1998

An integrated multiple capillary array electrophoresis system for high-throughput DNA sequencing

Xiandan Lu
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

An integrated multiple capillary array electrophoresis system for high-throughput DNA sequencing

by

Xiandan Lu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Analytical Chemistry

Major Professor: Edward S. Yeung

Iowa State University

Ames, Iowa

1998

Graduate College

Iowa State University

This is to certify that the Doctoral dissertation of

Xiandan Lu

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

ABSTRACT

Capillary array electrophoresis system was chosen to perform DNA sequencing because of several advantages such as rapid heat dissipation, multiplexing capabilities, gel matrix filling simplicity, and the mature nature of the associated manufacturing technologies.

There are two major concerns for the multiple capillary systems. One concern is inter-capillary cross-talk and the other concern is excitation and detection efficiency. Cross-talk is eliminated through proper optical coupling, good focusing and immersing capillary array into index matching fluid. A side-entry excitation scheme with orthogonal detection was established for large capillary array. Two 100 capillary array formats were used for DNA sequencing. One format is cylindrical capillary with 150 μm o.d., 75 μm i.d and the other format is square capillary with 300 μm out edge and 75 μm inner edge. The capillary arrays were sandwiched between two fused silica plates with immersion solution in between, which formed a channel to guide the light. Different immersion fluids were chosen for different capillary formats. Immersion oil with a higher refractive index was chosen for the cylindrical capillary due to a higher trapping efficiency. Fused silica index matching liquid was chosen for the square capillary array because of the elimination of stray light produced during the laser beam transmission through the two dielectric media of capillary walls and immersion oil. The waveguide cell is designed in such a way so that the highest

excitation efficiency is obtained. Comparing the two capillary formats, the square capillary array shows higher excitation efficiency and is scalable up to hundreds of capillaries which matches the requirement of high throughput for large scale DNA sequencing.

This project is focused on the development of excitation and detection of DNA as well as performing DNA sequencing. The DNA denaturing condition and injection schemes are discussed for the cases of single and bundled capillaries. An individual sampling device was designed. The base-calling was performed for a capillary from the capillary array with the accuracy of 98%.

GENERAL INTRODUCTION

Dissertation Organization

This dissertation starts with the general introduction. This chapter is composed of two sections. The first section is the introduction of the Human Genome Project, in which background knowledge of genes, DNA, genomes and their relationships are explained in detail. The second section overviews the high throughput CE and parallel technologies for DNA sequencing, which gives you an idea what role CE may play in the Human Genome Project. The first chapter documents the optimization of focused line excitation and detection scheme for multiple capillary DNA sequencing. The second chapter is based on the paper published in Applied Spectroscopy. The third chapter documented the research on the waveguide excitation and detection system for high throughput DNA sequencing. References for each chapter follow the contents of the chapter in which they are cited. A general summary presents the comments on this work and future directions. Finally, the dissertation concludes with a list of the cited references for the general introduction.

Human Genome Project

Human has long been seeking the answers such as what blueprint controls the color of the skin, eyes, hair in order to ultimate understand themselves and other organisms.

More than 100 years ago [1], Gregor Mendel discovered cellular units (later named as genes) which control such inherited traits. The development of modern instrumental and analytical techniques used in molecular biology laboratories and the realization that genes consist of simple polymers of nucleotide sequences gave way to spectacular advances in understanding gene structures and functions. Not only is it possible to establish the nucleotide sequence of virtually any gene, it is possible to synthesize genes. This makes it feasible to obtain the ultimate description of genes and DNA, the giant molecule from which genes are formed.

In 1988, Human Genome Project [2] was initiated by the United States Congress. The goal of this project is to map and sequence the whole human genome (about 3 billion base pairs) within 15 years at the cost of 3 billion dollars. Mapping, sequencing, informatics, ethical, legal and social issues form the backbone of the whole genome project. Since this project involves big money, big consequences, and big controversies, it has the "big science" image. In order to understand the importance of the Human Genome Project, one must first understand genes, DNA and genomes.

Genes, DNA and Genomes. Deoxyribonucleic acid (DNA), which is recognized as the genetic material, is the single most important molecule in living

cells and contains all of the information that a cell needs to live and to propagate. It is a very long double-stranded chemical polymer (Figure 1). Each DNA strand is composed of four different units called nucleotides (Figure 2), that are linked end to end by phosphoester bonds to form a long chain. Each nucleotide has the following three components. 1) A cyclic five-carbon sugar; 2) A purine or pyrimidine base attached to the 1'-carbon atom of the sugar by an N-glycosylic bond. The bases shown in Figure 2 are the purines, adenine (A) and guanine (G), and the pyrimidines, cytosine (C) and thymine (T); 3) A phosphate attached to the 5' carbon of the sugar by a phosphoester linkage. This phosphate is responsible for the strong negative charge of DNA. The nucleotides are covalently linked by a second phosphoester bond that joins the 5'-phosphate of one nucleotide and the 3'-OH group of the adjacent nucleotide to form an extended chain of one strand of a DNA. The bases of one strand are hydrogen-bonded to those of the other strand to form purine-to-pyrimidine base pairs A, T and C, G. Base-pairing is one of the most important features of the DNA structure because it means that the base sequences of the two strands are complementary. Thus, an A in one strand is paired with a T in the other. Whereas, a C in one strand is paired with a G in the other. It is always true that [A]=[T] and [G]=[C], but there is no rule governing the ratio of total contents of G and C to those of A and T in a DNA molecule. In fact, there are enormous

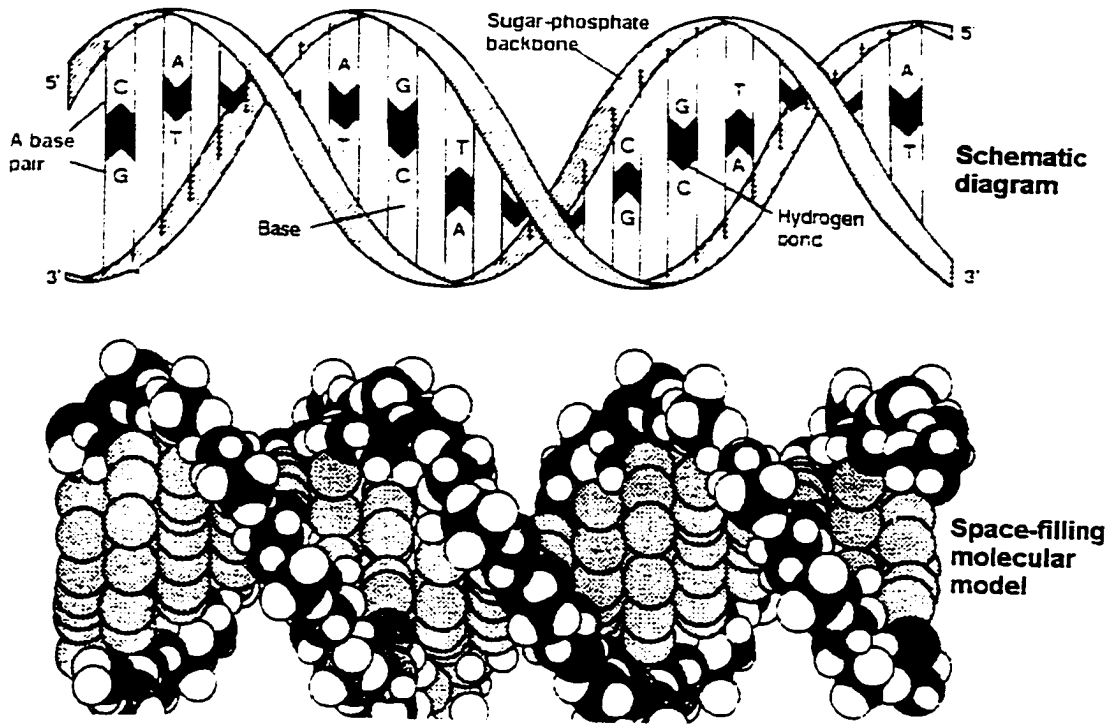
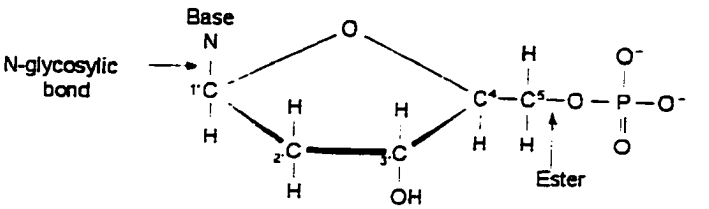
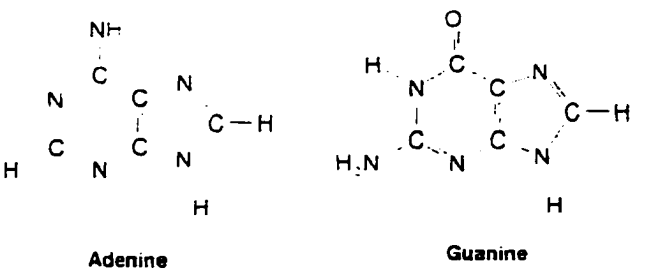
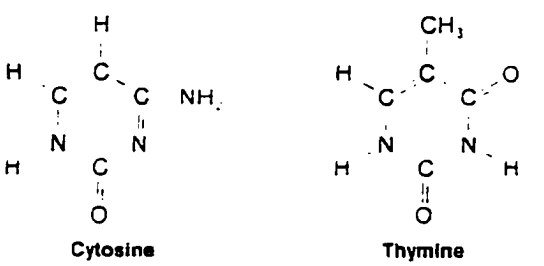


Figure 1. Schematic diagram of DNA molecule.



A





B

Figure 2. Structure of a deoxynucleotide and four bases found in deoxynucleic acids.

variations in this ratio. For most higher organisms, G+C content is near 50%. For the simple organisms, the percent of G+C varies widely from one genus to another. For example, in bacteria the content of G+C is 27% for the genome Clostridium and 76% for the genus Sarina. E. coli DNA is around 50% G+C [1].

One DNA molecule, which together with some proteins constitutes a chromosome, differs from one to another in sequence and in length. Each DNA molecule contains many genes, which are its functional units. Each gene contains thousands of the nucleotide pairs. All the genes are in a defined order along a DNA molecule. In general, each gene codes for the production of a single protein.

In simple organisms, the coding sequences of a gene are continuous, but in mammals, the coding segments in a gene (called exons) are separated from one another by non-coding segments (called introns). Many exons have been found to be part of a family of related coding sequences that are used in the construction of many different genes [1]. The human genome is composed of 24 kinds of DNA molecules, 24 types of chromosomes. Humans are diploid organisms, containing two sets of genetic information, one set inherited from the mother and one from the father. So each cell has 22 pairs of chromosomes (one member of each pair from each parent) and two sets of chromosomes (a x and a y chromosome in males and two x chromosomes in females) which totals 46 DNA molecules. Because human chromosomes exist in pairs that are almost identical (when they are not identical, the person is said to be heterozygous for the pair), the human genome is thus said to contain 3 billion nucleotide pairs instead of 6 billion nucleotide pairs.

the person is said to be heterozygous for the pair), the human genome is thus said to contain 3 billion nucleotide pairs instead of 6 billion nucleotide pairs.

Mapping. Although the sequences that we each inherit are unique (identical twins are exceptional), the genome order and structure of genes do not vary between individuals. Consequently, simple one dimensional maps can specify the genetic organization of the human, as well as other species.

Several technologies, including recombinant DNA (Figure 3), methods of cloning DNA molecules at specific sites, and pulsed-field gel electrophoresis (PFGE) [3-5] have all contributed to present mapping capabilities.

Basically, there are two kinds of mapping methods: genetic linkage mapping and physical mapping.

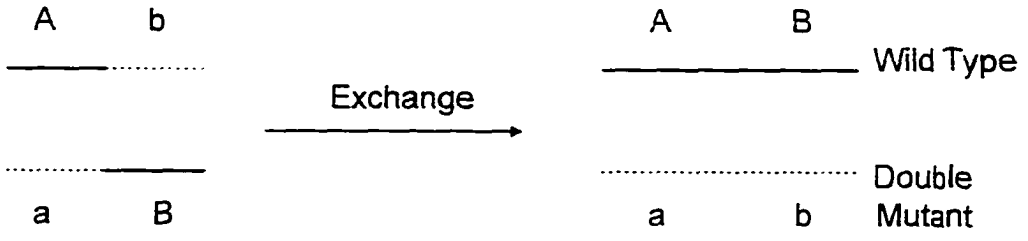


Figure 3. Genetic recombination: genes or mutations which are initially on different chromsomes onto a single chromsome. A,B are two wild type genes, and a,b are their mutants. Two genes or mutants are initially on two different chromsomes. A cut is made in both chromsomes, and the four fragments are then joined together to form two new combinations of genes.

Genetic linkage mapping is using the technology of genetic recombination (Figure 3) to study families and measure the frequency with which any two different traits or genes are co-inherited or linked. So the genetic linkage maps describe patterns of inheritance.

