using the 60 cm capillary, the separation took almost twice as much time as compared to a separation using the 35.6 cm capillary (see Fig. 3.6). However, the 48 cm and the 60 cm capillary show much better separation for all fragment length (see Fig. 3.6.1 and Fig. 3.6.2). Fig. 3.6.3 shows the plots of N_p vs. fragment length N (a) and R vs. fragment length N (b) for three different capillary lengths. There is a significant increase in the efficiency

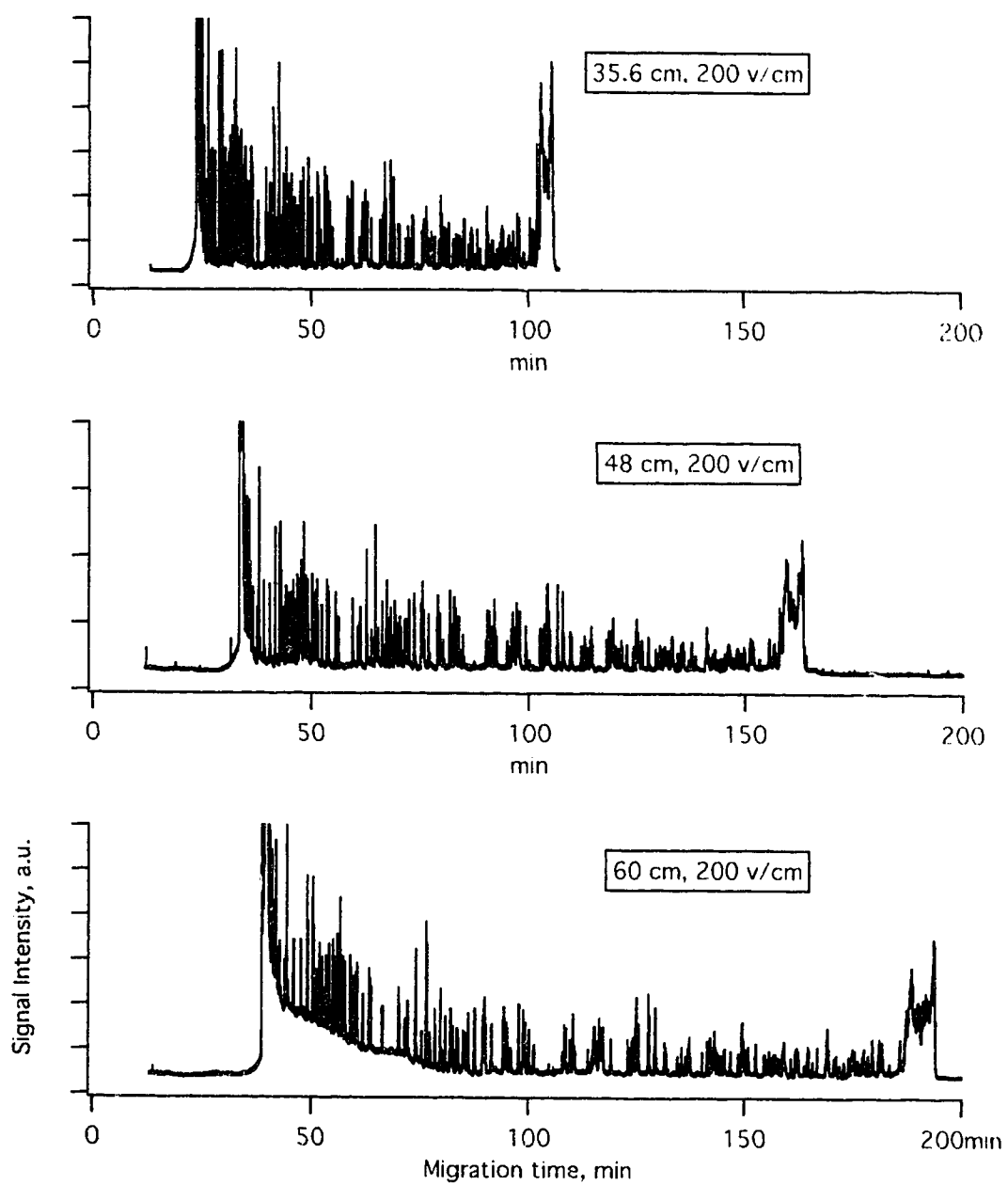


Figure 3.6 Separation of single base terminated sequencing fragments with capillary lengths ranging from 35-61 cm