Physical maps are obtained from chemical measurements such as fluorescence in situ chromosome hybridization [6], sequence-tagged sites (STS) [7] by polymerase chain reaction (PCR) [8-9] made on the DNA molecules. Physical maps specify the distances between landmarks of cleavage sites of restriction enzymes along a chromosome. The distances are measured in nucleotides. So the map provides a direct description of a chromosomal DNA molecule.

Although the order of loci in physical and genetic maps will be the same, there is no easy way of converting from one to the other. A map unit of 1cM is used by genetic linkage maps to indicate a 1% chance of recombination. Recombination frequencies vary with chromosomal region and sex, as well as genome size. Genetic distance does not have a constant relationship with physical distance. But the generation of two kinds of maps have their own significance. The generation of genetic linkage maps can advance our understanding of basic human biology as well as to direct improvements in the diagnosis and treatment of many diseases. Two major forces drive the generation of physical maps. The first is to find genes. The second is to prepare materials ready for sequencing. Both maps share the common goal of placing information about human genes in a systematic linear order according to their relative positions along each chromosome.

The ultimate map of the human genome is the nucleotide sequence in which the identity and location of each of the 3 billion nucleotide pairs are known. Only such a sequence reveals all or nearly all the information in the human genome.

DNA Sequencing. The nucleotide sequence of a genome is the physical map at the highest level of resolution (one base pair).

There is an argument why the entire human genome has to be sequenced instead of the estimated 5% of useful protein-coding area. There are two reasons. Firstly, it will also take a huge effort to identify the 5% of the genome that actually encodes proteins before sequencing. This effort will not be less difficult than to sequence the entire genome. Secondly, only sequence cDNAs will not be able to reveal the range of sequences having important functions such as promoters and enhancers (which are elements that RNA polymerase to initiate gene transcription), gene regulation area and chromosome organization area. It is believed that the complete sequencing of the whole genome is the only approach that guarantees nothing significant will be missed [10]. Of course, the information-poor regions such as short-tandem repeated DNA sequences are not worth sequencing completely.

The DNA sequencing project is diverse. Any successful scheme will represent the combined efforts of multidisciplinary fields such as physics, analytical chemistry, organic chemistry and molecular biology. Currently, the sequencing technology is the bottleneck of the Human Genome Project. Tremendous efforts [11,12] are being put into each step of the DNA sequencing to achieve the 'dream' system which is described by Lloyd Smith (University of Wisconsin, Madison, WI,

USA) , a turnkey system, where DNA is put in at one end of the system, and finished separation, assembled DNA sequence, comes out at the other end. Clearly, in order to reach this goal, automation at all stages starting from clone picking to oligonucleotide synthesis, DNA extraction, template preparation, PCR, sequencing reactions, sample loading and detection, base-calling and cleaning, regeneration each part of the system to be ready for the next run. This dissertation is fully devoted to the instrumentation of high throughput capillary electrophoresis sequencer for DNA fragments or PCR sample separation and detection.

Consequences of the Human Genome Project. The success of the Human Genome Project will have profound impact on many areas in the biological sciences. It will free the researchers from having to track down and independently sequence the gene, allowing them to devote their efforts to more scientifically productive tasks from the beginning. Scientists will be able to find out the origins of human genetic diseases that are inherited as single-gene traits [13]. Eventually, multigenic traits will be resolved into single-gene components, and the environmental and genetic factors in complex diseases such as cancer, Alzhermer's disease, Huntington's disease will be disatangled. Great advances will be made in diagnosing many human diseases. Comparisons between members of gene families, between one gene family and another, and between corresponding genes from different organisms can be made. These comparisons will allow a much deeper understanding of the evolution of genes and regulatory elements [14]. Once all human genes are identified, mapped, and sequenced, it will be possible to define

the "transcript map" of an organism [14]. Such a map will describe not only the chromosomal locations and sequences of the genes but also their patterns of expression and functional importance during the development of an organism. The human genome sequence will have important applications in medicine, and provide the basis for gene therapy [15].

Sequencing Technologies and The Role of CE in The Human Genome Project

Sequencing Chemistry. DNA sequencing first became possible with the discovery of restriction enzymes and DNA polymerases around 1970. For the first time well-defined fragments could be derived from a larger molecule. The ability to sequence large DNA molecules via the manual radioactive methods became a reality when two independent techniques developed: Sanger's enzymatic sequencing [16] (Figure 4) and Maxam and Gilbert's [17] chemical sequencing (Figure 5).

The Sanger method [16] (Figure 4) utilizes the principle of chain-terminators, which are analogous to the four DNA nucleotides (dNTPs: A, G, C, T) except that terminators do not have 3'- OH group (ddNTPs). When the terminators are incorporated into DNA by DNA polymerase, synthesis of the growing DNA chain will be terminated. The Sanger reaction must be processed in four separate tubes in order to get all four nested DNA fragments which are terminated by four different ddNTPs. For example, in order to obtain the nested DNA fragments which are terminated by ddATP, a short oligonucleotide (primer) which is labeled by ^{32}p

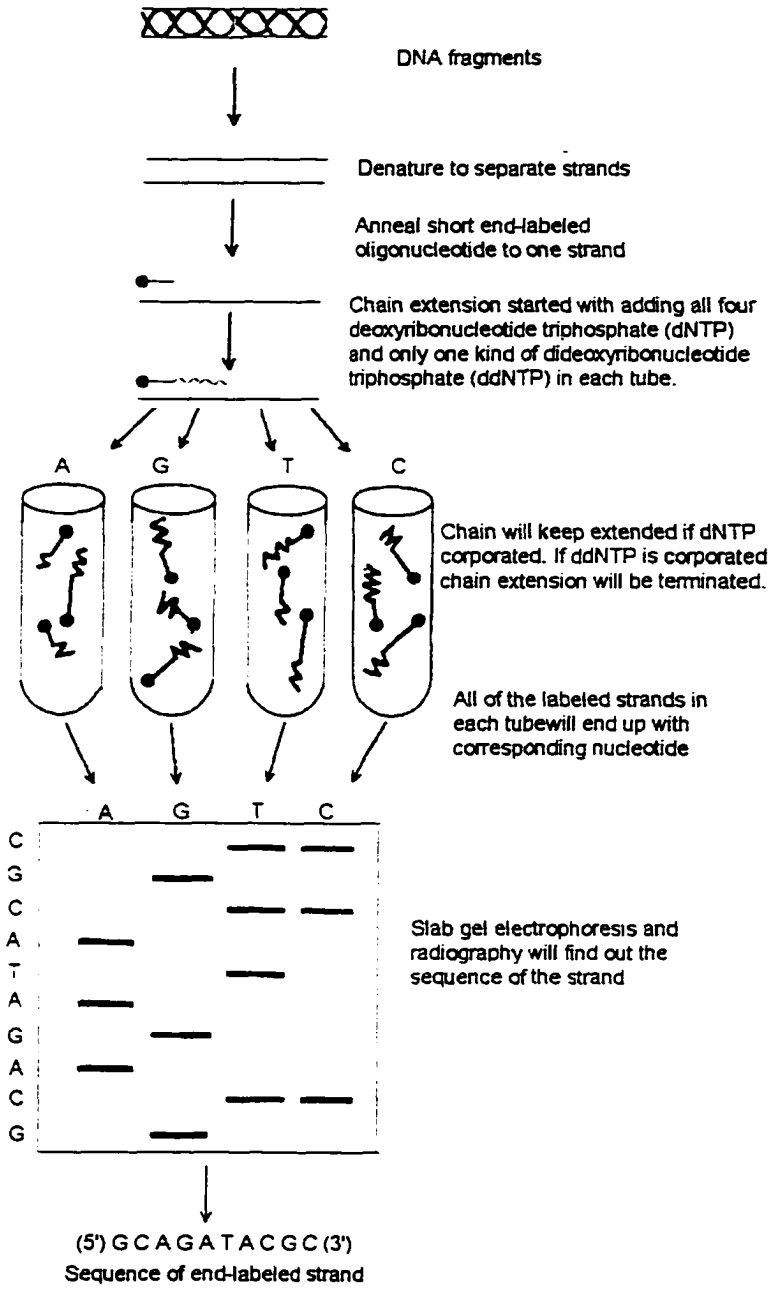


Figure 4. DNA sequencing by Sanger's Method

(normally,16-20 mer, shorter than that, melting temperature will be lower than the temperature required for annealing; longer than 20, not necessary) is first synthesized and annealed to the template of interest. The chain extension is carried out in the presence of all four types of dNTPs, polymerase, related reagents, and a small amount of ddATP. Once a ddATP is picked up, the chain extension will be stopped. Nested DNA fragments which are terminated by ddATP will be produced. The denatured DNA fragments are then loaded onto a gel to run electrophoresis. Bands corresponding to the different lengths of DNA fragments are produced and detected by autoradiography. If fragments of four different reactions run in four different lanes on one slab gel, all the sequences which are complementary to the template will be obtained.

The Maxam-Gilbert [17] (Figure 5) method uses chemicals that break the DNA chain at specific bases. The DNA molecule is labeled at one end with a radioactive tag. It is then chemically cleaved in such a way that only a few breaks are generated for any given base. As in the enzymatic sequencing methods, the DNA fragments are separated according to their sizes, and the sizes are correlated with the bases that are cleaved. To determine the entire sequence, four reactions are required. This method is generally more time consuming than the Sanger method. At present, the Sanger method dominates DNA sequencing applications.

Both methods generate mixtures of specific DNA fragments that are separated by polyacrylamide gel electrophoresis. When radioactively labeled DNA fragments are used, they are detected by exposing the gel to an X-ray film. The film,

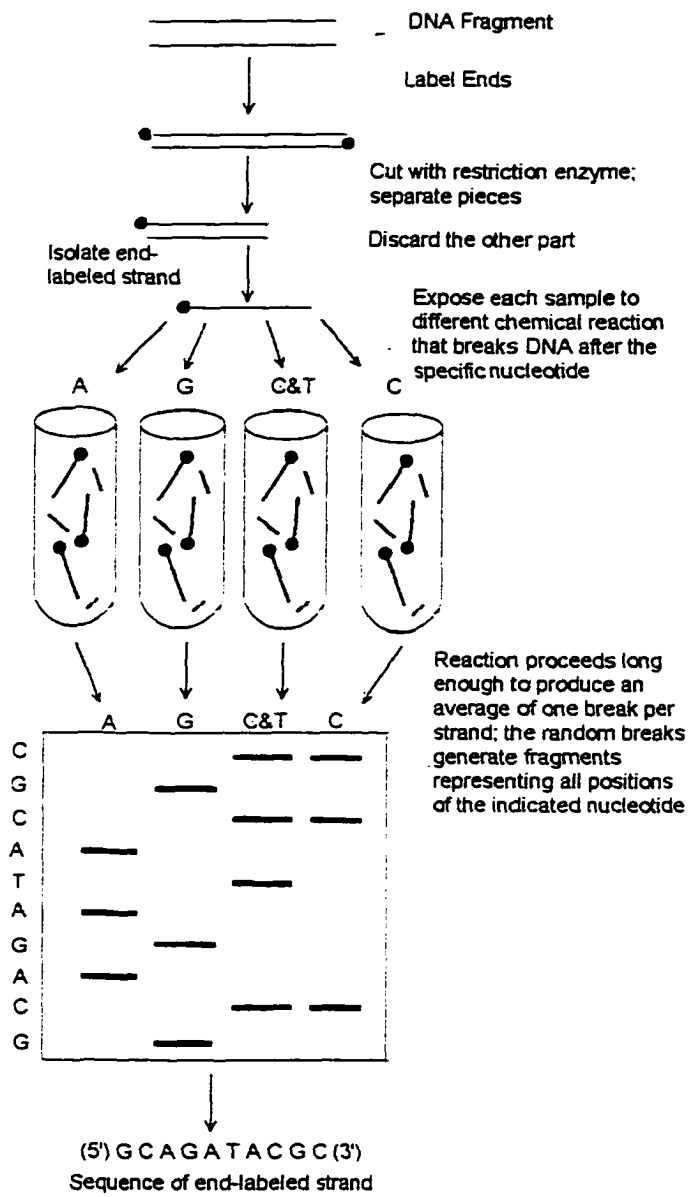


Figure 5. DNA sequencing by the chemical method.

which has an imprinted ladder distributed over four lanes representing the four nucleotides of DNA, must be interpreted by tracking the lanes together at one band a time. The data are then entered into a computer. No matter which method is used, the whole process is labor intensive, time consuming and expensive, requiring highly skilled scientists and the use of hazardous chemicals and unstable radioisotopes. These factors have inhibited serious consideration of very large scale sequencing projects such as the analysis of the human genome.

Another breakthrough in 1986 was the introduction of fluorescence-based dye into the Sanger reaction by Hood and co-workers [18]. This led to real time detection of sequenced fragments as they electrophoresed past a detector and an automatic sequencing system [19-21]. James M. Prober [22] and co-workers developed another set of dyes which is suited for excitation by an argon ion laser operating at 488 nm. Dyes suitable for high throughput genome-scale DNA sequencing should have the combined properties of physical stability, minimal overlapping emission spectra, high fluorescence intensity, no impairment in sequencing reactions, and uniform electrophoretic mobility. Commercial dyes for sequencing are fluorescein and rhodamine dyes (FAM, JOE, TAMRA, and ROX for dye labeled primer; R110, R6G, TAMRA and ROX for dye terminators). Although physically stable, these dyes do not meet the remaining criteria and require software corrections to produce optimal data [23]. Those dyes do not have same absorption coefficient under the excitation wavelength. Primers or terminators labeled with rhodamine dyes yield much weaker signals than those labeled with

fluorescein dyes. To get around this problem, more template DNA in the reactions involving rhodamine dye-labeled primers can be used. One also can use multiple lasers [24,25] to obtain even signals from four fragments. The other approach is to improve the spectroscopic properties of the dyes used for labeling. Recently, Richard N. Methies [26,27] developed a set of energy transfer dye to tag primer, which increased fluorescence signal 2- to 14-fold compared with conventional dyes for primers. Gichard A. Gibbs and co-workers [28] have identified another new family of novel fluorophores, dipyrromethеneboron difluoride (BODIPY), that show uniform electrophoretic properties under a variety of polyacrylamide gel conditions, narrower emission spectra and improved photostability relative to conventional fluorescein and rhodamine sequencing dyes. It is claimed that this set of dyes has overcome the problems presented by conventional and other energy transfer dye primers [26,27]. Single and double BODIPY dye primers were characterized in commercially available DNA sequencers and showed uniform electrophoretic mobility, high fluorescence intensities and improved spectral purity. The improved physical properties of BODIPY dye primers were demonstrated by direct base-calling from the unprocessed fluorescent signals. The high sensitivity of BODIPY dye primers requires at least 33% less reagent consumed per reaction than conventional dye primers, which should lower the costs and benefit large genome-sequencing efforts.

Sequencing Technologies. The sequencing technology reviewed here focuses on the post-reaction steps in the sequencing process: the electrophoretic

separation of the DNA fragments produced in the sequencing reactions and the interpretation of the separation patterns.

With the introduction of fluorescence-based dye into the Sanger reaction, a number of automated slab-gel DNA sequencer systems have become commercially available. These systems are: the Applied Biosystems, Inc. Model 370A, 373A, and 377A; the Du Pont Genesis™ 2000 DNA Analysis System; the Hitachi SQ3000 DNA Sequencer; and the Pharmacia Automated Laser Fluorescent (A.L.F.) DNA Sequencer™ [19-21]. These instruments combine modifications of standard polyacrylamide gel electrophoresis used in manual DNA sequencing with real-time detection of fluorescence-labeled DNA fragments. The advantages are duo to the automated data collection and determination of nucleotide sequence (base-calling) and to the increased number of samples that can be processed. The challenge is that real-time detection is not forgiving. Everything must be arranged appropriately to get successful runs. With autoradiography, the signal is accumulated over a controlled period of time. If the amount of isotope present in the gel is not sufficient to produce a detectable image on the initial film, a new film can be exposed to the gel for a longer period. But in the real-time detection system, the label is detected as it passes through a fixed-point detector. Therefore, only a period of time is available for signal capturing. If the signal is too low to be analyzed accurately, the data that is obtained may not be interpreted. This characteristic of the real-time detection has provided a challenge to the design of instruments as well as the development of reliable, reproducible protocols. The commercial systems are similar

in gel matrix (cross-linked polyacrymide) for DNA fragments separation and air-cooled argon lasers to excite the dyes. But the detection systems and the method used to identify the base are different.

The Applied Biosystems Inc. (ABI) models 370A/373A have similar detection systems (Figure 6) with the newest model 377A being different (Figure 7). The system originally described by Smith et al. [18] used four fluorescent dyes (a fluorescein, tetramethylrhodamine, 4-chloro-7-nitrobenzo-2-oxa-1-diazole (NBD), and Texas red) that were attached to the 5' end of the sequencing primers. ABI uses a different set of dyes. For the Taq dye deoxy terminators, they use R110, R6G, Tamra, and Rox which are all rhodamine dyes. For the dye primers, they use FAM, JOE, Tamera, and ROX which are fluorescein and rhodamine dyes. Due to the big differences in the excitation maxim of the dyes, the system uses two wavelengths excitation (488nm and 514nm). The fluorescence signals are detected by a photomultiplier tube (PMT) for models 370A/373A. A set of four filter wheel is used to discriminate the wavelengths of light that are being emitted. The laser focusing optics, filter wheel, and the PMT are mounted on a translator that travels across the face of slab gel. Since the system uses a single PMT, the gel must be scanned four times for each complete data point. The model 377A uses the scanning spectrograph (diffraction grating) and a charge coupled device (CCD) system instead of PMT. This allows detection of multiple fluorescent emissions with a single pass of the scanner.

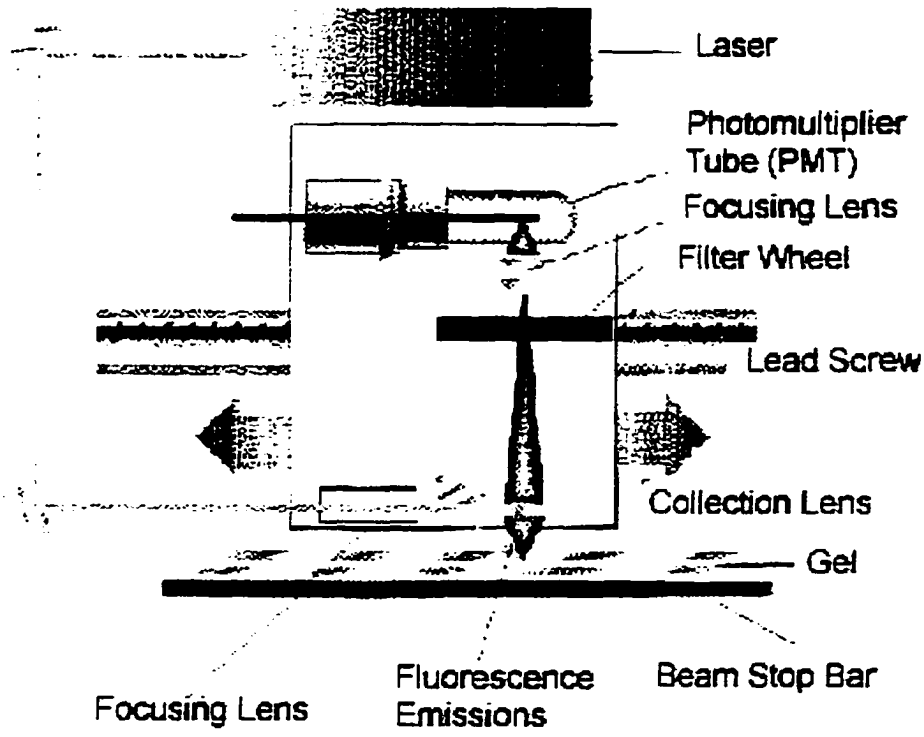


Figure 6. Diagram of the Laser/Scanner Assembly (Top View).

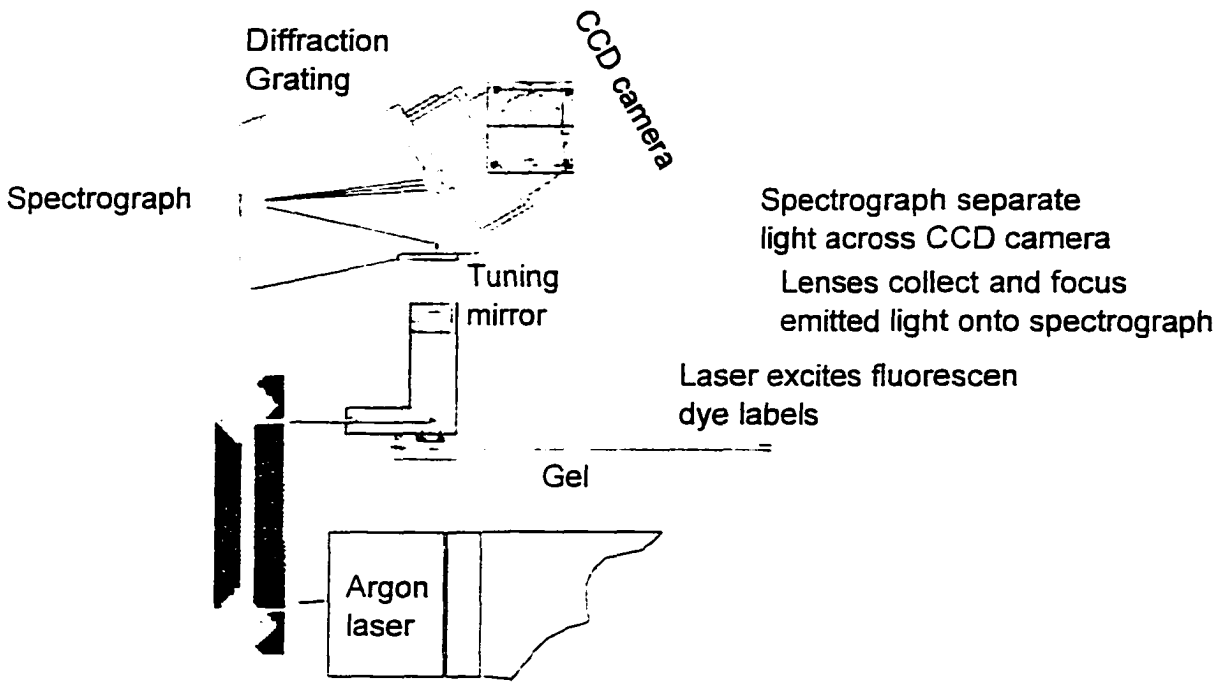


Figure 7. The ABI Prism 377 scanning/detection system.

Du Pont's Genesis 2000 DNA Analysis System (Figure 8) uses four fluorescent labels for primer. They developed a family of succinyl fluoresceins [22] that differ from each other based on the methyl substitutions shift the emission maxim of the dyes. The emission maximums range from 505 to 526nm. All four dyes can be efficiently excited by a single 488nm laser line. In this system, a mirror mounted on a high resolution stepper motor sweeps the laser beam across the gel face, and the signal from the fluorophore is detected by two PMTs with different filters.

The Hitachi SQ3000 DNA Sequencer [29] (Figure 9) uses a single color labeling (fluorescein isothioyanate) and four lanes electrophoresis. The 488nm argon ion laser is used to excite the bands from the side of the thin gel, and the emitted light is detected with a vidicon camera instead of PMTs. With the side-on excitation, there are no moving parts. And a larger band is excited compared to the scanning system. Since this is a single dye system, four lanes are needed for each sample. Accurate base assignment relies on the equivalent rates of electrophoresis between the four lanes. But the rate of electrophoresis varies from lane to lane due to differential rates of heat dissipation, differential amounts of salt in the sample, and the non-uniformity within the gel matrix.

The Pharmacia A.L.F. system (Figure 10) is based on the European Molecular Biology Lab's system [30]. This system also uses a side-on laser, a single-dye tag, and therefore four lanes per sample. The difference from other systems is that the emitted light is detected by a linear array of solid state

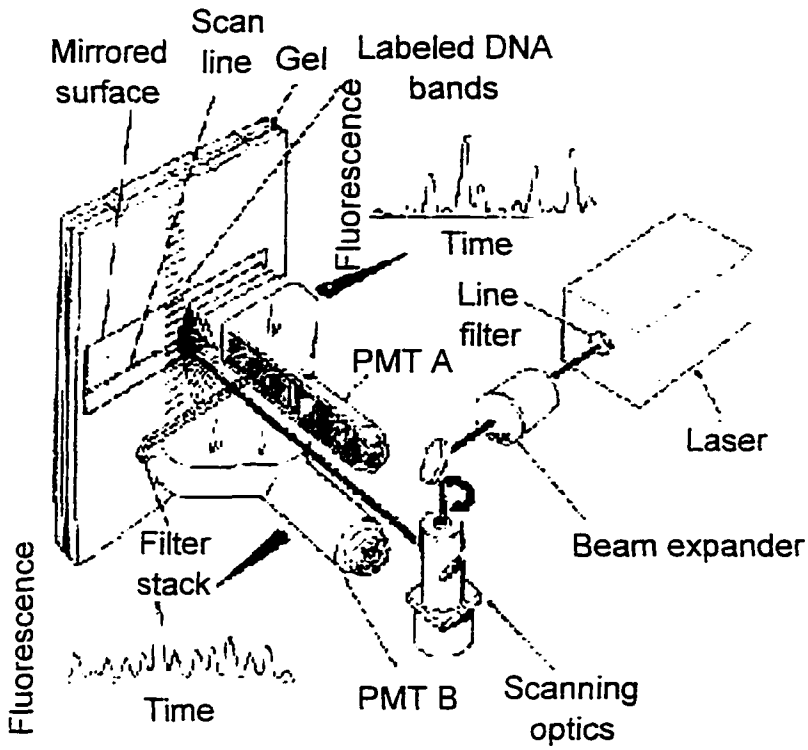


Figure 8. Du Pont's Genesis 2000 DNA Excitation and Detection System.