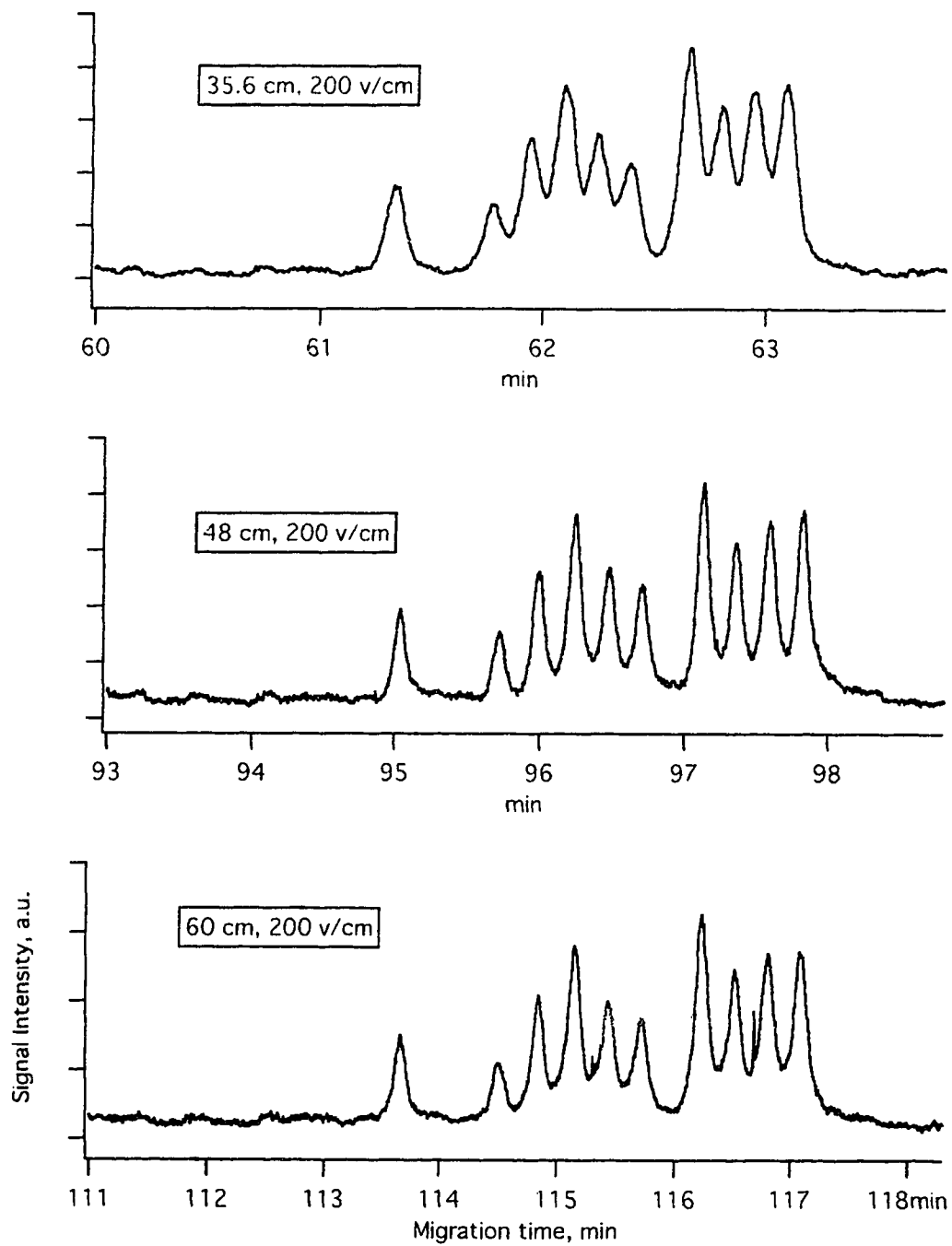


Figure 3.6.1 Zoomed view of Fig.3.6 at about 320 base pairs

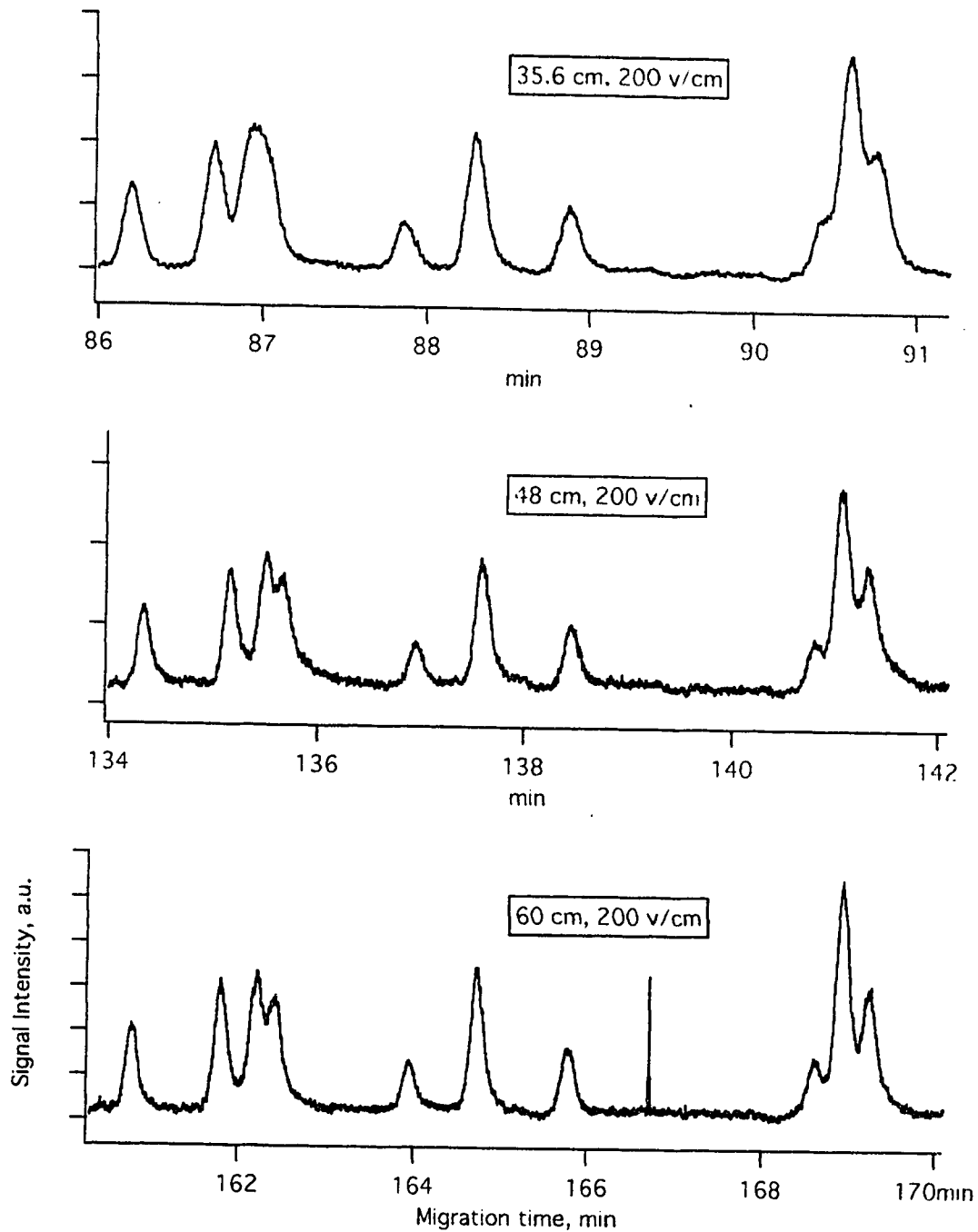


Figure 3.6.2 Zoomed view of Fig.3.6 at about 560 base pairs

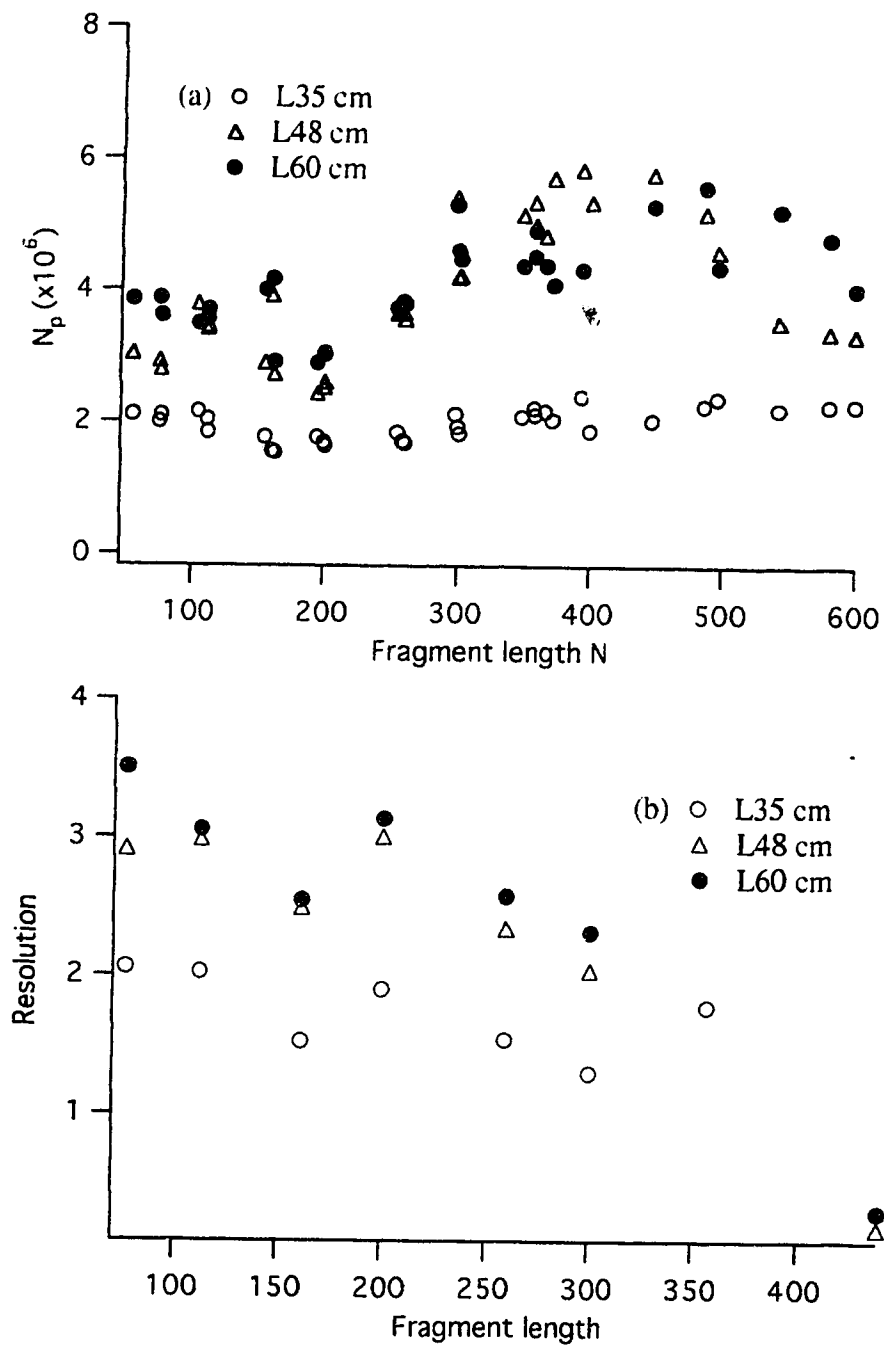


Figure 3.6.3 Effect of capillary length. (a) Plot of separation efficiency vs. fragment length. (b) Plot of resolution vs. fragment length

and resolution for the capillary length from 35.6 cm to 48 cm. However, a further increase in the capillary length shows no advantages in efficiency (Fig. 3.6.3 a) or resolution (Fig. 3.6.3 b). The results show that a 48 cm long separation column is adequate for separation of the sequencing fragments while a 60 cm column does not generate a big advantage but does require a longer separation time.

3.4 Sample injection related parameters

Sample injection parameters include (1) the minimum sample volume needed for a successful injection and (2) the optimum sample size required to generate a good separation. The sample size is determined by the sample concentration and injection time with a given injection voltage.

3.4.1 Minimum sample volume requirement

The usual DNA sample volume used for injection was 3-4 μl . Because of the cost of preparation of real DNA samples, the sample injection volume needs to be minimized as much as possible.

The injection current is used to monitor the injection. A significant drop of the current during injection indicates the failure of the injection, when very small sample volume are used (see Table 3.1). In this table, the injection current is monitored for several different sample volumes in both a single capillary and five capillary configuration. The current is about 1 μA per capillary when large sample volumes are used. As the sample volume decreases below 1 μL for the five capillary instrument, the current also drops. We believe this loss in current arises because most of the ions in the sample are loaded onto the capillary. The depletion of ions from the reservoir results in an increase in resistance and

Table 3.1 Minimum sample volume for adequate injection (100 V/cm).

Sample Volume (μl)	Injection Current I (μA)	
	Five Capillary	One Capillary
3.0	5.18	1.00
1.5	5.20	0.99
1.0	5.22	1.00
0.5	2.15	1.01
0.3	1.39	1.00 (2nd trial)
1 \times TBE (500)	5.26	1.00

a drop in current in the constant voltage injection.

The results show that 1 μl is the minimum volume for proper injection with the five capillary system. It is only 25% of the volume we used in our original experiments. This decrease in sample size allows us to split the products from a single reaction into several aliquots so that we can carry out more studies with one sample. By using aliquots from a single reaction, we eliminate one possible source of variance in these parameter studies.

An alternative injection procedure is to re-inject samples from the same vial in replicate experiments. However, a systematic decrease in signal with consecutive injections from the same sample vial was observed in our studies and have been reported by other researchers [32]. Therefore, it is advantageous to split the sample into multiple aliquots before any replicate studies are done.

3.4.2 Sample loading effect on the separation

The amount of sample that can be injected into the capillary is limited by the requirement for not causing significant zone broadening from sample introduction. The total (final) zone width is determined by [67]:

- 1) The width of the starting zone,
- 2) Zone broadening during the separation, caused by diffusion, Joule heating, conductivity difference between a solute zone and the surrounding buffer and the diffusion coefficient,
- 3) Zone broadening during detection due to the volume and time-constant of the detector.

The injection procedure may be the only extracolumn contribution that needs to be taken into account. The injection variance contributing to band dispersion has been

extensively studied in chromatography [68] and has also been adapted to capillary electrophoresis [69, 70]. For an ideal plug input shape, the second moment can be expressed as

$$\sigma_{inj}^2 = h^2 / 12 \quad (3.4)$$

where h is the initial width of the sample plug, which can be estimated for electromigration from the injection voltage, the total capillary length, and the injection time [28].

The sample amount loaded into the capillary needed to be optimized in order to achieve the optimum separation. Overloading will deteriorate the resolution while underloading may result in a low signal-to-noise ratio.

The sample load is determined by the sample concentration, the injection time, and the injection voltage. Significant zone broadening and distortion could occur due to improper sample injection. This can be caused either by distortion of the applied electric field when the concentration of sample ions is too high or simply by injection of too large a zone [22].

Fig. 3.7 presents the sample concentration effect on separation with a given injection condition. The results show that sample concentration is an important factor for separation. The experiment was carried out by using ABI standard samples with different concentrations, including the concentrated ABI standard and 2×, 4×, 8×, and 16× dilutions. The injection conditions were identical for the five samples, that is, 20 second at 100 V/cm.

The concentrated sample gave very broad and distorted peak shape and showed very poor separation, which indicated overloading of the sample. Much better separation was

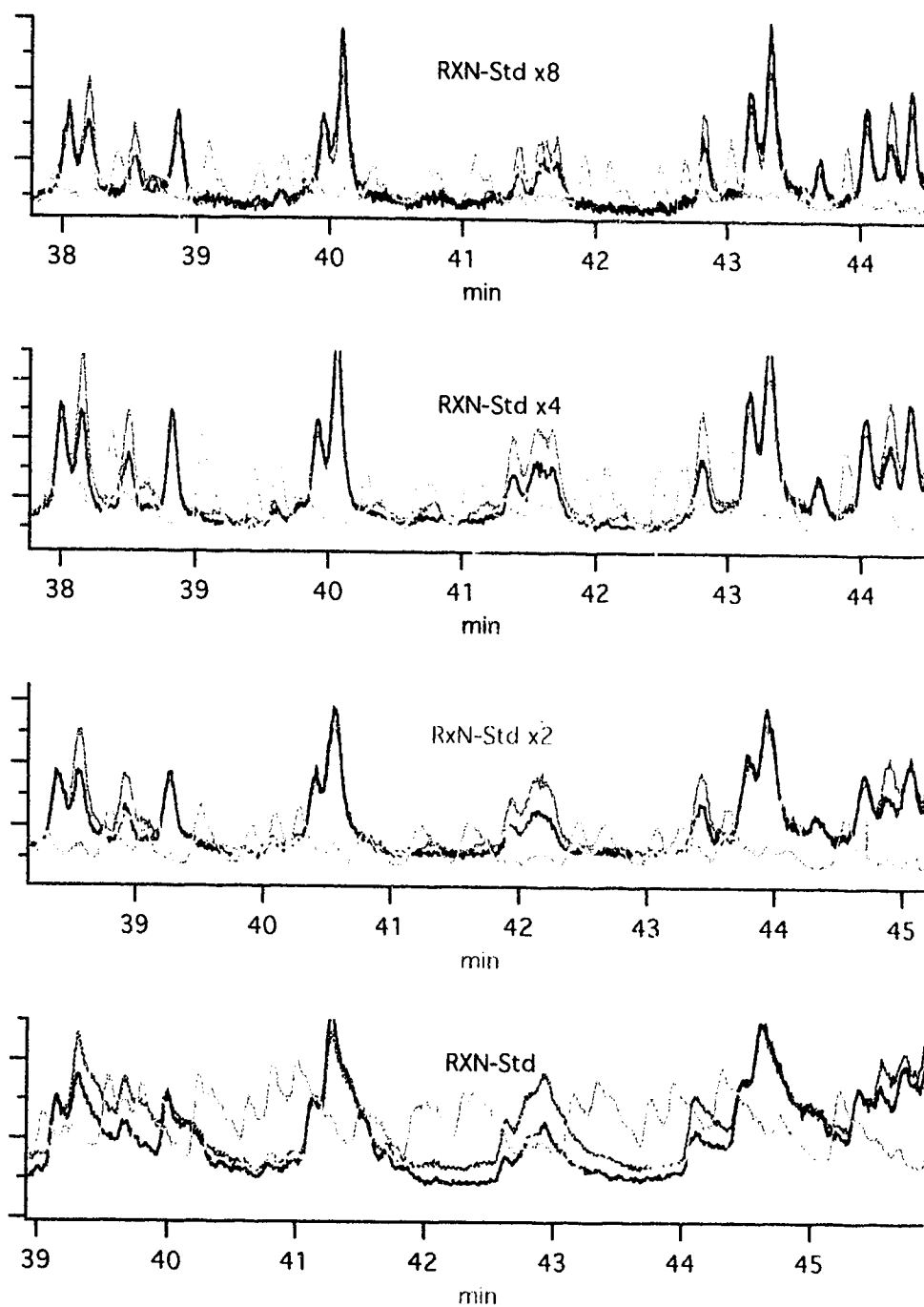


Figure 3.7 Effect of sample concentration on separation with a given injection condition.

obtained for the diluted samples. The 4× and 8× dilutions show good results and are reasonable concentrations for the injection conditions used. For the 16× dilution, almost no signal other than the primer peak is observed (the result is not presented here), which means that the concentration is too low to provide useful sequence information.

If low concentration samples have to be injected, longer injection time may be used in order to introduce more sample ions. However, a very long injection time may also cause overloading. Overloading the capillary by introducing a sample plug that is too large results in extra zone broadening. Such overloading will deteriorate resolution too [32]. Fig. 3.7.1 shows the effects of injection time on separation efficiency N_p (a) and resolution R (b). There is no significant difference in separation efficiency for injection times of 30 and 60 seconds, but when the injection time increased to 90 seconds or 2 minutes, the plate count decreased. This decrease is probably due to the width of the starting band. A similar trend is observed for the effect of injection time on resolution R (Fig. 3.7.1 (b)).

Sample injection is a critical issue for separation and deserves special attention. In addition to the sample concentration and injection time, several peculiar effects related to sample introduction and matrix effects have been reported [32]. Grushka and McCormick [71] found that the insertion of the capillary into a sample vial may result in an inadvertent injection. A detailed study of this effect, caused by an interfacial pressure difference at the inlet of the capillary (across the curved surface of the droplet at the end of the capillary, causing a small amount of sample to be pulled into the capillary), was recently published by Fishman et al. [72]. This “ubiquitous” injection occurs when the capillary just touches a sample solution. Peak broadening due to penetration of residual