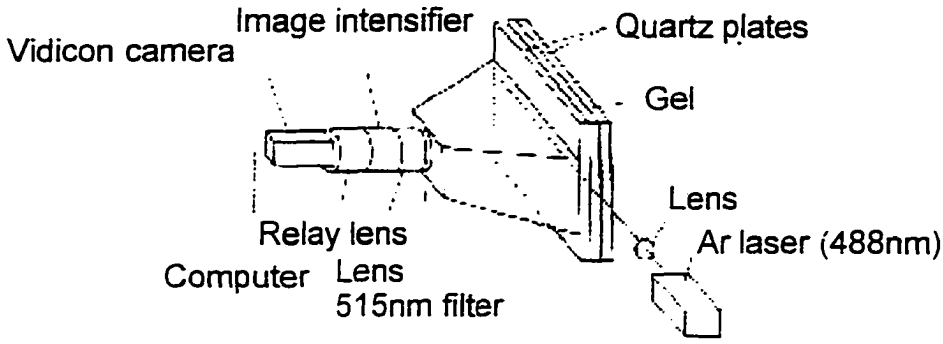


Figure 9. Hitachi SQ3000 DNA Sequencer.

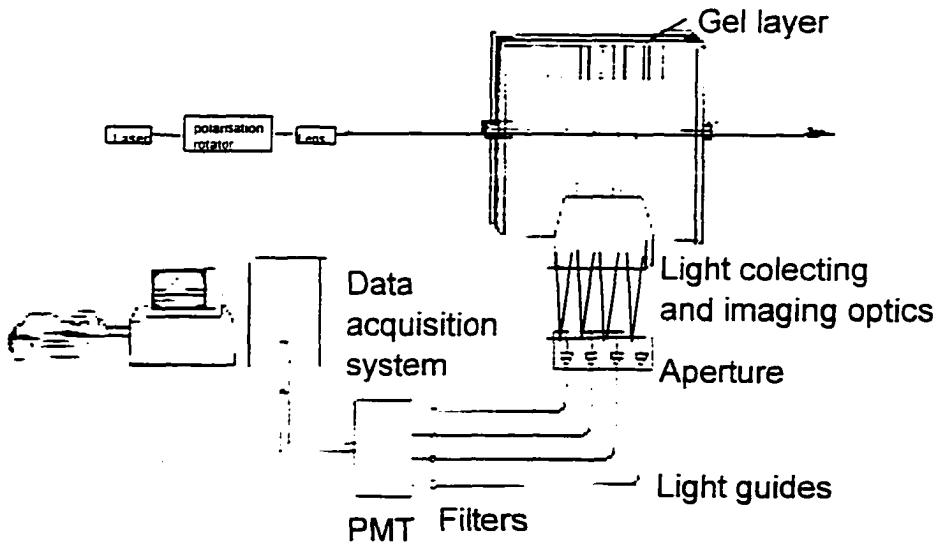


Figure 10. The Pharmacia A.L.F. System.

photodiodes, one for each lane. As in the Hatachi SQ3000 system, there are no moving parts. However, there is a chance for DNA band to miss the detector due to the distorted path of electrophoresis.

Currently, the ABI 373 slab gel sequencers are widely used in the high-throughput automated DNA sequencing and analysis core facilities [31]. The genome center at Lawrence Berkeley National Laboratory sequences about 2 million base pairs per year. At this rate, it would take hundreds of years to sequence one human genome. It is true that hundreds, thousands of machines can be used to improve the high throughput, but the cost (estimates run anywhere from $1-$6 a base pair) will never be reduced. Recently, ABI placed Model 377 on the market. The ABI PRISM ™ 377 DNA Sequencer can run up to 36 lanes instead of the 24 lanes used by the Model 373 by utilizing a diffraction grating and CCD camera in scanning mode to detect and display laser-excited fluorescence emitted from dye-labeled DNA molecules being electrophoresed in slab gels. With one set of ABI 377 automated sequencer, approximately 54,000 base pairs of sequence per day can be achieved (from the statistic data of Iowa State University Nucleic Acid Facility Center), but this is still not good enough for large scale sequencing.

Commercially available instruments will not do the job. Faster and cheaper DNA-sequencing techniques must be developed to make large-scale sequencing projects possible and to provide a routine tool for diagnostic medicine. Present

methods are about 100-1000 times too slow and at least 100 times too expensive for routine use. This is where the challenge of the Human Genome Project lies.

Several new technologies are emerging. Basically, there are two kinds. One is called conventional technology, which is gel electrophoresis based sequencing. The other is several radically new technologies that could change things dramatically. Unconventional technologies will be reviewed first.

Single-Molecule Detection Based Sequencing [32,33]. A strand of DNA is synthesized complementary to the template DNA using raw materials of nucleoside triphosphates (dNTPs) which are bound to dye molecules. The selected DNA strand is suspended in a gently flowing stream of water, which carries the nucleotide to a detector after being sequentially cleaved by an exonuclease. The nucleotides are removed in order and if they can pass through the detector without scrambling the order, and the detector is sensitive enough to detect and identify four different kinds of dye molecules, and the sequence is obtained immediately. In order to make this scheme a reality, several technological advances must be first established. The first is to find four kinds of dyes to tag dNTPs. This set of dyes must guarantee the complete synthesis of the DNA strand, can be identified by the detector. The second is to find "a handle" which can hold DNA strand in the flowing water. The third is to find a way to pick out the bead which carries one DNA molecule. The fourth is to find an enzyme which can cleave bases one by one in an efficient way (complete cleavage?). The final step is to find a flow system which can carry the labeled bases one by one through a laser beam, where each base will be

identified by the laser-induced fluorescence of its dye label. So far, none of the prototypes have been set up for the above requirements.

Sequencing by Hybridization (SBH) [34-37]. SBH is based on the fundamental chemistry of DNA itself. Because of DNA base pairing specificity, a large set of synthetic oligonucleotides of specified length can be used to "read" DNA sequences by identifying overlapping blocks of oligonucleotides that form perfect duplexes with the target DNA. This method is expected to be much faster than the conventional sequencing because there is not an electrophoretic separation step. Two formats have been proposed for this method. One is that unknown DNA samples are attached to a support and hybridized successively with every possible oligonucleotide probe. The other is arraying all the possible oligomer probes on a support and hybridizing the target DNA to the array. In principle, it is very simple and attractive. In practice, several technical issues arise. No matter which format, highly discriminative hybridization is required to distinguish between a perfect DNA fragment and an oligonucleotide complementarity and one or more nucleotide mismatches. Generally, mismatches can be decreased by using probes of a different length and adjusting hybridization and washing conditions. But because base pairing strength is much higher for G and C (three hydrogen bonds) than for A and T (two hydrogen bonds), a mismatch in G, C-rich region of the DNA will still give a stronger signal than a true AT match. In order to read simple sequences precisely, a large array of probes must be synthesized. In the case of an 8-mer probe, this could require as many as 65,000 kinds of probes. For 10-mer probes,

the number rises to more than a million. The cost per base pair is $0.75 at Iowa State University's Nucleic Acid Facility Center. It will cost about $1 million for synthesis. Applying all those probes to the silicon substrate without errors is another issue. At the same time, the large computation in order to find the overlap sequences is necessary. Another characteristic of DNA is that there are lots of repetitive DNA sequence regions in each chromosome fatal SBH to be able to serve as a primary sequencing tool for sequencing. Instead, it can serve as a DNA diagnostic tool, such as mutation detection or as a tool for sequencing of short, non-repetitive DNA fragments used in identifying genes, repeats and along with primary sequence data derived by other methods to provide a rapid means of confirming and correcting sequence data.

Sequencing By Mass Spectrometry. During the past decade, tremendous effort has been made on applying mass spectrometry to biopolymers [38]. However, there are two major difficulties: developing a reliable method to deliver these nonvolatile biopolymers to the mass spectrometer's ionization zone and achieving ion production without significant fragmentation in either the desorption or the ionization steps. Several ionization methods such as fast atom bombardment [39], thermospray ionization [40] electrospray ionization (ESI) [41], laser desorption/ionization [42] and matrix-assisted laser desorption/ionization (MALDI) [43] have been developed. Electrospray ionization and MALDI have been considered to be the best choices for biopolymers. Recently, ESI [44,45] and MALDI [46-49] coupled with MS [46-49], FTMS [44] and MS^n [45] have provided

limited sequence information for small synthesized DNA fragments. It turns out that resolution of the oligonucleotides is always lower than that of proteins of similar masses [50]. This mainly comes from the peak broadening due to alkali cations in samples and nucleotide base losses during the evaporation and ionization process [51]. So far, the technique still has to be further improved at least an order of magnitude in both mass range and sensitivity in order to play a key role in DNA sequencing.

CE Technology is Seeking a Practical Role in DNA Sequencing. Even though gel-based techniques could never achieve the dramatic speed improvement as those of unconventional methods, it is the most practical and mature technique for DNA sequencing thus far.

As mentioned above, Applied Biosystems Model 377, which is based on slab-gel electrophoresis, is widely used for DNA sequencing. Each machine runs 36 lanes simultaneously to produce sequencing rates of 5,400 bases/h/slab. Increased sequencing rates can be produced either by running electrophoresis at higher electric field strength or by running more lanes simultaneously. Unfortunately, neither can be really done on the slab gel. The heat produced by running electrophoresis at higher electric field strength than 50 v/cm on conventional 200-400 μm thick gels will affect the separation due to distortion of the gel band or smearing the gel band. Several labs are trying to push this technique further by using ultra-thin slab gel [52] and developing coupling techniques for the sample

loading [53]. However, the difficulties of pouring gel uniformly over a large area, uniform cooling and sample loading automation prevents this technique from high multiplexing and automation [54].

Capillary gel electrophoresis, which is based on the same separation principle as conventional slab gel electrophoresis, became an attractive alternative to slab gel for DNA sequencing. All the techniques which have been proven in the slab gel method such as gel matrix (cross linked polyacrylamide), buffer, separation mechanism, sequencing chemistry and dye tagging chemistry can be directly transferred to CE without further development. The superior heat dissipation property of CE due to large surface to volume ratio allows use very high electric field strengths. A roughly 25 times increase in sequencing speed compared to slab gel has been achieved by using 400 V/cm [55] for one capillary. The format of the capillary is perfect for the large array. Typical capillaries for DNA separations are 50-75 μm i. d. and 150-360 μm o. d. If capillaries are packed side by side [56] without spacing in between. A 100 capillary array will only cover about 1.5 cm. A 1000 capillary array will cover 15 cm, which is still manageable. At the same time, the use of a new gel matrix, such as PEO [57] and non-cross-linked polyacrylamide [58,59] has overcome the original problems encountered in crossed-linked gels. In addition, this new gel can be made quickly and reproducibly. Instruments incorporating CE are being developed.

In high throughput DNA sequencing, the challenge lies in the ability to efficiently couple laser power to each capillary. This dissertation is devoted to this

topic. It is important to prove that the whole system including excitation and detection scheme, gel concentration, gel viscosity, electric field for separation can actually function well for multiple capillary sequencing. What works best for a single capillary may not be fit for a large-array. The optimization of all these parameters are addressed in the following chapters to lead to an integrated successful approach for DNA sequencing.

CHAPTER 1

INVESTIGATION OF FOCUSED LINE EXCITATION AND DETECTION SCHEME FOR MULTIPLE CAPILLARY DNA SEQUENCING

Introduction

Ueno and Yeung [1] investigated different on-column excitation schemes for large capillary arrays. This investigations finally led to a two-color two-window excitation and detection scheme. Although actual DNA sequencing had not been done with this system, the potential for 100 capillary array has been demonstrated through detection restriction enzyme fragments. The 100 capillaries were set into 100 triangle-shaped grooves machined on an aluminum block. Each of the two laser beams (488 nm and 514 nm) was expanded and then focused into a fine line with a cylindrical lens to cut across the capillary array for excitation. The system can be expanded to accommodate larger number of capillaries such as 1000 by using higher laser power, a larger format camera lens, and a faster camera. The work demonstrated the major features of a truly multiplexed CE system without moving parts. However, before it can be adapted to the high throughput DNA sequencing, many major steps need to be established, proven and optimized. This chapter documents the optimization of this system in the following aspects: First, optical

coupling arrangement of a capillary onto CCD camera. The right optical coupling arrangement is important to decrease the cross-talks between capillaries.

Second, experiments are undertaken to improve the excitation efficiency. Third, an individual sampling device is designed.

Experimental Section

An argon ion laser (Uniphase, San Jose, CA, Mode 2213-150ML or Coherent Innova 90) was used for excitation. The laser beams (488 nm and 514 nm) were expanded by a circular lens or a beam expander (Oriel Corp., Stratford, CT) and focused onto the capillary array with 10-cm f.l. planar-convex cylindrical lens (Melles Griot Model 01LCP155, f.l. 10 cm). The fluorescence signals were detected by a charge coupled device (CCD) (AT-200 serial Photometrics CCD) or charge injection device (CID technologies Inc., Liverpool, New York) with a Nikon F1.4/28 mm wide angle camera lens. Round fused silica capillaries of 150 μm o.d., 75 μm i.d. with polyimide coating (Polymicro Technologies, Phoenix, AZ) were used for all the experiments. For the cross-talk experiment, short capillaries were used. All the array capillaries were packed by 3M double-sided tape and fixed onto a microscope slide covered on the top of the mount. A $1x10^{-3}$M fluorescein (Sigma Chemical Co., St. Louis, MO) was used as the stock solution. The samples were prepared from the dilution of the stock solution into pH 10.0 3 mM Na_2HPO_4 (running buffer) to the desired concentrations. For illumination uniformity experiment, the capillaries were bundled into 4 groups with 25 capillaries each. Sequentially flush through buffer and fluorescein from the last bundle to the first bundle. Signals from the array were

obtained by subtract blank background counts from the flushing fluorescein counts. For DNA sequencing experiments, 65 cm total with 50 cm efficient length capillary was used. Before filling a capillary with gel, the capillary was flushed with methanol. The gel used for separation is the mixture of 1.5% 8,000,000 MW polyethylene oxide (PEO) and 1.4% 600,000 MW PEO in 1xTBE with 3.5 M urea. Before DNA sample was injected, the capillary pre-ran 10 minutes. DNA sequencing samples (PGEM/U) were prepared from the Sanger reaction according to the standard protocols (ABI, DyeDeoxy Terminators and cycle sequencing with Taq polymerase) in the DNA Sequencing Facility of Iowa State University. Each of two tubes of dried PGEM/U sample was resuspended in 4 μL of 5:1 (v/v) of formamide and EDTA (10 mM) solution. Combine two tubes into one. The sample was heated at 95 °C for 4 minutes to denature the DNA and put into a -10°C freezing well to cool it down rapidly. Sample injection was performed at -5 kV for 1 minute. Electrophoresis was running at -9.7 kV via a high voltage power supply (Spellman, Plainview, NY). DNA sequencing was run on five capillary array which covers 15 mm range (the width of 100 capillary array) with the laser output 300 mW. The 514 nm laser wavelength (Innova 90, Coherent) was chosen for excitation. Figure 1 shows the schematic diagram of the experimental setup. A 514 nm Raman-edge long pass filter (Kaiser Optical Systems, Ann Arbor, MI) was attached onto the camera lens (Nikon 28 mm/1.4, filter size 72 mm diameter). After a run was completed, the gel matrix was pushed out by compressed N_2 gas at 350 psi. The capillary can be regenerated by circulation of methanol at high pressure (thousand of psi).

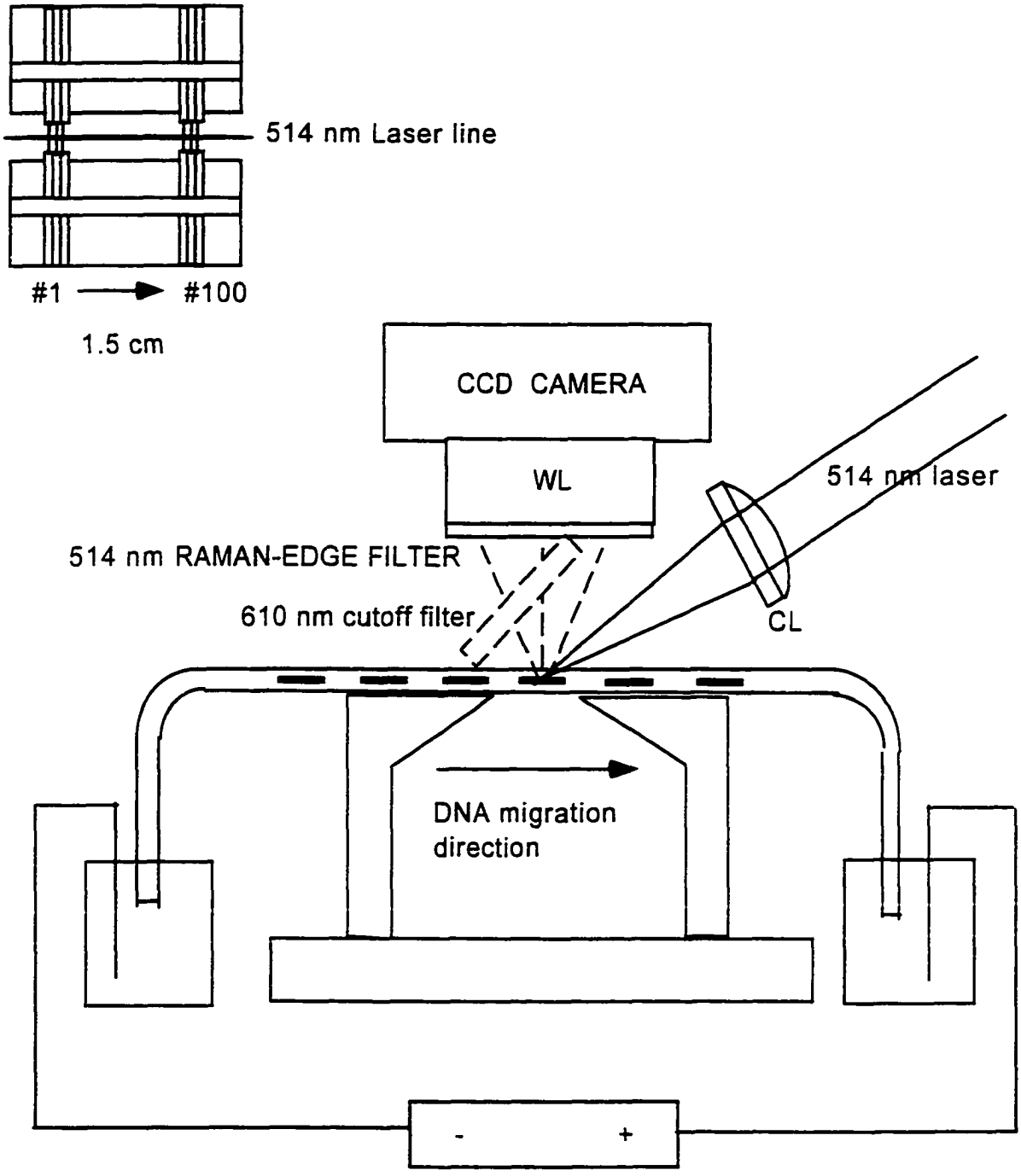


Figure 1. Experimental set-up for focused line excitation and detection
WL, wide-angle lens; CL, cylindrical lens.

Ann Arbor, MI) was attached onto the camera lens (Nikon 28 mm/1.4, filter size 72 mm diameter). After a run was completed, the gel matrix was pushed out by compressed N_2 gas at 350 psi. The capillary can be regenerated by circulation of methanol at high pressure (thousand of psi).

Results and Discussions

Cross-Talk. In multiple capillary system, decrease in cross-talk is critical for the effort of the instrumentation. Cross-talk is the false peaks which are observed in nearby capillaries. It must be reduced or eliminated to a level which will not affect the base calling scheme before DNA sequencing application can be addressed. Tayor and Yeung have observed the cross-talk problem in capillary array system [2]. They tried to reduce the problem by putting the spacers, 150 μm o.d. capillaries coated with black ink, between each of the separation capillaries. Ueno and Yeung [1] adapted this method to his system. Instead of capillaries, the 100 capillaries were set into the triangle-shaped grooves which were machined on the surface of the aluminum block. Because of irregularities of the grooves on the aluminum block, however, the cross-talk is still about 10%. Using deeper grooves to isolate the capillaries has been proposed. However, this requires a larger imaging area, results in less efficient utilization of the laser beam and requires a larger memory for data storage. Binning the CCD camera decreased the data size, but the correct optical coupling arrangement has to be established, or serious cross-talk will result [1].

Ueno tried 2:1 binning to make each capillary span 1.83 pixels which is the main reason that 10% cross-talk was observed in this compared with 1% in the work of Taylor and Yeung.

Cross-talk results from a number of factors including signal light refraction from the cylindrical walls of adjacent capillaries, poor image quality, and scatter from optics. Methods for minimizing cross-talk will be discussed below.

Cross-talk produced from the capillary wall refraction can be solved by using right optical coupling arrangement or immersing the capillary array into a refractive index matching fluid [3]. The latter method will be discussed in the following chapter. As for the focused line excitation scheme, from the excitation point of view, there is no reason to put the capillary array into the fluid. Putting the capillary array in air is the simplest case with the lowest background. Elimination of refraction is not realistic. Optical arrangement can be optimized to reduce the cross-talk in air.

Packing capillary side by side has been tested with the optical coupling arrangement of 2:1 and 3:1. For the 2:1 arrangement (Figure 2), core of the capillary was focused onto one pixel. The walls of two adjacent capillaries were focused onto the other pixel of which the signal was void, so the refracted fluorescence from adjacent capillaries were spatially blotted. In order for the core of the capillary (which is 75 μm) to cover exactly one pixel, the packing quality and the focus alignment is very important. The capillary array was mounted onto a one dimensional stage with the adjustment orthogonal to the capillary array. Nonetheless, about 6% of cross-

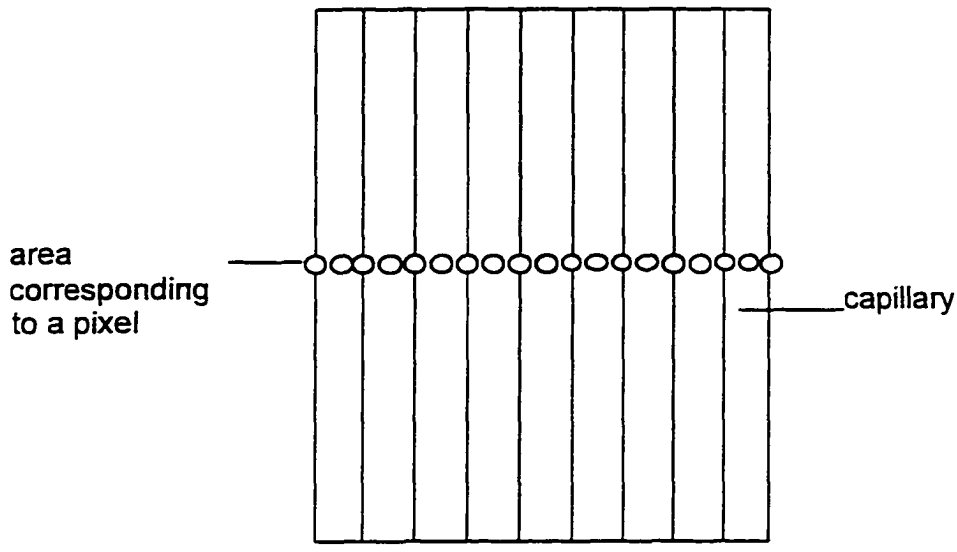
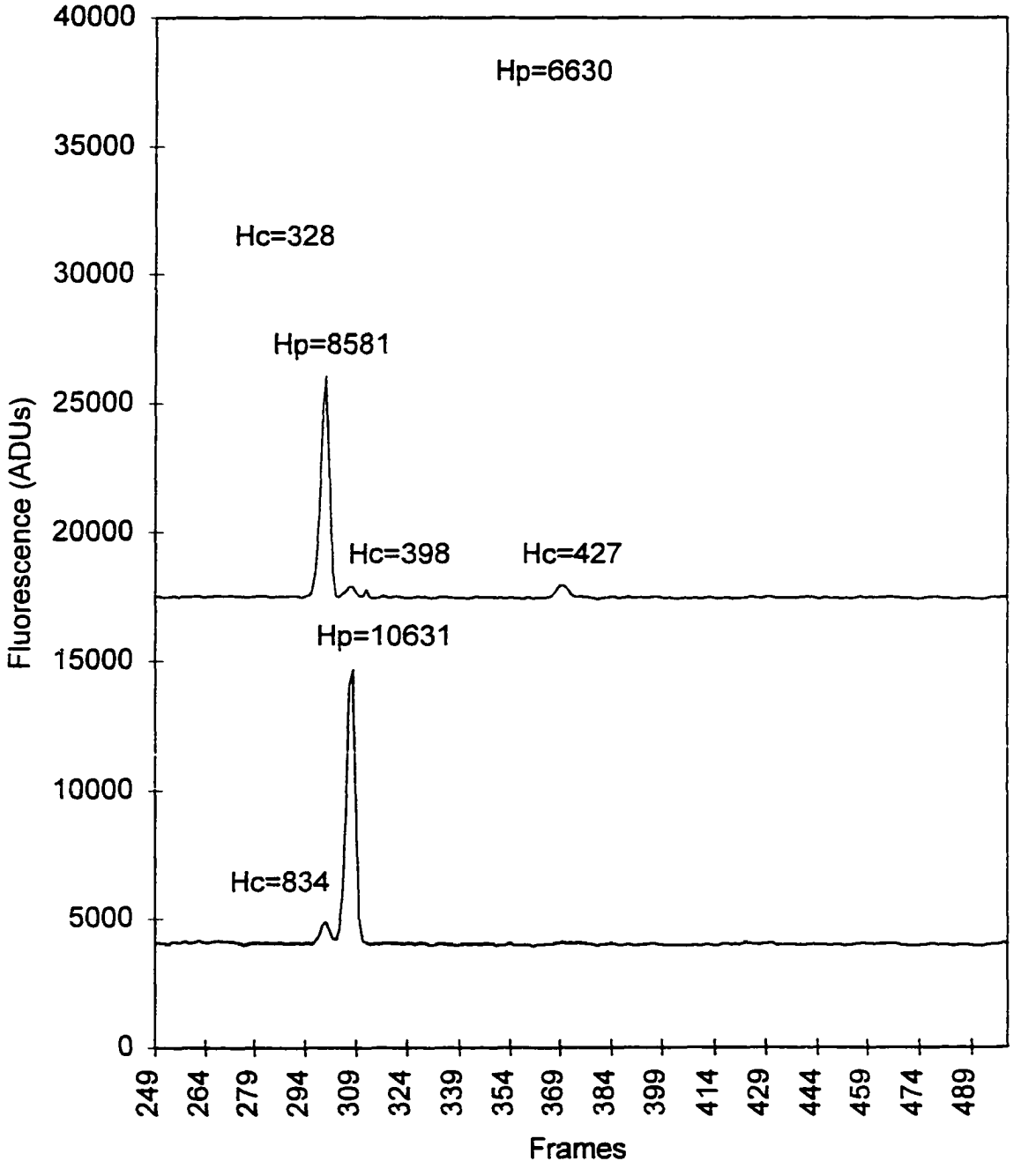


Figure 2. Overhead view of the 2:1 optical coupling arrangement.