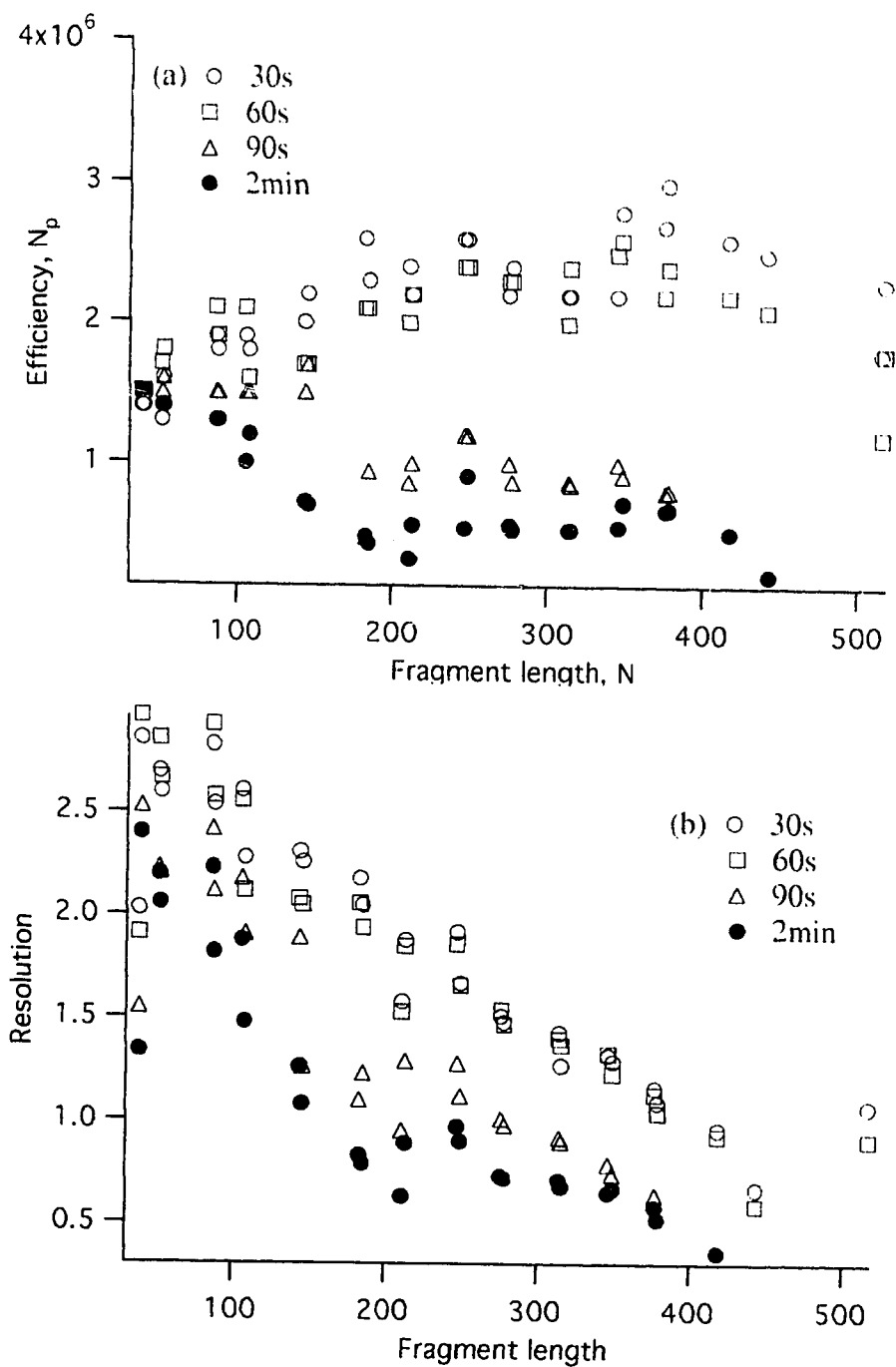


Figure 3.7.1 Effect of injection time on (a) separation efficiency N_p ; and (b) resolution R for different fragment length

analyte from the outside of capillary was described by Lux et al. [73]. During injection, “carryover” liquid at the end of the capillary results in peak asymmetry. The authors suggested an additional rinsing step in the sample introduction procedure. By simply dipping the capillary in buffer after sample introduction, the problem can be prevented. Ermakov et al. [74] observed artifactual peak splitting due to interaction of sample and background electrolyte.

Substantial peak distortion and impaired separation performance can occur if the tip is not cut perpendicularly to the capillary axis at the inlet. Hence, careful cutting and examination of the capillary tip prevents any undesirable peak distortion and consequent resolution loss [32].

After studying different separation parameters, the optimum separation conditions for DNA sequencing with the five capillary CZE system were obtained and are summarized in Table 3.2. These conditions are to be used for sequencing real samples.

3.5 The reuse of the separation matrix

Column lifetime is an important issue. Early reports on capillary gel electrophoresis separations of oligonucleotide standards stated that at least a hundred separations could be performed without replacement of the gel [35]. However, no group has been able to reuse gel filled capillaries for many separations when dealing with DNA sequencing samples. At electric field strengths greater than 300 V/cm or after repeated use, the separation performance deteriorates. Analyte mobilities increase, resolution decreases, and sometimes the gel fails completely. The methods to assess the stability of gel include monitoring the current and measuring the resolution of adjacent peaks [4].

Table 3.2 Separation conditions used for DNA sequencing by the multiple capillary CZE system.

Gel:	5 % T, 5-7M urea, 1× TBE
Capillary:	50 μ m (I.D.) L= 40-48 cm
Running buffer:	1× TBE
Injection Time:	10 ~ 60 seconds
Injection Voltage:	100 ~ 150 V/ cm
Running voltage:	150 V/cm-200 V/cm

Bubble formation in the capillary during polymerization of the polymer matrix or during later use is one catastrophic failure mode. The use of DNA sequencing samples creates additional problems. The high molecular weight and high charge template DNA has been shown to produce a band of high conductivity in the gel [75]. The slow-moving template generates a hot region in the capillary, which results in instability in the gel. This is observed as a dramatic decrease in current. As a consequence, this localized heating may reduce the resolving power of the capillary gels and impair the signals [4] or even damage the gel. Attempts to load increasing amounts of such samples, in order to increase the readability of sequence, resulted in diminished resolution and reproducibility, especially for large molecular weight fragments. The complications that arise from loading real DNA samples make the re-use of capillary difficult [75].

The reuse of the separation column is in practical concern. Reusing the capillary decreases cost by eliminating instrument re-alignment and gel preparation in addition to saving the capillary and gel. One of the reasons to prevent the capillary from reuse is the formation of ion depletion zone in the gel-buffer interface during separation, which results in a decrease in current. One method used for reuse of capillaries was allowing the capillary to recover overnight. The recovery is consistent with diffusion of ions into the depletion zone [76]. This method works, but it limits the instrument to one separation per day. Our group has found it helpful to reverse the electrical polarity for a brief period after each separation. I tried this method in the four color sequencing experiment. After the first separation, an electric field of 200 V/cm with opposite polarity is employed for 30 minutes. This reversed polarity returns the current back to its original value. This procedure allow to do a second run right away, which increased instrument throughput remarkably.

Fig 3.8 presents the normal current and subsequent reverse-polarity current. The electrophoretic current decreases gradually during the run, presumably due to the induced

depletion zone of ions build-up at the gel-buffer interface [76] that substantially reduces the effective electric field strength to drive the electrophoresis. The current recovered to its original value after a 30 minute run at reversed polarity.

Fig 3.8.1 shows the electropherograms of two consecutive runs for a M13mp18 sample. There is a 30 minute polarity reversal between the first and second run. The results show that both runs have good resolution. The pentet T (red peaks) and quartet T (red peaks) around 320 base pairs are almost baseline resolved, and the sequence can be read up to 550 base pairs. Repeated operation of the same capillary allows generation of 1100 bases in 4 - 5 hours for one capillary, or 5500 bases sequenced with the five capillary system for the same time period. The results are very encouraging. However, more experiments need to be done to find the limits for the reuse of the capillary by polarity reversal for DNA sequencing with the multiple capillary system.

There are some other methods that have been suggested for improving the lifetime of the gel. Buffer additives [59], e.g. acrylamide monomers and formamide etc., have been added to match the organic environment of the gel with that of the running buffer for preventing formation of a liquid junction potential across a zone depleted of ions. The depletion zone causes heating and damage to the gel. The buffer additives counteract the depletion of ions occurring at the gel-liquid interface and increase the stability of the gel substantially. Removal of excess template DNA after the performance of the sequencing reactions was also shown to be essential for gel stability and to obtain multiple runs on the same capillary [75, 77]. One of the methods used in our experiments was solid phase purification [77].

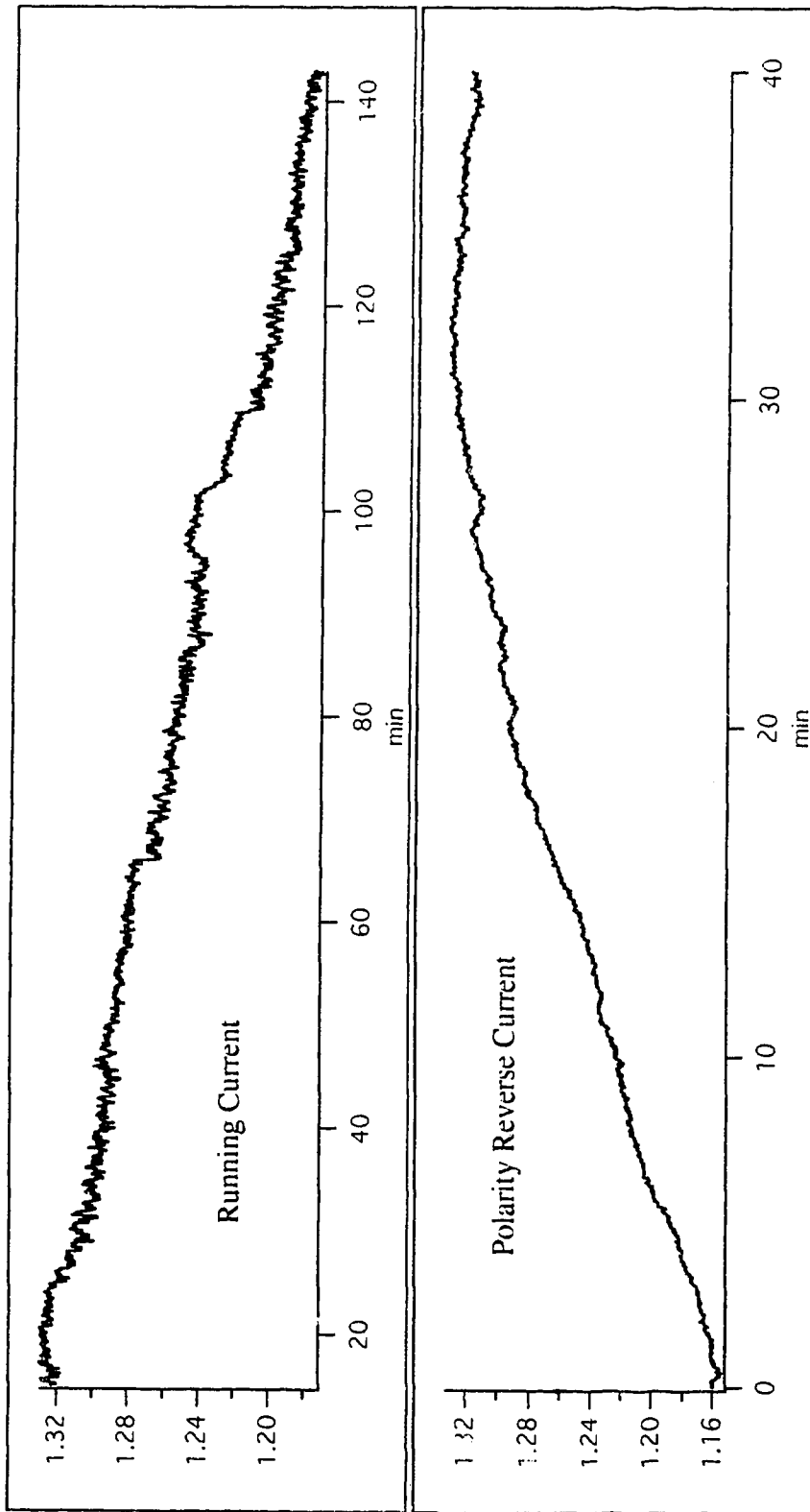


Figure 3.8 The normal running current and reverse polarity current

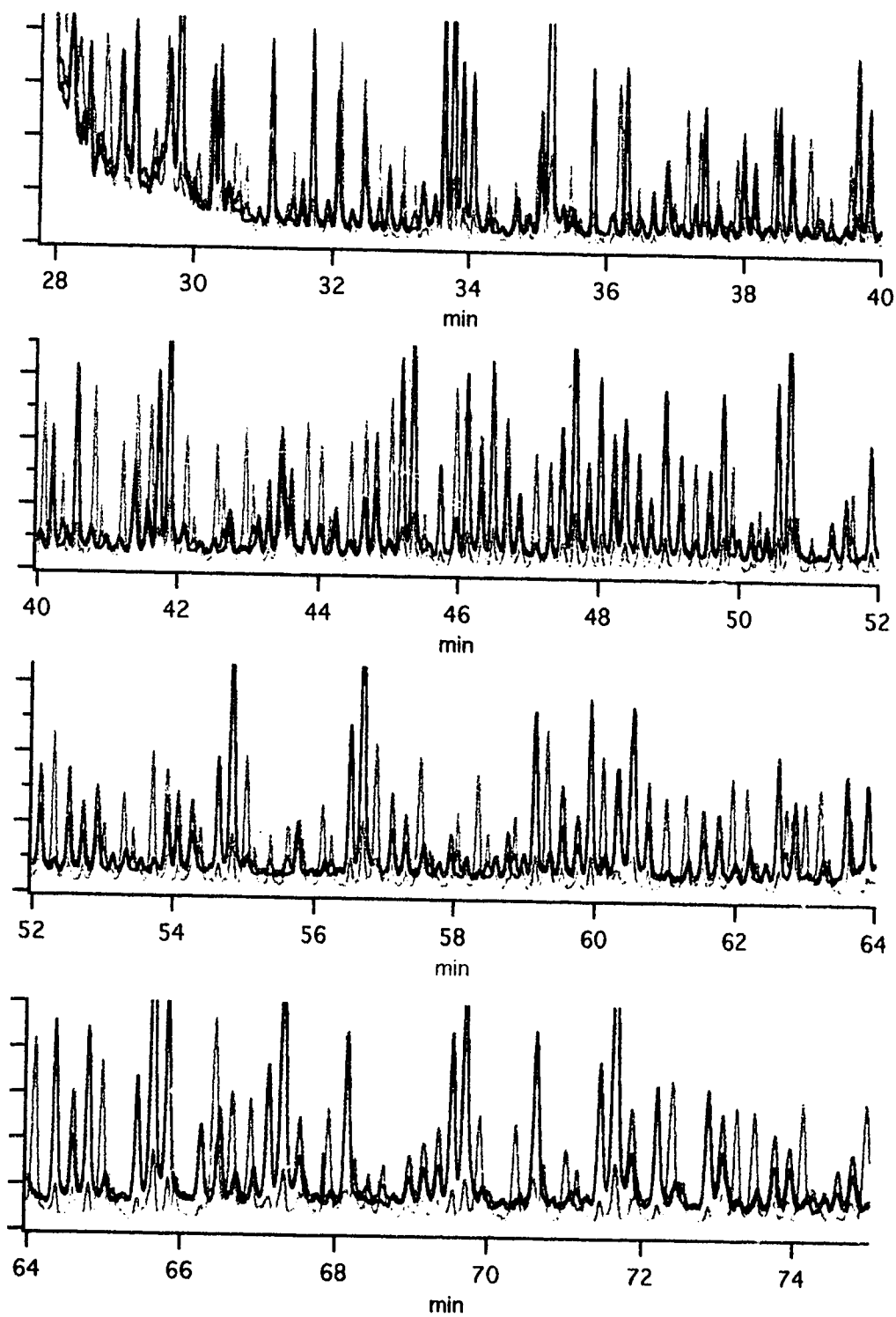


Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18
(a) first time run

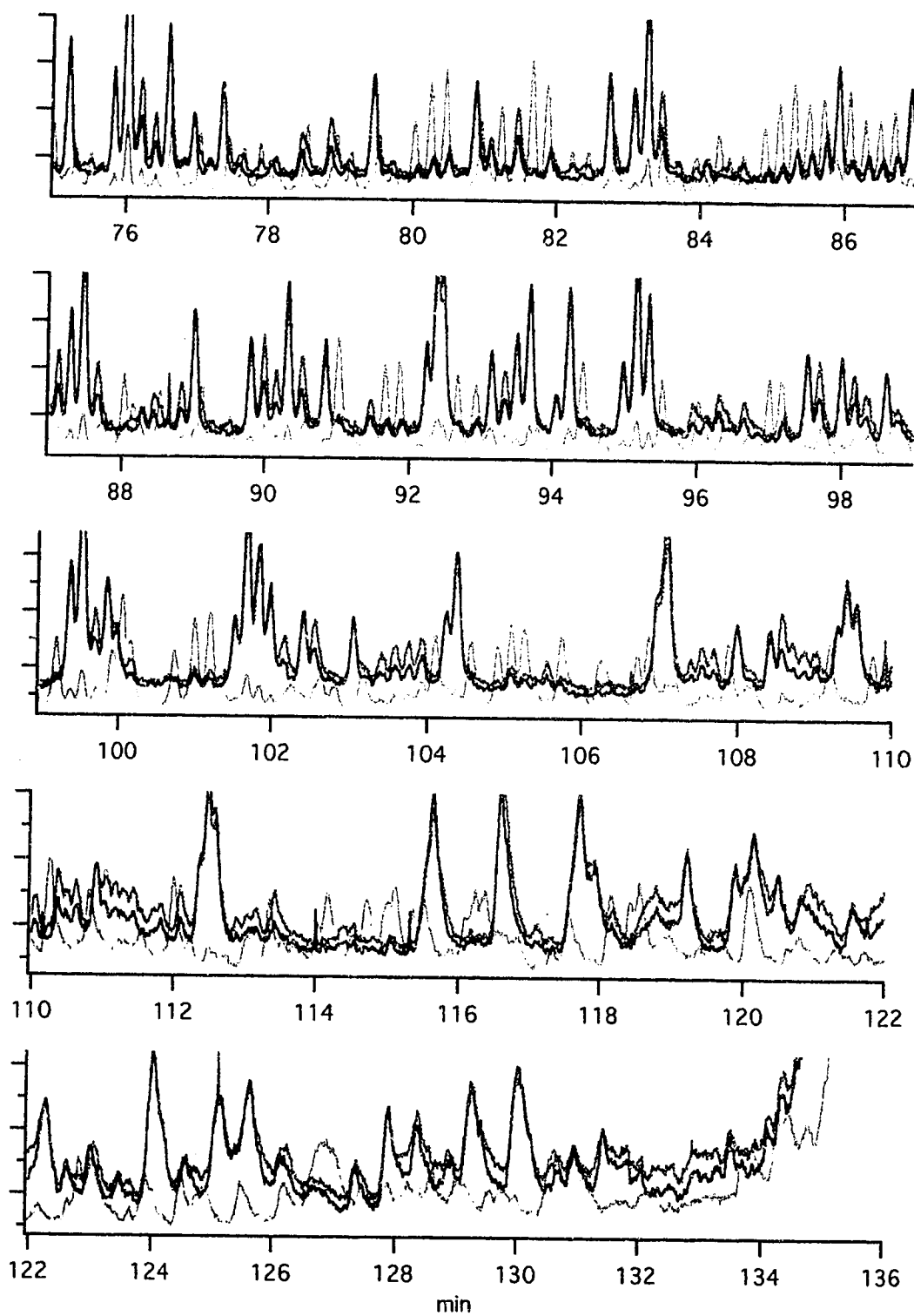


Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18
(a) first time run

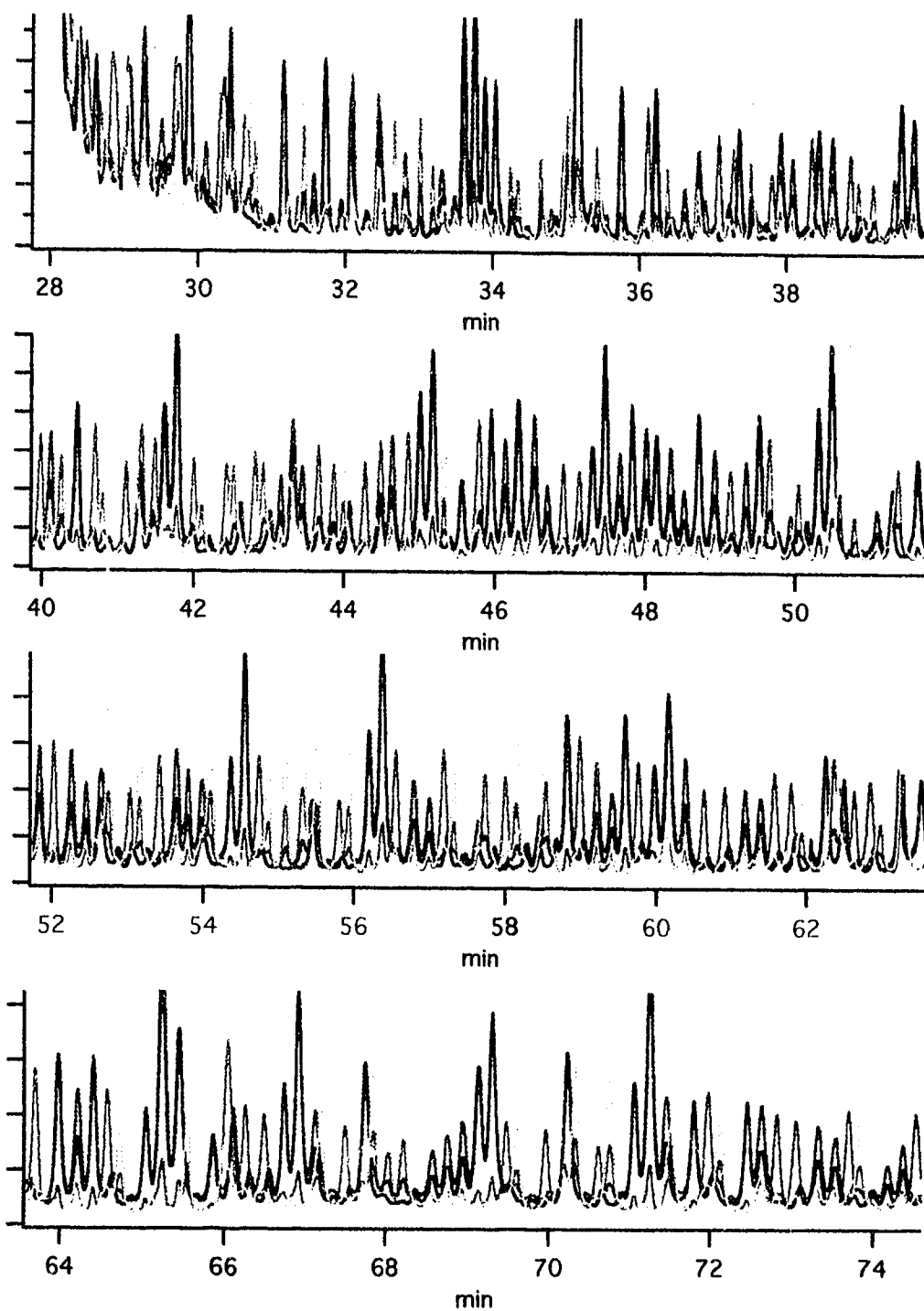


Figure 3.8.1 The electropherograms of two consecutive runs for sample M13mp18
(b) second time run after 30 min. polarity reversal

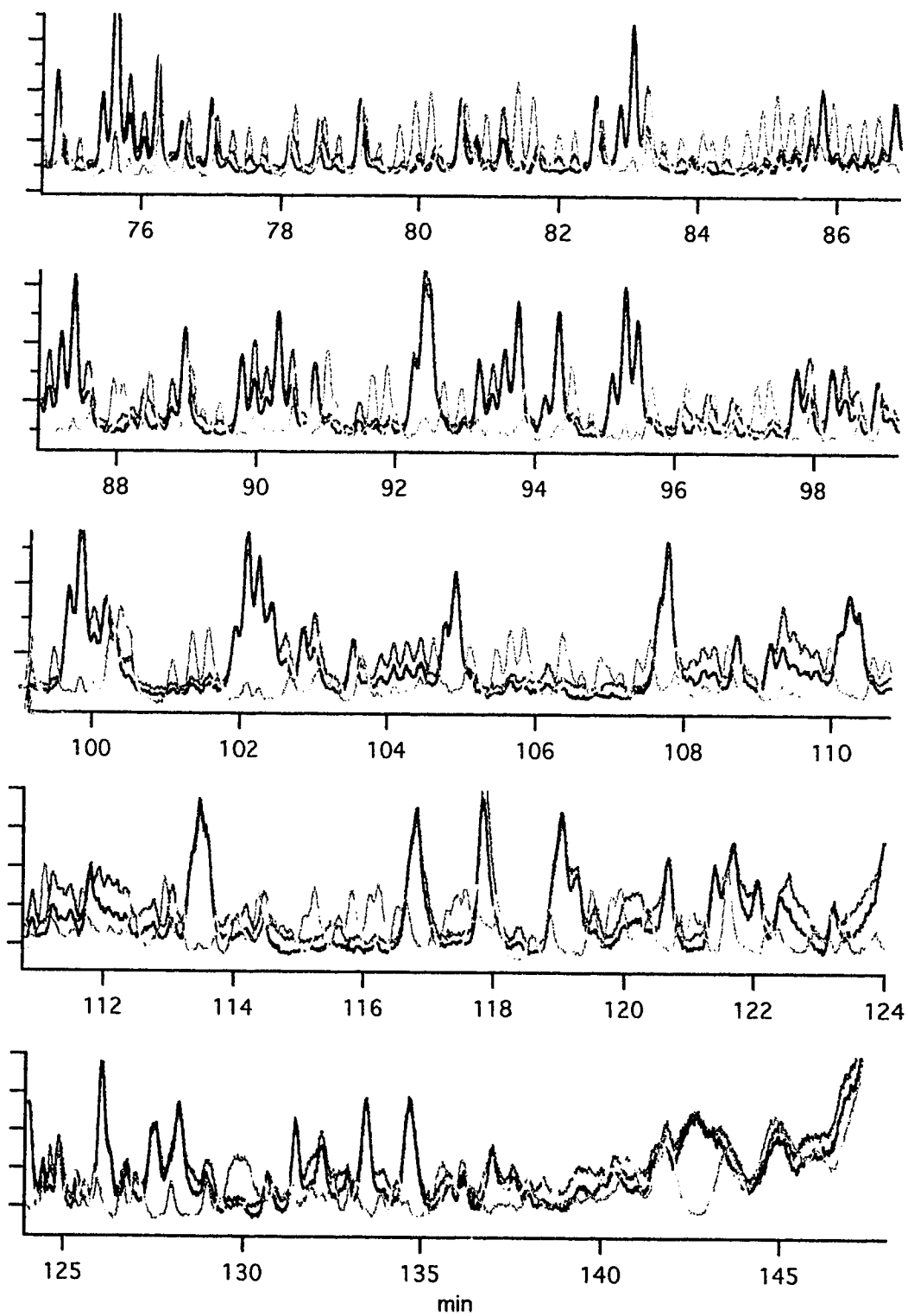


Figure 3.8.1 The electropherograms of two consecutive runs for sample M13 mp18
(b) second time run after 30 min. polarity reversal

The solid phase method is carried out using a Dynabead template preparation starter kit [78]. This kit uses streptavidin-coated magnetizable particles. Immediately after lysis of a single bacteria colony, amplification through PCR is carried out by two universal primers. One of the primers used for the PCR is biotinylated, resulting in an amplification product with one strand having a biotin group at the 5' end. The DNA duplex is selectively captured and immobilized onto Dynabeads and purified by magnetic separation. The complementary strand eluted into the alkaline supernatant and was removed. The immobilized single-stranded DNA template obtained from the primer set is subsequently sequenced using a forward sequencing primer. After the sequencing reaction, the extended material is purified by magnetic separation, eluted from the Dynabeads with formamide and loaded directly onto a sequencing gel. The sample obtained by this way has very high purity.