Capillary, 150 μm o.d.; 75 μm i.d.

Figure 3. Cross-talk in air with 2:1 optical coupling arrangement. The camera aperture was set at 1.4. A piece of 550 nm glass filter was attached in front of CCD camera to cut off the laser light. Hp is the height of the fluorescein peak detected in a capillary; Hc is the cross-talk signal observed in adjacent capillaries.


40000
35000
30000
25000
20000
15000
10000
5000
0
Fluorescence (ADUs)
Hp=6630
Hc=328
Hp=8581
Hc=398
Hc=427
Hp=10631
Hc=834
249
264
279
294
309
324
339
354
369
384
399
414
429
444
459
474
489
Frames

talk was observed (Figure 3). Furthermore, the packing quality requirement is very high in order for the core of the capillary (75 μm) cover one pixel. Even though the capillary array was mounted onto a one dimensional stage with the adjustment orthogonal to the capillary array, the alignment is still not perfect. This is proven by different levels of cross-talks in the center capillary from two nearby capillaries.

In another 3:1 optical arrangement (Figure 4), each capillary covers 3 pixels so that 50 μm instead of 75 μm accounts for one pixel. It gives greater flexibility of alignment and makes focusing a lot easier. Even though one-third of the signal will go to void pixels, putting the CCD camera closer to the capillary array will increase the light collection solid angle and collection efficiency. When each capillary covers 3 pixels, the cross-talk goes down to the average of 1.3% with a maximum of 1.9% (Figure 5). The 3:1 optical coupling arrangement is right for the capillary array in air.

Unclear focusing of the capillary array is another reason to cause the cross-talk. There are two reasons for bad focusing: objective reason and subjective reason. The quality of the focusing lens accounts for the objective reason. In our application, either Canon F1.4/24 mm or Nikon 1.4/28 mm wide angle lens is used. The largest aperture can be obtained by using f-number 1.4. It gives the biggest solid angle and the highest light collection efficiency. At the same time, it requires a fast shutter. The field depth is shallow. It is not easy to get clear focusing for the whole image range. You can change focusing to change the clearest focusing parts within the image range. For the large array capillary detection, clear focusing in the

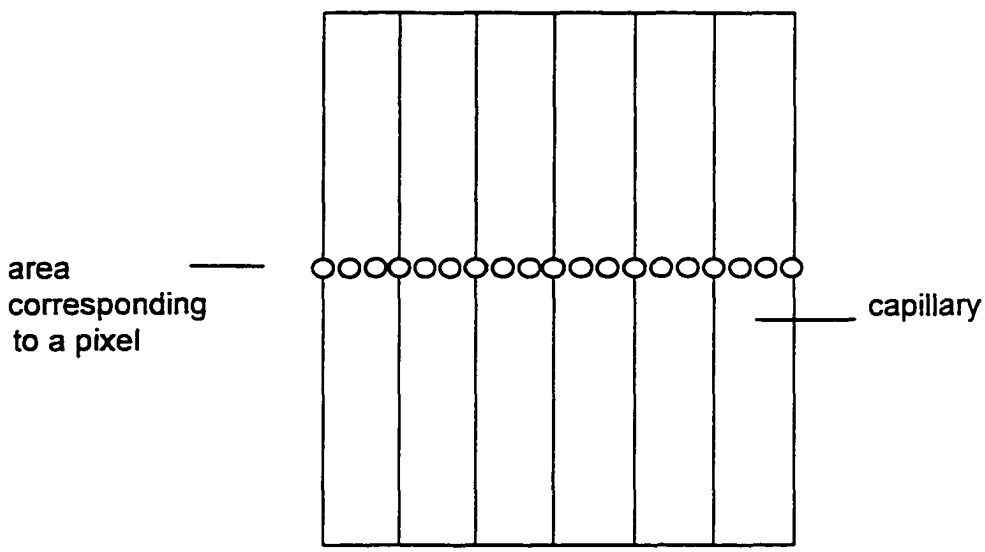
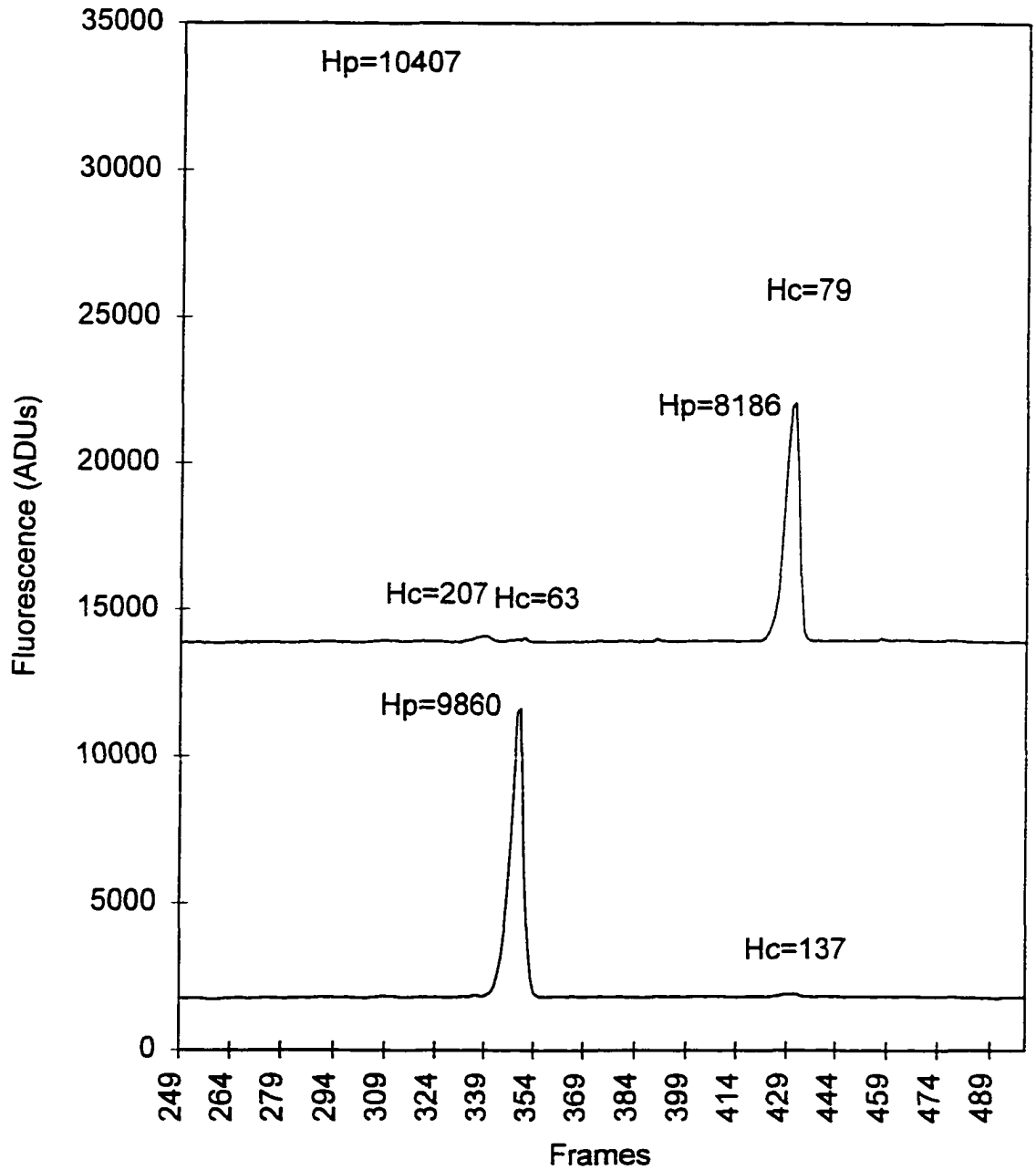


Figure 4. Overhead view of the 3:1 optical coupling arrangement.

Figure 5. Cross-talk in air with 3:1 optical coupling arrangement. The camera aperture was set at 1.4. A piece of 550 nm glass filter was attached in front of CCD camera to cut off the laser light. Hp is the height of the fluorescein peak detected in a capillary; Hc is the cross-talk signal observed in adjacent capillaries.


35000
30000
25000
20000
15000
10000
5000
0
Fluorescence (ADUs)
Hp=10407
Hc=79
Hp=8186
Hc=207
Hc=63
Hp=9860
Hc=137
249
264
279
294
309
324
339
354
369
384
399
414
429
444
459
474
489
Frames

whole range is the basic requirement. Stepping down the aperture is necessary. After the lens aperture is reduced, the field depth is increased, the focusing becomes sharper for the whole range. Aperture 2 is good enough for the focusing quality concern. At the same time, only 2 times decrease in the light collection efficiency can be obtained.

Human's action accounts for the subjective reason. Basically, there are two ways. One is when you image capillary array, before filling any solution into the capillary array, you do focusing either under the room light for CID or fairly dark room for CCD. Keep adjusting the focusing until both sides of capillary walls can be recognized clearly. The other way is filling dye solution into several dummy capillaries among the capillary array. Keep adjusting the focusing until you see the highest fluorescence intensity from one of the center pixels. Clear focusing can be achieved through the control of the lens aperture and the quality of the human action.

Scatter from the optics is another source. Optics here can be lens for light collection or the capillary surface. We cannot do much for the quality control of the capillary surface except by avoiding scratching the fused silica surface when the windows are created. Various methods may be used to remove the polyimide coating from the surface of the capillary [4].

Open flames (matches, lighters) may be used, but they tend to leave the glass surface brittle. Gas torches (eg. O_2/H_2) do a good job of removing polyimide

and leaving the glass strong. But these are difficult to use and sometimes dangerous. Oven may be used bake off the polyimide if the temperature is higher than 600°C. The polyimide will carbonize and flake off the glass. This generally takes 30 to 60 minutes. This method works well to remove polyimide from large sections of tubing. Electric coil heaters (nichrome wire wrapped around insulating quartz tube) burn off the polyimide rapidly. One must be careful not to touch the glass to the coil as this damages the glass surface. Electric arc, plasma may be used to burn off the polyimide in a very small region. Lasers may also be employed to remove the polyimide coating. Although lasers afford a clean source and fine control over the hot zone, they can be expensive and complicated to operate making them impractical for general laboratory use.

Chemical techniques may also be employed. Hot sulfuric acid (100°C) removes the polyimide in less than 20 seconds, but it is difficult and gangerous to work with. Strong bases such as sodium hydroxide will attack both the polyimide and the capillary surface, negating its utility.

Mechanical stripping using a wire stripper may also be employed, but the silica surface is frequently damaged in the process.