3.6 Effect of sample preparation conditions

Sample preparation is another factor affecting DNA sequencing. Both conventional isothermal sample preparation with Sequenase [79] and cycle sequencing with thermostable polymerases [80] have been used for these experiments. In the Sequenase method, a few micrograms of template is needed for a reaction. In comparison, cycle-sequencing requires much less template (a few nanograms) to produce ample quantities of sequencing products.

For the Sequenase method, the sequencing reaction is carried out in 40 mM MOPS buffer, pH 7.5, 50 mM NaCl, 10 mM MnCl₂, and 15 mM sodium isocitrate. For the first reaction, 2 µl of 5 µM tetramethylrhodamine labeled primer (Applied Biosystem-21M13 TAMRA) is annealed to 1 µg of M13mp18 single stranded DNA at 65°C for 2 minutes followed by slow cooling. A mixture of deoxy- and dideoxynucleoside triphosphates is added with 7-deaza-2'-deoxyguanosine-5'-triphosphate used in place of dGTP. After the

mixture is warmed to 37°C, 6 units of Sequenase Version 2.0 and 0.006 units of pyrophosphatase are added. Incubation continues at 37 °C for 30 min., after which the DNA is precipitated with ethanol. Identical experimental conditions are used with a sulforhodamine-labeled primer (Applied Biosystems -21M13 ROX), FAM, and JOE primers. The pooled samples are resuspended in a 49:1 mixture of formamide: 0.5 M aqueous EDTA. Three microliter aliquots are taken from the resuspended samples, heated to 95°C for two minutes, and injected onto the capillary.

A Sequitherm long-read cycle sequencing kit is used for cycle sequencing with FAM, JOE, TAMRA, and ROX labeled primers (Applied Biosystems). Cycle sequencing reactions are carried out according to the protocols recommended by the manufacturer. Cycle sequencing was performed using 30 cycles on a PTC-100 programmable thermal controller (MJ Research, Inc). Each cycle consists of 15 seconds at 95 °C and 90 seconds at 70 °C. Reaction products are pooled and immediately ethanol precipitated. The dried pellet is resuspended in 1.5 µl of formamide and kept in a refrigerator, ready for injection.

The first peaks that usually showed up in the electropherogram was the unextended primer. The later peaks showed a roughly exponential decrease in signal with time. This decrease in signal presumably is due to the first-order kinetics that describe the incorporation of chain-terminating dideoxynucleotides into the growing sequencing fragment [63].

DNA sample preparation can directly affect the quality of the sequencing results. For example, a gap or small peaks were occasionally observed in the electropherogram, this is due to illegitimate termination [81]. A small peak or a shoulder shown up in the electropherogram usually suggests the presence of minor derivatives [36]. Compared to single terminator (single base) reactions, the recognizable sequence patterns for four

terminators (four color) samples are often shorter because of congestion in the electropherogram.

3.6.1 Effect of ddNTP:dNTP ratio and with or without Mn^{+2}

The presence of some metal ions such as Mn^{+2} or Mg^{+2} in the sample preparation reaction may affect the relative peak height of different DNA fragments. Addition of Mn^{+2} into sequencing reaction mixture can produce electropherograms of much more uniform peak height and result in better accuracy in base calling when Sequenase is used as the DNA polymerase. Fig. 3.9 shows the comparison of two samples prepared with or without Mn^{+2} . The samples used for this section are prepared using different conditions that are listed in Table 3.3.

Significant variations in peak intensity were observed for a single terminator reaction performed using Sequenase in the absence of Mn^{+2} . In contrast, a remarkable similarity in the size of neighboring peaks is observed in electropherograms of reaction mixtures containing Mn^{+2} (see Fig 3.9). In addition to the large peak variation, false terminations in the absence of Mn^{+2} may occur. The false terminations and the lack of uniformity in peak intensity observed in the absence of Mn^{+2} could cause sequence reading error and preclude the possibility of sequence determination using peak height analysis.

Another important factor in optimizing the sequencing reaction is the proper selection of the ddNTP to dNTP ratio. Fig 3.10 shows the comparison of two samples with different dNTP:ddNTP ratios, 100:1 and 200:1. As the relative concentration of terminator, ddNTP, is increased (i.e, from 200:1 to 100:1), the extended primer is consumed more rapidly as indicated by the much smaller primer peak in Fig. 3.10 (b). Consequently, this results in a faster decrease in signal. Additionally, the earliest eluting

Table 3.3 Sample preparation conditions for Mn^{2+} effect and ddNTP:dNTP ratio effect.

sample	ddNTP:dNTP ratio	Mg^{2+}	Mn^{2+}
S1	1:100	yes	no
S2	1:100	yes	yes
S3	1:200	yes	no
S4	1:200	yes	yes

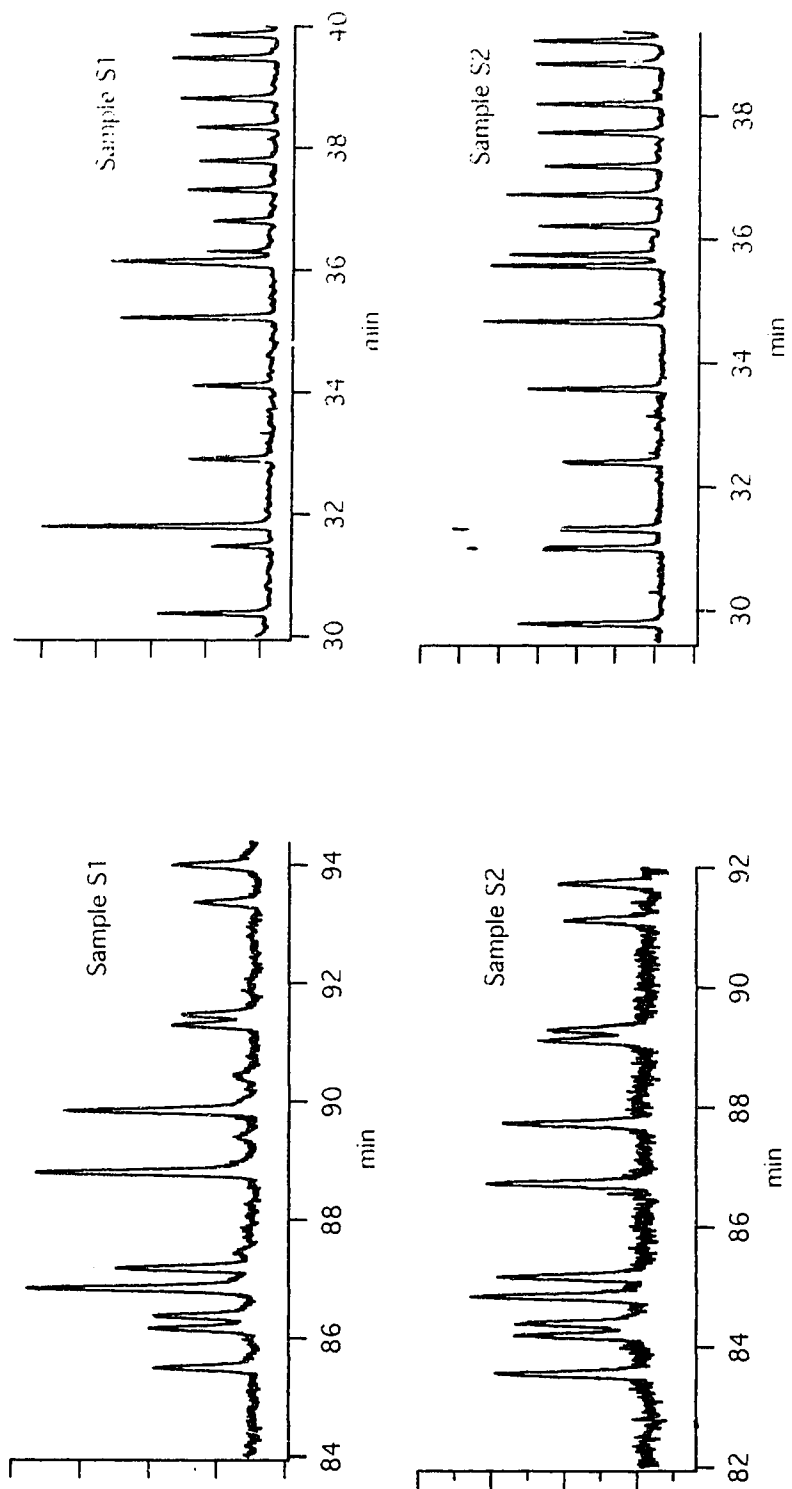


Figure 3.9 Comparison of two samples with or without Mn^{2+} . Zoomed view. Sample S1 has no Mn^{2+} . Sample S2, with Mn^{2+}

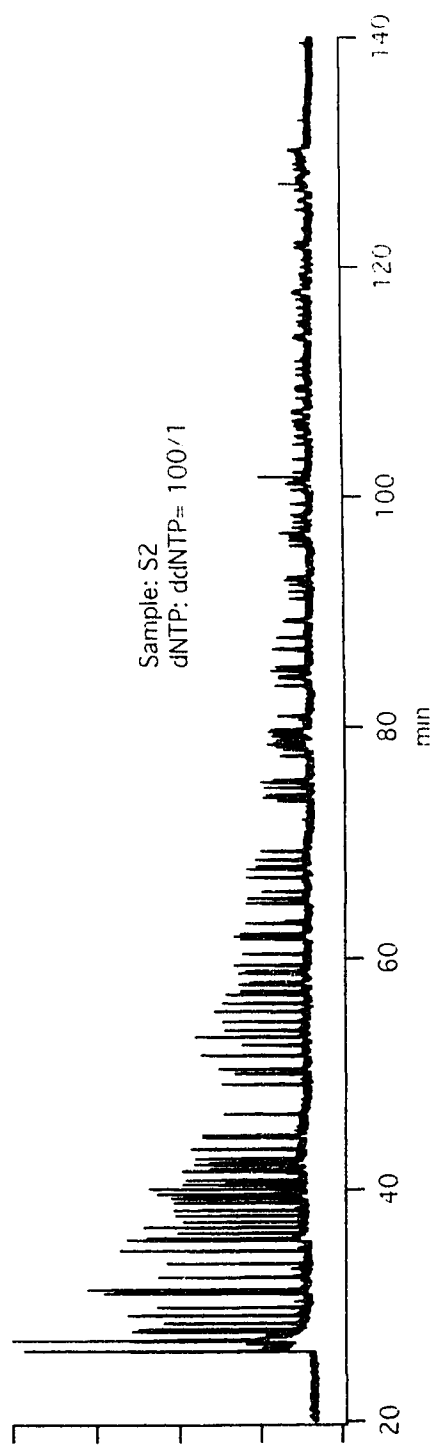
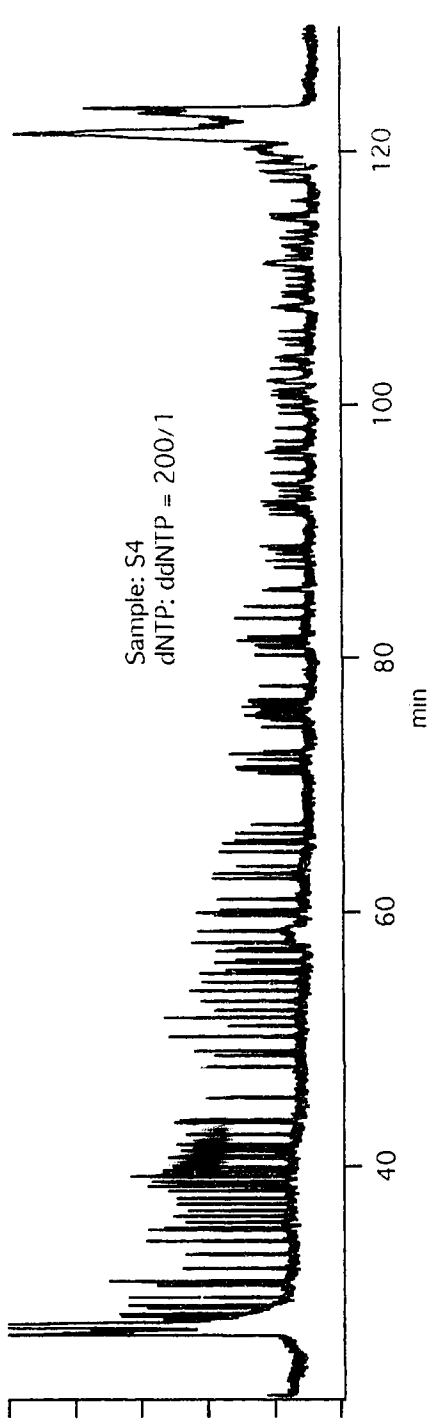


Figure 3.10 Comparison of two samples with different dNTP: ddNTP ratios, 100/1 and 200/1

peaks are observed to be more intense due to the high degree of ddT incorporation at the first few termination sites. For a dNTP:ddNTP ratio of 100:1, the terminal peak is hardly seen, while a big sharp terminal peak is observed for a ratio of 200:1. A smaller ratio of dNTP:ddNTP (e.g. 50:1) can even result in the loss of signal-to-noise ratio prior to loss of gel resolution. When a ratio of 500:1 is used, a significant amount of dNTP is wasted in the unresolved portion of the trace at the very end of the electropherogram [23].

The optimum ratio of dNTP:ddNTP should provide long readability without wasting dNTP.

3.7. Summary

Experimental conditions affecting DNA sequencing by the five capillary CGE system are studied and optimized. The non-crosslinked gel, prepared at room temperature, of 5 %T with 7M urea and 1x TBE is used as separation matrix. Capillaries of 50 μm i.d. and 40 cm length give optimum separation efficiency and speed with the running voltage around 150 V/cm - 200 V/cm. The 1x TBE is used as running buffer.

DNA samples prepared with different conditions are tested by the five capillary CGE system. The results show that the samples prepared with presence of Mn^{+2} and with the ddNTP:dNTP ratio of 1:200 gives uniform signals across wide range of fragment lengths.

DNA sample cleaning process is very important in order to obtain good separation results and read length.

Conditioning the capillary by polarity reversal seems helpful for re-using capillary in DNA sequencing.

The amount of sample injected into the capillary needed to be controlled properly by controlling the injection voltage, injection time, and sample concentration.

Chapter 4

Sequence data processing and base calling

DNA sequencing is a multistage process that includes the preparation of a DNA sample, the electrophoretic separation of the fragments, analysis of the data to determine the nucleotide sequence, and the interpretation of the resulting sequence information. As discussed in section 3.5, the current five capillary system can generate a DNA sequence at a rate as high as 1100 bases/hr. A two dimensional array system with hundreds of capillaries is under development in this group. The amount of sequencing data generated by such a system is enormous and thus an efficient data processing procedure must be developed.

An electropherogram consist of fluorescence peaks corresponding to one of the four bases, plus noise and interference from adjacent peaks. The sequence of the bases has to be recognized (called) from the electropherogram. Conventionally, the base calling procedure is performed by a human reading the electropherogram. The pattern recognition capability of an experienced person for the electropherogram is exceptionally good. However, manual base calling is tedious and time-consuming. Sometimes it is also error-prone when the operator becomes tired. The automation of base calling is an important aspect of sequencing technology. Especially, since the high throughput of the multicapillary DNA sequencing system demands high speed data handling and base calling methods [82].

A base calling computer program was developed by Larry Coulson and Dr. Xiang Rong Liu. Currently, the program is customized to work with the sequencing data of the five capillary system. The data processing and base calling program is written with

LabVIEW (National Instruments, 6504 Bridge Point Parkway, Austin, TX 78730-5039, USA) in order to use the built-in graphic functions of that program.

The data processing involves several steps in order to generate the final DNA sequence. These steps are channel separation (or decimation), data editing, spike detection and removal, signal/noise enhancement, resolution enhancement, normalization, and base calling.

4.1 Channel separation and data editing

The data is collected using a program written with LabView. Usually there are six analog input channels: five capillary channels and one current channel. The data is collected at about 8 Hz/channel (i.e. 2 data points/s/color) and is saved in an interleaved format, as illustrated below:

```

1.....7.....13.....19.....25.....30
0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5
  Y-Y-Y-Y-Y   G-G-G-G-G   B-B-B-B-B   R-R-R-R-R   Y-Y-Y-Y-Y

31.....37.....43.....49.....55.....
0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5-0-1-2-3-4-5-
  G-G-G-G-G   B-B-B-B-B   R-R-R-R-R   Y-Y-Y-Y-Y   G-G-G-G-G

61.....67... (Data number in the data file)
0-1-2-3-4-5-0-1. (Channel Number for each filter)
  B-B-B-B-B   R   (Filter's color: Y(Yellow), G(Green), B(Blue), R(Red))

```

Channel 0 is reserved for a measurement of the total current flowing through the capillaries; channels 1-5 represent data generated from the five capillaries. The data decimation process breaks the six channel interleaved data file into six individual channel data files. For example, the data file for channel 1 after decimation will consist of data points #2, #8, #14, #20, #26, #32, #38, #44, #50.....n6+2. Each channel's data file itself contains signals of the four colors in an interleaved format in the order of Y-G-B-R..... Y-G-B-R -..... Y-G-B-R -..... This procedure is done by the data processing program.

4.2. Spike peak detection and removal

Sometimes, the electropherogram contains spike peaks, which may be due to the scattering of the laser beam by small particles or bubbles in the sheath flow cuvette. These spikes are usually very intense but have a short duration. It is very easy to identify them by visual examination of the data. If not removed properly these spike peaks may cause errors at the base calling stage. The amplitude of the spikes can be reduced by low pass digital filtering. However, they can not be removed completely by low pass filtering because such signals have very broad frequency ranges.

A simple method is used to detect and remove the spike peaks automatically. In this method the ratio $I_n/(I_{n-1}+I_{n+1})$ is used to identify the spikes. I_n is the intensity of point n, I_{n-1} is intensity of point n-1, and I_{n+1} is the intensity of point n+1. For the sequence data, a threshold of 0.75 is adequate to distinguish the spike from normal peaks (if $I_n/(I_{n-1}+I_{n+1}) > 0.75$ then the data point corresponds to a spike). Once the spike is detected it is then removed by replacing the intensity of point n with the average of points n-1 and n+1.

4.3 Bandpass digital filtering

To ensure accurate base calling with simple logic, it is essential to pre-process the data for signal-to-noise (S/N) enhancement and resolution enhancement. Usually, there is a trade off between the S/N enhancement and resolution enhancement. With our DNA sequencing data, a simple bandpass filter is sufficient to achieve this goal.

The filtering process can be performed by convolution in the time domain (equation 4.1) or by FFT (Fast Fourier Transform) route in frequency domain (equation 4.2).

$$a(t) * b(t) = c_{ab}(\tau) \quad (4.1)$$

$$\downarrow \text{FT} \quad \downarrow \text{FT} \quad \uparrow \text{FT}^{-1}$$

$$A(f) \times B(f) = C_{AB}(f) \quad (4.2)$$

In the above equation, $a(t)$ is the time domain signal, $b(t)$ is the filtering function, $c_{ab}(\tau)$ is the filtered signal. $A(f)$ is the Fourier transform of the signal $a(t)$. $B(f)$ is the frequency domain filtering function, and $C_{AB}(f)$ is the product of $A(f)$ and $B(f)$.

Figure 4.1 shows a single color electropherogram, its Fourier transform, and the frequency domain filtering function. By multiplying the filtering function and the Fourier transform of the electropherogram, both high frequency and low frequency components of the signal are suppressed. The useful signal information is mostly within the bandwidth from 0 to 0.2 Hz. Most noise signals in our system have a white spectrum. By removing the high frequency components, the noise levels are reduced significantly. The suppression of the low frequency components helps to make the baseline flat and to separate some unresolved peaks. Similar filtering effects can be achieved by convolving the time domain signal with a Savitzky-Golay filter [83, 84].

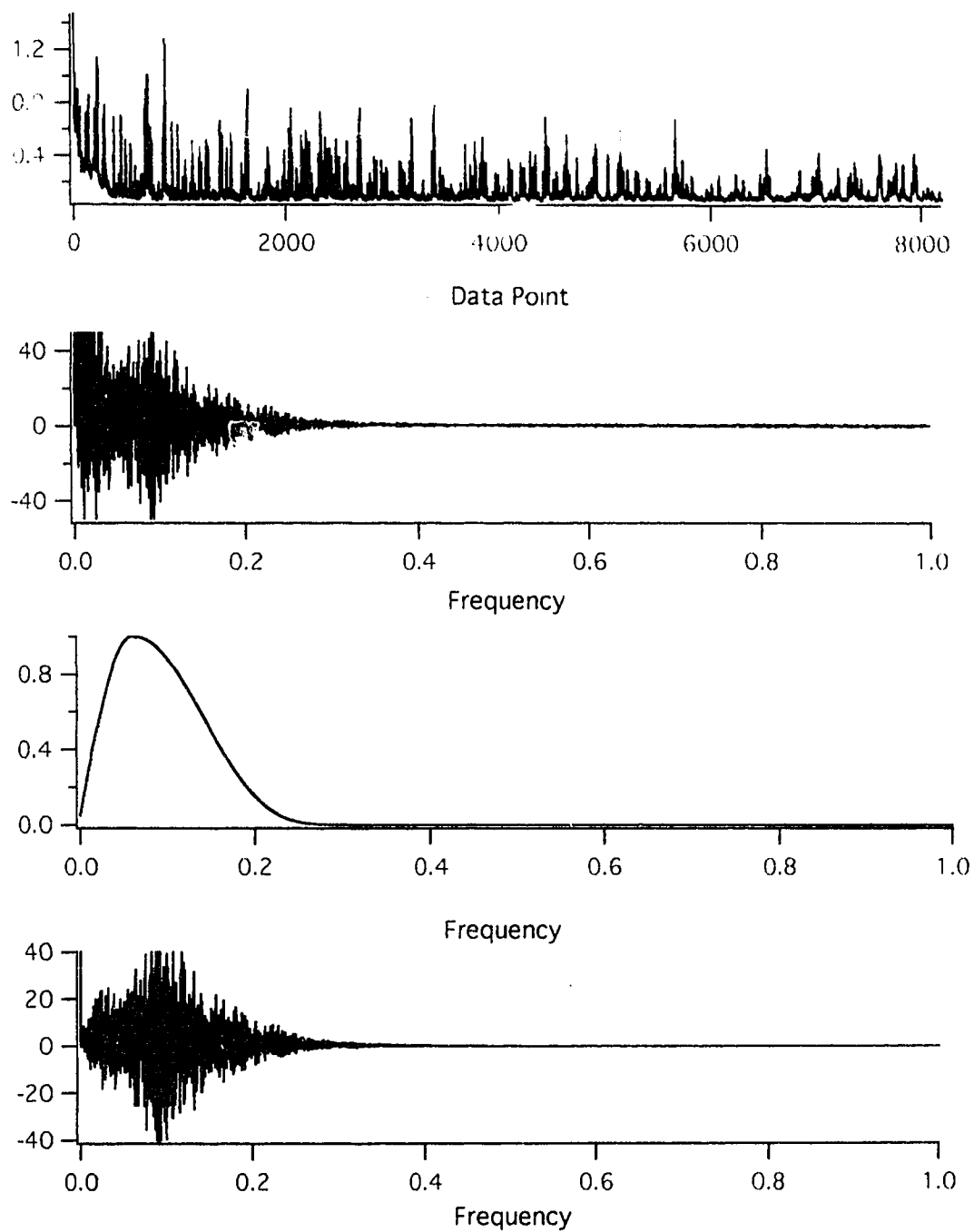


Figure 4.1 A single color electropherogram, its Fourier transform and the frequency domain filtering function

The peak width of the signal increases with migration time. That means the low frequency components of the signal increase towards the end of the electropherogram. On the other hand, the S/N usually decreases with the migration time because of a drop in signal. Therefore, it is desirable to gradually lower the high cut-off frequency of the filtering function as the migration time increases. In this way, the S/N ratio and the resolution are optimized for different regions of the electropherogram. The convolution approach is preferred for this purpose because it is simpler to implement and can provide smoother transition from region to region. This modification is done by simply incrementing the number of the points in the filtering function as a function of retention time. (e.g. increment for every 500 data points). A base line cut-off process is applied after the filtering process. A triangular convolving function, with the sum of index equal to zero, is used for filtering the data in the time domain.

Due to the variation of the sample preparation parameters, the relative signal intensity of the four colors may change from sample to sample. It is necessary to adjust the relative signal intensities of the four colors in order to facilitate effective base calling. This adjustment is done by normalizing each color signal to either its standard deviation or average peak height.

The signal intensities usually decrease with the migration time due to the decreased concentration of the larger fragments. However, for the base calling purpose, it is desirable to have a uniform signal intensity through out the electropherogram. Therefore a scaling process is applied to produce a more uniform signal.

Figure 4.2 shows a portion of the beginning and end of an electropherogram after the above mentioned preprocessing. The raw electropherogram is also plotted for comparison. The digital filtering process is especially helpful for the data with poor S/N and poor resolution. The data are then ready for base calling.