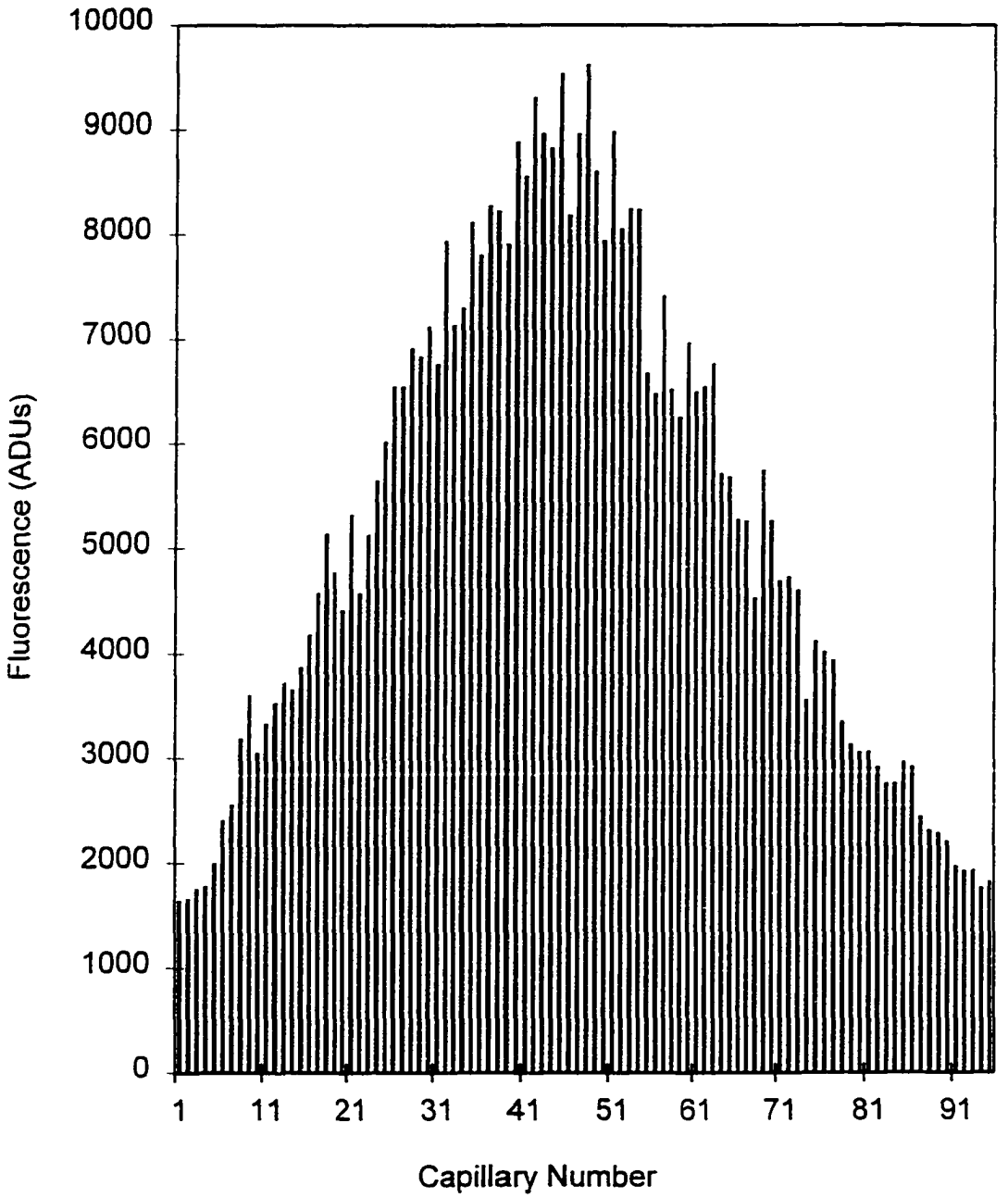
In the experiments in this chapter, an electric coil heater is used to remove the coating of the capillaries for the single capillary experiments or the small array experiments. This procedure was found to yield clean glass windows with a minimum of carbon deposits between the glass and the polyimide coating. This

characteristic is very important in order to have flat packing for the multiple capillary array.

The question is what percentage of cross-talk is acceptable in multiple capillary system? The answer will be different for different applications. For most CE applications, if the retention time of the samples for different channels do not overlap, cross-talk is easily identified and eliminated without causing any problem. For DNA sequencing, peaks come out consecutively. How many percentages of cross-talk will matters? Different base-calling scheme, different enzyme used for sample preparation and different dye set will result in different tolerance to the cross-talk. In our system [5], base-calling is based on two channel ratio-gram for dye-tagged terminators. There is no change required in the dye chemistry and sample preparation protocols. If the cross-talk can be controlled under 2%, it will not affect the base-calling [5, 6].

Excitation efficiency is automatically increased by closely packing the capillary array. The laser light will not be wasted on the gaps between capillaries which was the case in the set-up of Ueno and Yeung [1]. Since the argon ion laser's intensity distribution is Gaussian, close packing will also help the uniformity of the signals from the capillary array. Figure 6 shows the signal distribution from the 96 capillary array when laser was expanded to a 2.5 cm line. The signal from the side capillary is about 6 times lower than that of the center one. The line can be stretched further to obtain more even signals. But it sacrifices more of the laser

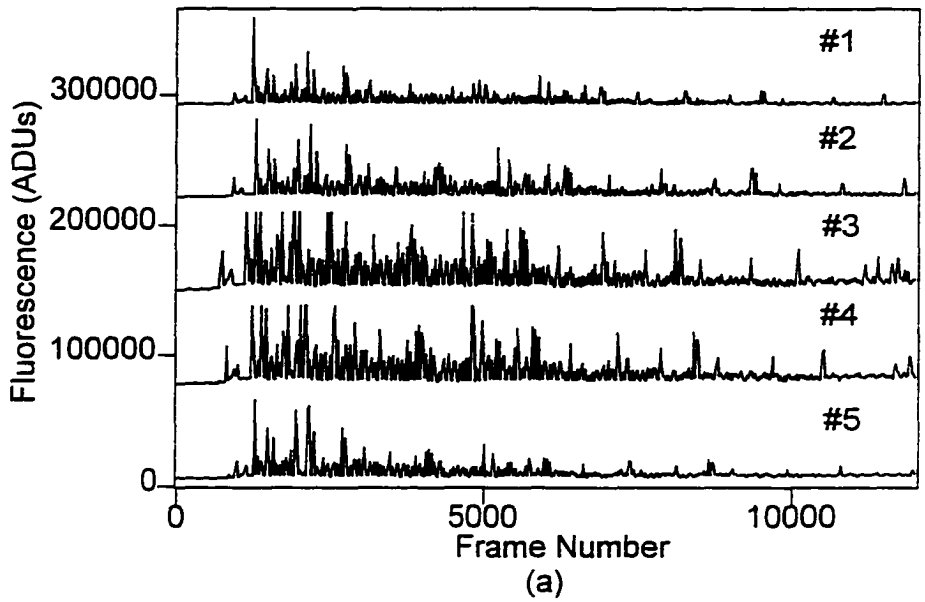
Figure 6. Excitation uniformity of the focused line geometry. 10 mW 488 laser was expanded by 10 cm lens and focused down into 2.5 cm line by a 10 cm cylindrical lens. Lens aperture, 2.0. Fluorescein concentration, $1x10^{-7}$ M. Exposure time is 100 millisecond.


10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Fluorescence (ADUs)
1
11
21
31
41
51
61
71
81
91
Capillary Number

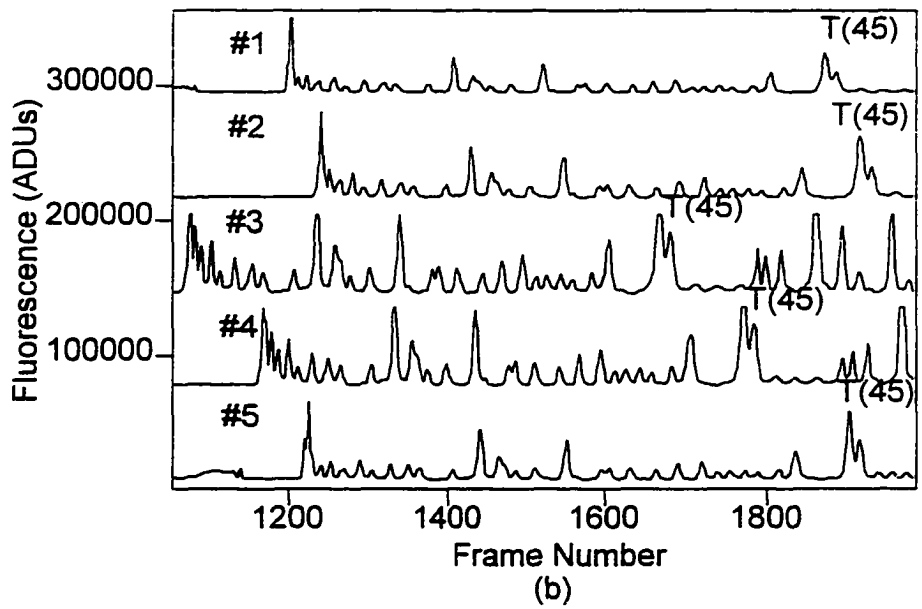
power. The bottom line is that the side capillary should have enough signal to noise ratio to do base-calling without saturation of CCD or CID camera with the center capillaries' signals.

Figure 7 presents a DNA sequencing run performed on 5 capillaries. Capillary #1 and #5 are side capillaries. Capillary #3 is the center one. Capillary #2 and #4 are in between. The #5 capillary is 15 mm away from the #1 capillary. The excitation and detection setup is shown in Figure 1. From the intensity distribution of the capillary array (Figure 6), 5 to 6 times increase in power is necessary in order for side capillary has the same signal level. It is not the absolute signal level that will count. Even though signal level decrease 5 times, If the noise level also decreases 5 times, same detection limit can be obtained from the center and side capillaries. In the above run, peak T of the 45th base is used to compare the detection limits from these 5 capillaries. The results are shown in Table 1. From the comparison of the peak heights of T, the array is put off the center of the laser line. If 100 capillary array is positioned better, the signal level from the side capillary will be higher than capillary #1. The noise level from the side capillary is lower than that from the center capillaries. Even though the signal form capillary #3 is about 6.5 times of that from capillary #1, the detection limit is about 3 times better than that of capillary #1. This run was done in one channnel detection mode with the Raman Edge filter in front of the CCD camera. In order to do base-calling, two-channel [5] detection is necessary. A 610 nm glass filter is tilted underneath the Edge filter to split the signals into two

Figure 7. PGEM/U DNA sequencing data on a 5 capillary array which covers 15m (100 capillary array) range. The experiment was done by utilizing the AT200 CCD camera. Laser power, 300 mW; Lens aperture, 2.0; Exposure time, 300 ms; Other parameters were described in experimental section. (a) whole electropherograms of the five capillary array. (b) part of the electropherograms of the five capillary array .


#1
#2
#3
#4
#5
300000
200000
100000
0
Fluorescence (ADUs)
0
5000
10000
Frame Number

(a)


#1
#2
#3
#4
#5
T(45)
T(45)
T(45)
T(45)
T(45)
300000
200000
100000
Fluorescence (ADUs)
1200
1400
1600
1800
Frame Number

(b)

channels (Figure 1). The red channel has to be favored (about five times of the signal which goes to green channel) to have similar signal level as the green channel. For two channel detection, at least 1.5 W laser is required in order for the array to have same signal distribution and signal to noise from both channels. If more uniform signal to noise from the array is required for base-calling, the laser spot can be further expanded to a longer line.

Table 1. Comparison of the detection limits of the 5 capillary array in focused line excitation geometry by 45th T base of PGEM/U

Capillary	Peak height of T (Hp)	Peak to peak noise (Hn)	Relative detection limit (Hp/Hn)
#1	6300	70	90
#2	20437	120	170
#3	40727	120	339
#4	40792	170	240
#5	25435	110	231

Injection. In the multiple capillary system, different samples will be injected into different capillaries. A scheme for the individual injection without cross contamination is necessary. Two schemes have been proposed and tested. One is an injection device with 100 needles built on one aluminum plate which are the individual electrodes for different samples (Figure 8). This plate can be made