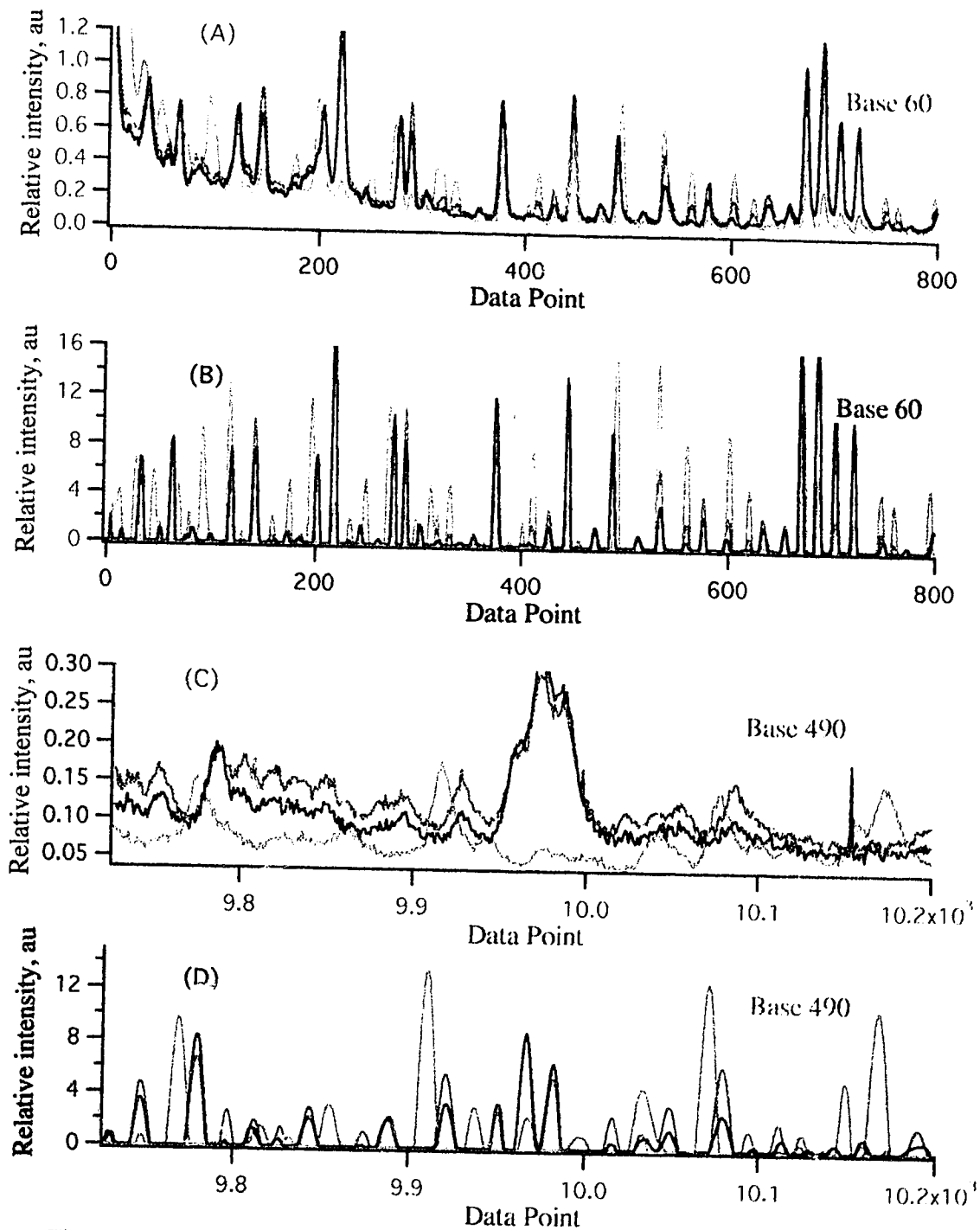


Figure 4.2 A portion of a four color electropherogram (A and C) and the processed data (B and D). (A), (B) the beginning of the electropherogram; (C), (D) the end of the electropherogram

4.4. Base calling

Because each base is tagged by one of four different fluorescence dyes, fluorescence is detected in four spectral channels. As mentioned in Chapter 1, the fluorescence labels used are fluorescein (FAM), modified fluorescein (JOE), tetramethyl rhodamine (TAMRA) and Texas red (ROX). The four different dyes allow all four termination reactions to be separated in one column. Each dye represents chain termination of a specific base. That is FAM for C, JOE for A, TAMRA for G, and ROX for T. The four fluorophores are denoted by four different colors in our electropherogram. Blue for C (FAM, 540 nm), green for A (JOE, 560 nm), yellow for G (TAMRA, 580 nm), and red for T (ROX, 610 nm).

In this program, the relative fluorescence intensity is used as the base calling criteria. From Fig. 1.5 of chapter 1, it is clear that the fluorescence generated by any of the four labeling fluorophores will pass through each of the four bandpass filters with different efficiencies. Therefore, base A will generate signals in all four channels with relative intensities of green>blue>yellow>red. For base C we will see signals with relative intensities of blue>green>>red>>yellow. For base G, the relative intensity is yellow>red>green>blue. For base T, the order of the signal intensities would be red>yellow>green~blue, see Figure 4.3. The relative signal intensities may be changed from sample to sample due to the changes of the sample preparation conditions, which may invalidate the base calling logic. As mentioned in section 4.3, such undesirable variation is effectively minimized by applying a normalization step after the bandpass filtering.

There are some peaks which will not satisfy any of the above conditions. In most of the cases the GC compression is responsible for such peaks. The program can check and

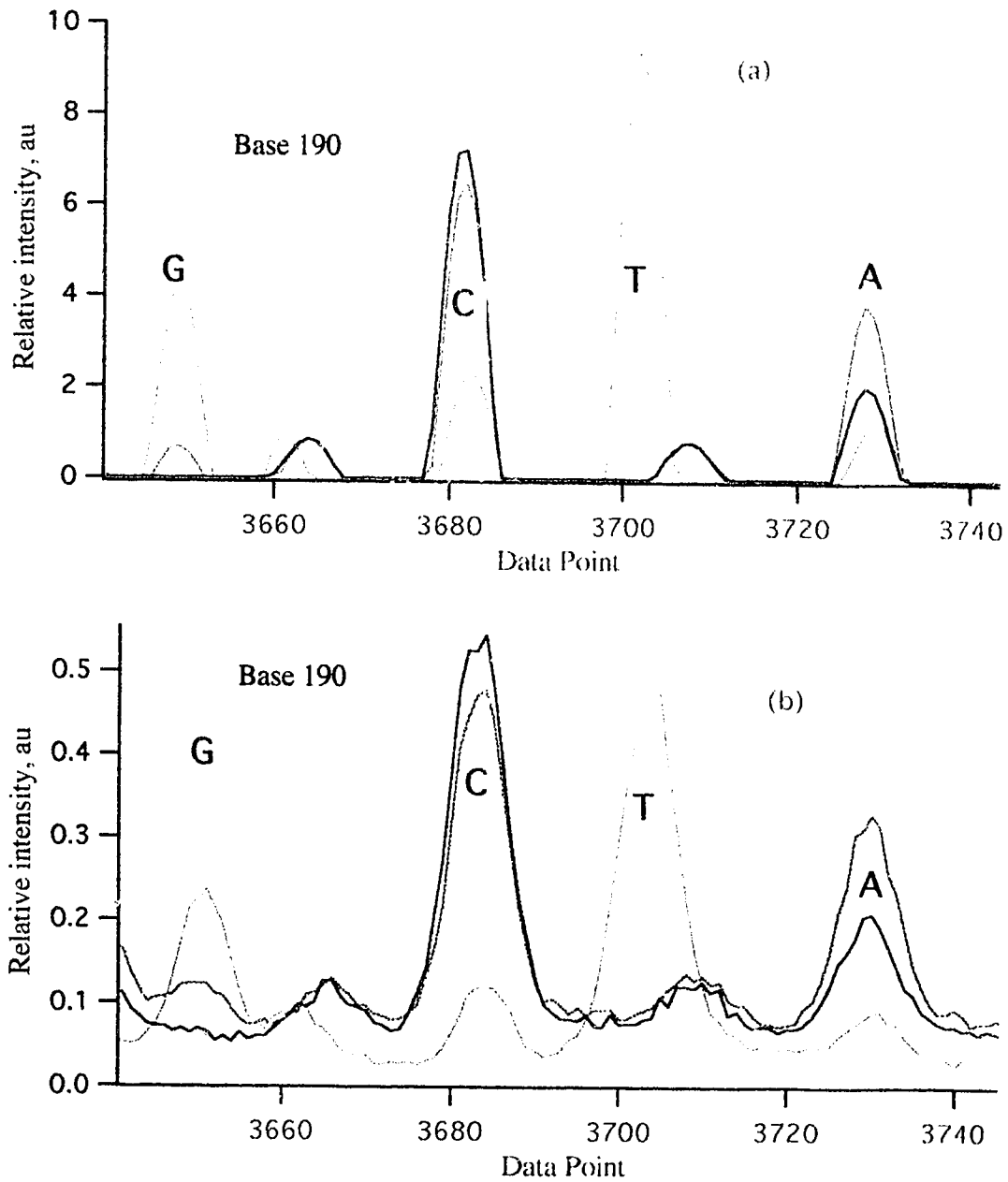


Figure 4.3 Relative fluorescence intensities of the four bases on the four bandpass filters. The bases are called according to their relative intensities after normalization and filtering (a). The raw data is plotted in (b)

call such peaks.

Spurious peaks are occasionally observed in the electropherogram, especially for the cycle-sequencing samples. Such peaks may result in errors in base calling. After the bases are called, the sequences are double checked for these spurious peaks by checking the spacing and the neighboring bases and intensities.

4.5 Summary

The base calling program has been successfully used to call the sequences for our five capillary CZE DNA sequencing data. It can generate sequences in about 10 seconds/channel in Mac Quadra 950. The base calling accuracy is dependent on the quality of electropherogram and region of fragment size. Fig. 4.4 shows a plot of base calling accuracy vs. fragment size. The accuracy for the both ends of the electropherogram needs to be improved. This can be overcome by using overlapping sequences so that the bases at the beginning or end of a primer template can get shifted into the optimal region (size) for accurate sequencing.

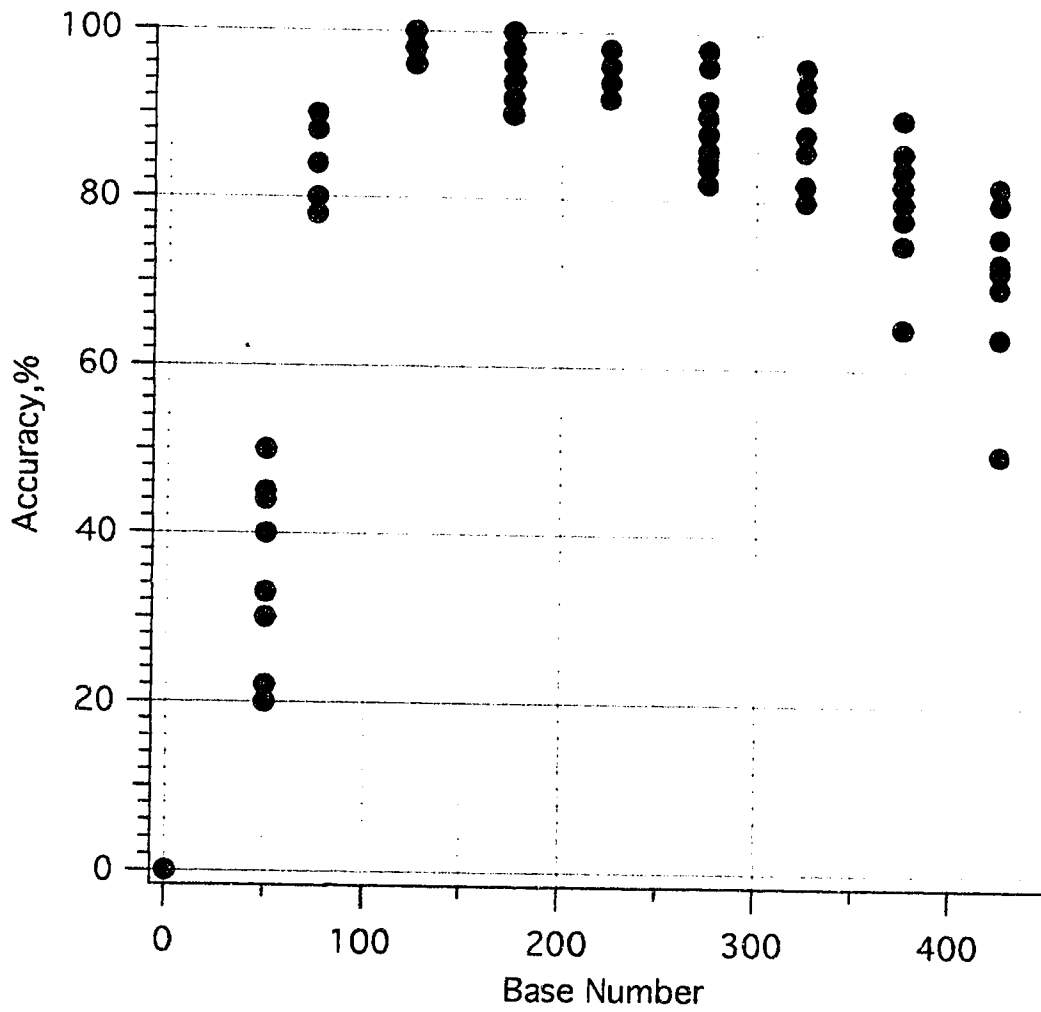


Figure 4.4 A plot of base calling accuracy vs. fragment size

Chapter 5

Summary and Future work

5.1 Evaluation of the system

During the course of this work, the five capillary system has shown great promise for rapid, sensitive and high throughput DNA sequencing. By operation at an electric field of 200 V/cm, It is possible to separate DNA sequencing fragments up to 5500 bases in length in about 4 - 5 hours. The separation is performed with 5% T 0%C polyacrylamide. This non-crosslinked matrix has sufficiently low viscosity that it can be pumped from the capillary and replaced with fresh material when required [37, 59]. Simultaneous analysis of a number of capillaries increases the throughput of the system. The five capillary system is ideally suited for small-scale sequencing projects. The instrument is quite robust, has been routinely used, and has generated several hundred sequencing runs for over 10 months.

Different sequencing protocols have been evaluated by using this system. These included the use of dye labeled primers with both T7 and Taq polymerase for cycle and conventional sequencing. Most work focused on the use of ABI's dye labeled primers for DNA sequencing. Real samples were prepared using solid phase sequencing method, where PCR fragments are captured on streptavidin coated magnetic beads. Cycle sequencing was the method most commonly used to generate the reaction fragments.

Table.5.1 Total number of sample run

Month	June	July	Aug.	Sept	Oct.	Nov.	Dec.	Jan	Feb.	March	Total
	94	94	94	94	94	94	94	95	95	95	
Samples	14	25	61	48	83	92	83	48	60	59	573
Data sets	55	90	142	97	195	199	165	109	106	113	1271

Note that samples refer to different sequencing reactions; data sets refer to the number of electropherograms generated. The same sample would occasionally be injected onto several capillaries at one time, which makes the sample number smaller than the data sets.

The instrument has been running for over 10 months, and is quite reliable and robust. During the time from June 1994 to March 1995, a total of 573 different sequencing reactions have been run through the system and 1,271 sequencing data sets (1 data set/channel/run) have been collected. Table 5.1 summarize the workload of the instrument during that period.

The results of this work demonstrate the potential of multiple capillary system for DNA sequencing. Separation conditions have been optimized. A computer base-calling program has been tested and used for data processing and sequence reading. Sample preparation and injection conditions have been explored.

The biased reptation model has been discussed in this work. The classic biased reptation model predicts that $\beta=2$; the onset of biased reptation is predicted to decrease with the square of electric field. A new model, called biased reptation with fluctuations (BRF), predicts several regimes. In the case of very flexible molecules, the BRF model predicts that $\beta=1$. Our data is consistent with this regime of the BRF model. We observe no evidence for an inverse quadratic dependence, which is required by the classic biased reptation model.

5.2 Future work

Even though the five capillary system shows promise for DNA sequencing, there is still room for improvement. The next few paragraphs discuss some of the areas that could be explored in the future to further improve the performance of the system.

Enhanced separation of DNA molecules could be achieved by applying a nonuniform electric field (field strength gradient) as a function of time [35]. Pulsed-field electrophoresis [85, 86] (changing the direction and magnitude of the field in an oscillating manner) could be employed to take advantage of the elongated and oriented

configuration of large DNA molecules in gels. This method may be especially helpful for improving the separation of large fragments.

Running the separation at elevated temperatures may help to reduce the secondary structure of the DNA (GC compressions). Excellent separation has been achieved in this group at separation temperature as high as 60 °C. The gel composition and the optimum running conditions, such as the electric field, will need to be optimized for higher temperature.

Sample injection needs to be automated. Loading samples onto the gels is currently an extremely tedious and painstaking manual procedure. Moreover, the quality of loading significantly affects the band shapes and resolution and thus can complicate the automated sequence determination. The automated sample injection system may also help to improve precision.

The replacement of the capillaries and the subsequent realignment of the optical system is one of the most time consuming parts of the current setup. An in-situ gel refilling apparatus needs to be set up for the multi-capillary system. The use of elevated temperatures may make refilling easier due to the decrease in the viscosity of the gel at higher temperature.

As was demonstrated in this thesis, the multiple capillary system is useful in evaluating gel compositions for DNA sequencing. The system should be used in the future to study new separation media to improve the separation of sequencing samples.

References

1. T. A. Brown, *Gene cloning, 2nd edition*, Chapman & Hall
2. J. D. Watson, M. Gilman, J. Witkowski, M. T. Miller, *Recombinant DNA, 2nd edition*, W.H. Freeman and Company (1992)
3. T. Hunkapiller, R. J. Kaiser, B. F. Koop and L. Hood, *Science* 254, 59-67 (1991)
4. H. Swerdlow, K. Dew-Jager and R. F. Gesteland, *BioTechniques* 16 (4), 684-693 (1994)
5. A. S. Cohen, D. R. Najarian and B.L.Karger, *J. of Chromatography* 516, 49-60 (1990)
6. X. C. Huang, M. A. Quesada and R. A. Mathies, *Anal. Chem.* 64, 2149-2154 (1992)
7. T. Nishikawa and H. Kambara, *Electrophoresis* 12, 623-631 (1991)
8. N. J. Dovichi, *CRC Handbook of capillary Electrophoresis: A Practical Approach*, Chapter 14, 370-386
9. F. Franke and D. Keller, *Proceedings of Advances in DNA Sequencing Technology*, 1891, 78-84 (1993)
10. S. M. Lindsay and M. Phillip, *Genet. Anal. Tech. Appl.* 8, 8 (1991)
11. C. W. Chou, D. Dogruel, J. Krone, R Nelson, D. Schieltz and P. Willians, *DOE Human Genome Program, Conf. 9411116, Contractor-Grantee Workshop IV*, Nov. 13-17, P140(1994)
12. K. B. Jacobson, *Genomics*, 9, 51 (1991)
13. F. Sanger, S. Nicklen, A.R. Coulson, *Proc. Natl. Acad. Sci., U.S.A.* 74, 5456 (1977)
14. A. Maxam and W.Gilbert, *Proc. Natl. Acad. Sci., U.S.A.* 74, 560 (1977)
15. N. Arnheim, C. H. Levenson, *C&EN*, 36-47, October1 (1990)
16. R. A. Gibbs, *Anal.Chem.*, 62, 1202-1214 (1990)
17. B. L. Karger and F. Foret, *Capillary Electrophoresis Technology, PartI*, 3 (1993)

18. S. Hjertén, *J. Chromatography*, 270, 1 (1983)
19. V. Dolnik, *J. Microcolumn Separations*, 6 (4), 315-330 (1994)
20. X. C. Huang, S. G. Stuart, P. F. Bente III and T. M. Brennan, *J. of Chromatography*, 600, 289-295 (1992)
21. A. Guttman, A. S. Cohen, D. N. Heizer, and B. L. Karger, *Anal. Chem.*, 62, 137-141 (1990)
22. J. Macek, U. R. Tjaden and J. Van der Greef, *J. of Chromatography*, 545, 177-182 (1991)
23. S. L. Pentoney, Jr., K. D. Konrad, W. Kaye, *Electrophoresis*, 13, 467-474 (1992)
24. H. Swerdlow and R. Gesteland, *Nucleic Acids Res.*, 18, 1415 (1990)
25. H. Drossman, J. A. Luckey, A. J. Kostichka, J. D'Cunha, and L. M. Smith, *Anal. Chem.*, 62, 900 (1990)
26. J. A. Luckey, H. Drossman, A. J. Kostichka, D. A. Mead, J. D'Cunha, T. B. Norris, and L. M. Smith, *Nucleic Acids Res.*, 18, 4417 (1990)
27. A. E. Karger, J. M. Harris, and R. F. Gesteland, *Nucleic Acids Res.*, 19, 4955 (1991)
28. J. Liu, V. Dolnik, Y. Z. Hsieh, and M. Novotny, *Anal. Chem.*, 64, 1328-1336 (1992)
29. H. Swerdlow, J. Z. Zhang, D. Y. Chen, H. R. Harke, R. Grey, S. Wu, N. J. Dovochi, and C. Fuller, *Anal. Chem.*, 63, 2835 (1991)
30. D. Y. Chen, H. R. Harke and N. J. Dovichi, *Nucl. Acids. Res.* 20(18), 4873-4880 (1992)
31. N. J. Dovichi, J. Z. Zhang, in preperation
32. A. Guttman and H. E. Schwartz, *Anal. Chem.*, 67(13), 2279 (1995)
33. G. Schomburg, *Capillary Electrophoresis Technology*, 311 (1993)
34. B. L. Karger and A. S. Cohen, *U. S. Patent.*, 4, 865, 706 (1989)
35. A. Guttman, B. Wanders, and N. Cooke, *Anal. Chem.*, 64, 2348-2351 (1992)

36. T. Manabe, N. Chen, and S. Terabe, *Anal. Chem.*, 66, 4243-4252 (1994)
37. J. Sudor, F. Foret, P. Bocek, *Electrophoresis*, 12, 1056-1058 (1991)
38. N. Best, E. Arriaga, D. Y. Chen, and N. J. Dovichi, *Anal. Chem.*, 66, 4063-4067 (1994)
39. J. A. Luckey, L. M. Smith, *Electrophoresis*, 14, 492-501 (1993)
40. Slater, G.W., Mayer, P., Hubert, S.J., and Drouin, G. *Appl. Theoret. Electrophoresis*, 4, 71-79 (1994)
41. G. W. Slater, G. Drouin, *Electrophoresis*, 13, 574-582 (1992)
42. J. C. Giddings, E. Kucera, C. P. Russell, and M. N. Marcus, *J. Phys. Chem.*, 72, 4397-4408(1968)
43. O. J. Lumpkin, P. Dejaridin, and B. H. Zimm, *Biopolymers*, 24, 1573-1593 (1985)
44. P. Mayer, G. W. Slater, and G. Drouin, *Appl. Theoret. Electrophoresis*, 3, 147-155(1993)
45. T. A. Duke, J. L. Viovy, and A. N. Semenov, *Biopolymers*, 34, 239-247(1994)
46. C. Heller, T. A. Duke, and J. L. Viovy, *Biopolymers*, 34, 249-259 (1994)
47. T. Duke, and J. L. Viovy, *Phys. Rev. E*, 49, 2408-2416 (1994)
48. A. Tiselius, *Trans. Faraday Soc.*, 33, 524-531 (1937)
49. S. Wu and N. J. Dovichi, *J. of chromatography*, 480, 141-155 (1989)
50. Y. F. Chen, S. Wu, D.Y. Chen and N. J. Dovichi, *Anal. Chem.*, 62, 496-503 (1990)
51. J. Z. Zhang, *Ph.D. thesis*, University of Alberta (1994)
52. J. Z. Zhang, D. Y. Chen, H. R. Harke, and N. J. Dovichi, *Capillary Electrophoresis Technology*, 631-641 (1993)
53. C. A. Parker, *Photoluminescence of solutions*, New York, Elsevier, 411-426 (1968)
54. *CRC Handbook of Capillary Electrophoresis: A Practical Approach. Appendix I*, CRC. Press Inc., 617(1994)
55. M. J. Rocheleau, R. J. Grey, D. Y. Chen, H. R. Harke, N. J. Dovichi, *Electrophoresis*, 13, 484-486(1992)

56. T. L. Rapp, W. K. Kowalchuk, K. L. Davis, E. A. Todd, K-L Liu, and M. D. Morris, *Anal. Chem.*, 64, 2434-2437 (1992)
57. C. Gelfi, P. G. Righetti, *Electrophoresis*, 2, 213-219 (1981)
58. C. K. Mathews and K. E. Van Holde, *Biochemistry*, Benjamin/Cummings Publishing Company, Inc. 213 (1990)
59. M. C. Ruiz-Martinez, J. Berka, A. Belenkii, F. Foret, A.W. Miller, and B. L. Karger, *Anal. Chem.*, 65, 2851-2858 (1993)
60. Toyochi Tanaka, *Sci. Am.*, Jan. 124-140 (1981)
61. D. Tietz, and A. Chrambach, *Electrophoresis*, 13, 286-294 (1992)
62. M. J. Rocheleau, and N. J. Dovichi, *J. Microcolumn*, 4, 449-453 (1992)
63. J. A. Luckey and L. M. Smith, *Anal. Chem.*, 65, 2841-2850 (1993)
64. J. Yan, N. Best, J. Zhang, H. Ren, R. Jiang, J. Hou and N. J. Dovichi, *Electrophoresis*, in press
65. H. R. Harke, S. Bay, J. Zhang, M. J. Rocheleau and N. J. Dovichi, *J. of Chromatography*, 608, 143-150 (1992)
66. J. A. Luckey, T. B. Norris, and L. M. Smith, *J. Phys. Chem.*, 97, 3067-3075 (1993)
67. S. Hjerten, *Electrophoresis*, 11, 665-690 (1990)
68. J. C. Sternberg, In *Advances in Chromatography*, J. C. Giddings, R.A. Keller, Eds.; Marcel Dekker, New York, Vol.2, Chapter 6, P 226 (1966)
69. S. Terabe, K. Otsuka, and T. Ando, *Anal. Chem.*, 61, 251 (1989)
70. F. Foret, M. Deml, and P. Bocek, *J. Chromatography*, 452, 601 (1988)
71. E. Grushka, R. M. McCormick, *J. Chromatography*, 471, 421 (1989)
72. H. A. Fishman, N. M. Amudi, T. T. Lee, R. H. Scheller, R. N. Zare, *Anal. Chem.* 66, 2318 (1994)
73. J. A. Lux, H. F. Yin, G. Schomburg, *Chromatographia*, 30, 7 (1990)
74. S. V. Ermakov, M. Y. Zhukov, L. Capelli, and P. G. Righetti, *Anal. Chem.*, 66, 4034 (1994)

75. H. Swerdlow, K. E. Dew-Jager, K. Brady, R. Grey, N. J. Dovichi, and R. Gesteland, *Electrophoresis*, 13, 475-483 (1992)
76. D. Figeys, A. Renborg and N. J. Dovichi, *Electrophoresis*, 15, 1512-1517 (1994)
77. X. C. Tong and L. M. Smith, *Anal. Chem.*, 64, 2672-2677 (1992)
78. *Dynal production information, Prod. No. 612.02* (1991)
79. *United States Biochemical Corporation production information* (1991)
80. *Epicentre Technologies Production information, Catalog No. S36100* (1994)
81. S. Tabor and C. C. Richardson, *J. Of Biological Chemistry*, 265 (14), 8322-8328 (1990)
82. M. C. Giddings, R. L. Brumley Jr, M. Haker, and L. M. Smith, *Nucleic Acid Research*, 21(19) , 4530 (1993)
83. A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 36, 1627 (1964).
84. R. Betty and Gary Horlick, *Anal. Chem.* 49, 351 (1977).
85. J. Sudor and M. V. Novotny, *Anal. Chem.*, 66, 2446-2450 (1994)
86. Sharon Boots, *Anal. Chem.*, 61 (8), 553A (1989)