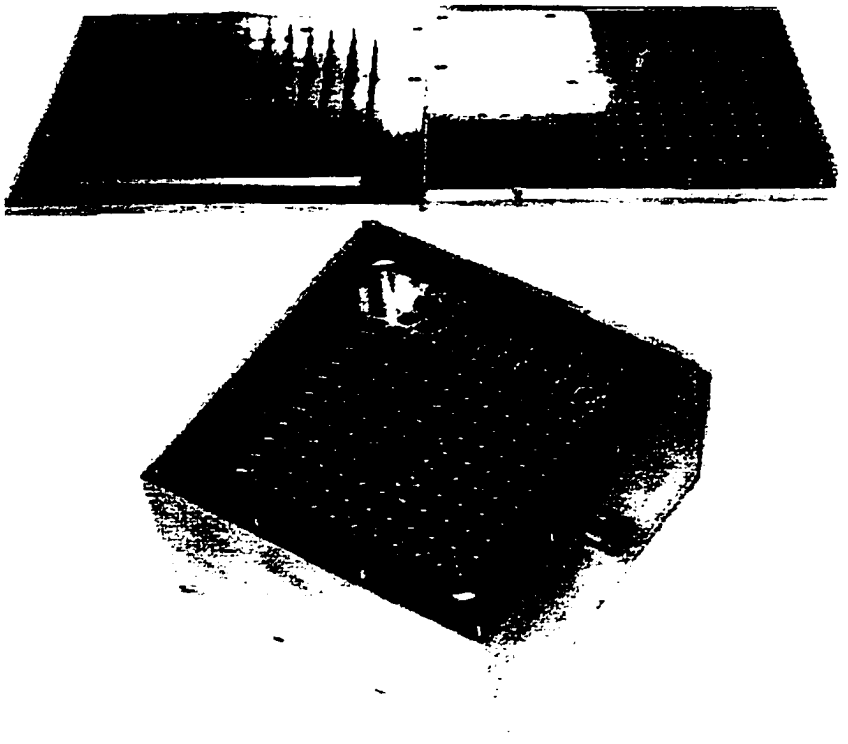


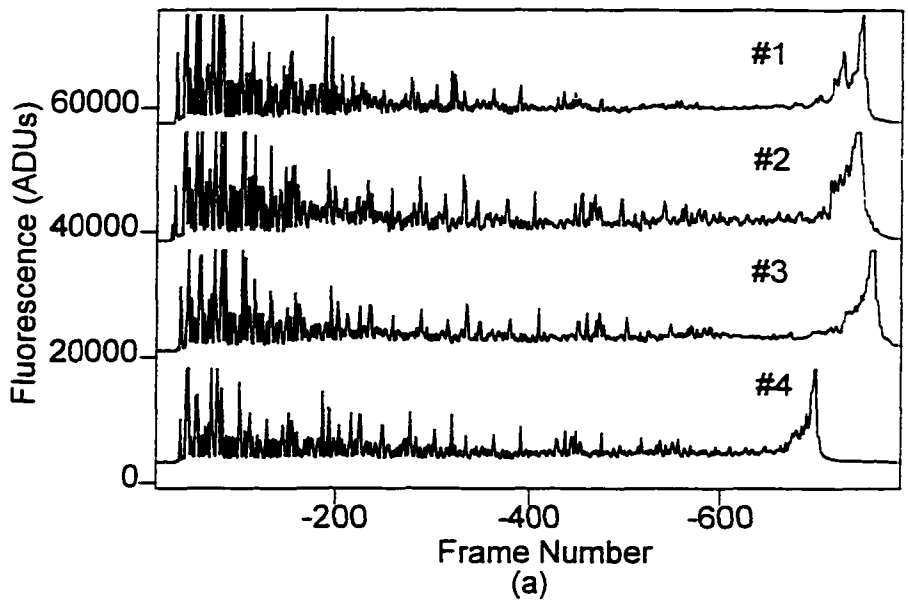
Figure 8. Multiple capillary injection device (100 needle plate and block).

disposable. The device is composed of three parts. The bottom is an electrode plate. The middle is the vial holders. The top plate with small holes can be used to hold capillary array and work as the guidance for dipping capillary into each vial. This device can also function as a heating block for the sample denaturing. The combination of denaturing process with hot injection process can be implemented.

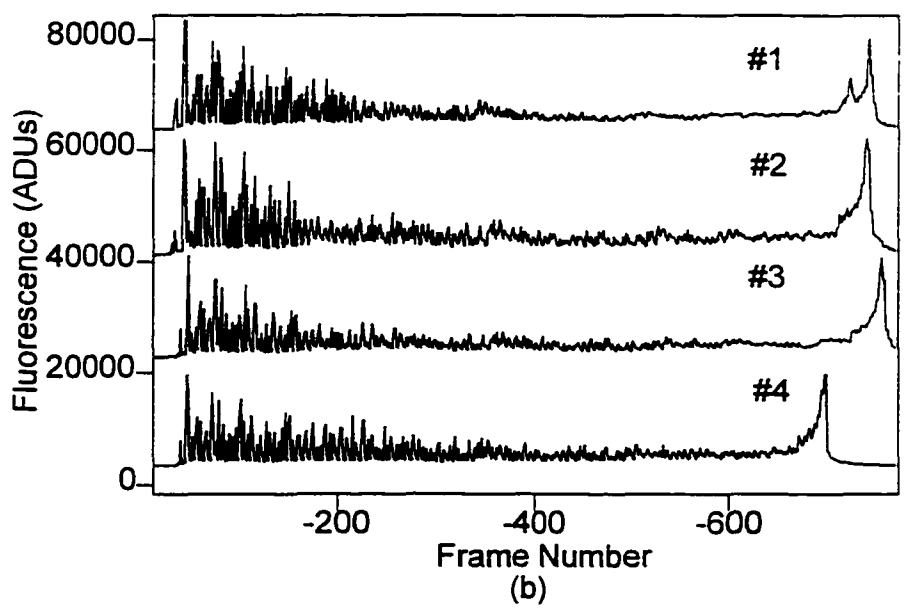
Figure 9 shows the results of test on the injection device by PGEM/U DNA sequencing from four capillaries. The capillaries are 65 cm long with 50 cm efficient length. Capillary is 360 μm o.d. 75 μm i.d.. 15 mW laser was focused onto the capillaries by 20 cm circular lens with the side entry excitation scheme (laser sequentially pass through capillary array) in air. Four samples have been prepared cooked and cooled down rapidly in a 50% (v/v) ethylene glycol/water mixture. The sample tubes were put into the vial holders and the needles may punch through the bottom of each vial. The sample was injected at 75 V/cm for 65 s, and electrophoretic separation was carried out at 150 V/cm. The detection scheme is shown in Figure 1. The result shows that the injection device may be used for individual sample injection into capillary array.

The other external injection scheme, external wires are connected together (Figure 10). Ten external electrodes have been connected to the center wire which goes to the high voltage end. Beside each electrode, there is another hole through which each capillary will go. The spacing is designed in such a way as to fit the 96 tube microtiter.

Figure 9. Individual Injection of PGEM/U DNA into 4 capillary array in air by 100 needle plate injection device. Injection, 65 s at 75 V/cm; Laser, 15mW; Exposure, 300 ms; Aperture, 2.0. (a) electropherograms of four capillary array in green channel (part only with the 514 nm Raman edge filter). (b) electropherograms of four capillary array in red channel.


Fluorescence (ADUs)
60000
40000
20000
0
#1
#2
#3
#4
-200
-400
-600
Frame Number

(a)


Fluorescence (ADUs)
80000
60000
40000
20000
0
#1
#2
#3
#4
-200
-400
-600
Frame Number

(b)

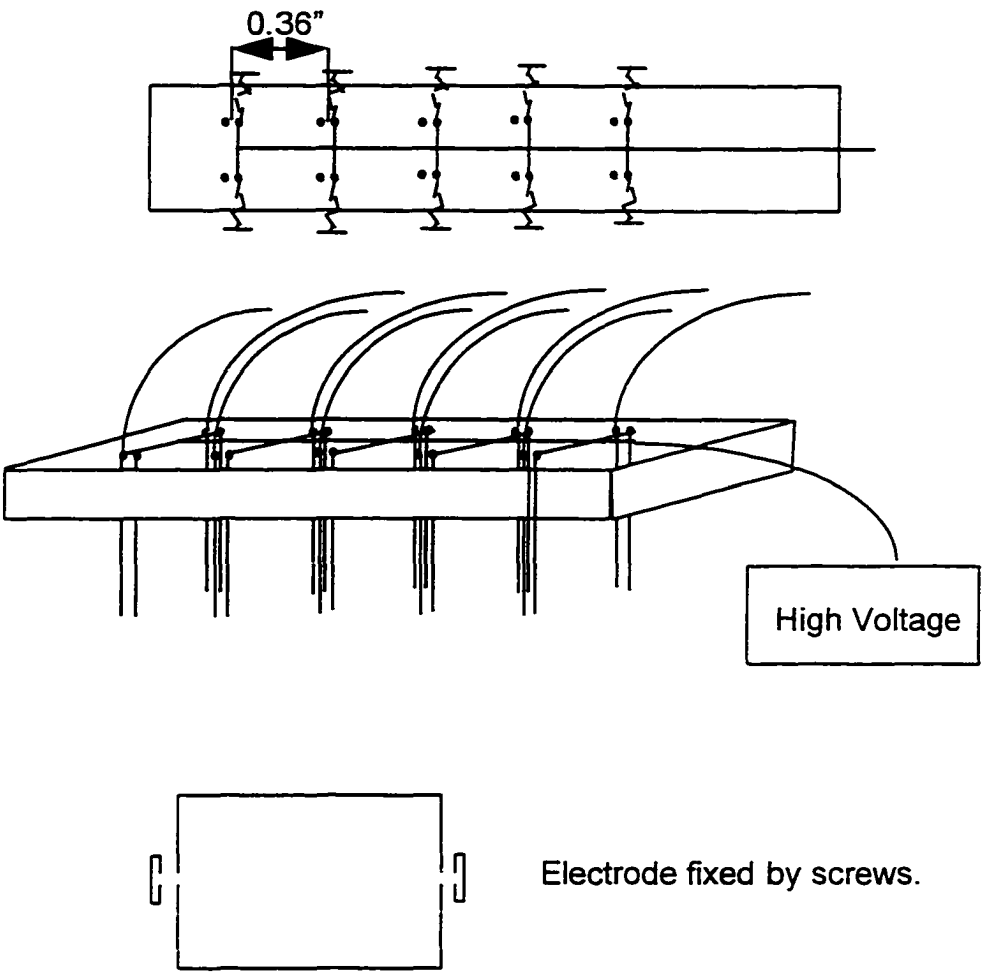
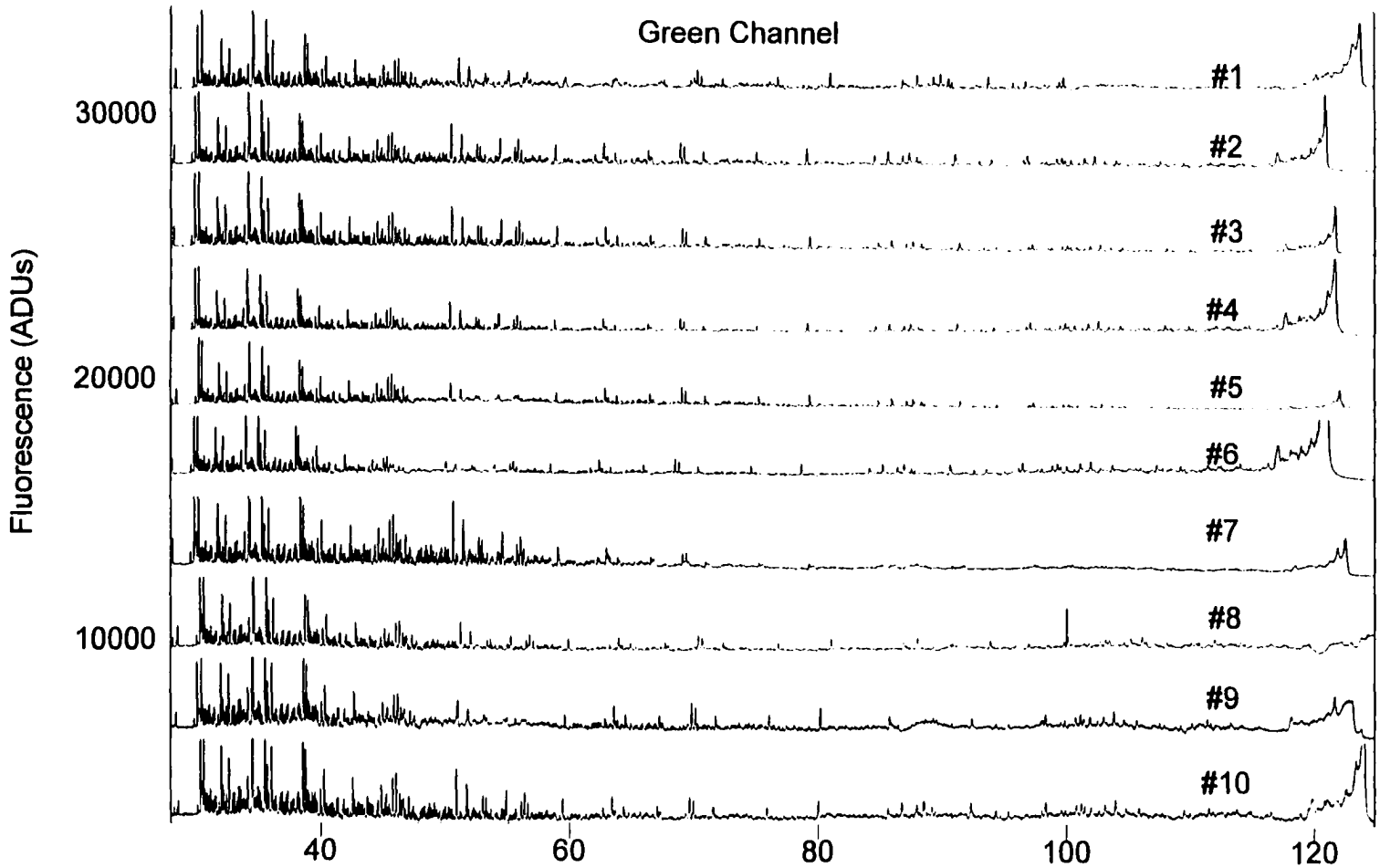


Figure 10. The 10 external electrodes injection device.

Figure 11. Injection of PGEM/U DNA to 10 capillary array by the 10 external electrodes injection device. Injection, 60 s at 65 V/cm; Laser, 15mW; Exposure, 400 ms; Aperture, 2.0. (a) electropherograms of ten capillary array in green channel (part only with the 514 nm Raman edge filter). (b) electropherograms of ten capillary array in red channel.


Green Channel
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
30000
20000
10000
Fluorescence (ADUs)
40
60
80
100
120
Minutes

(a)

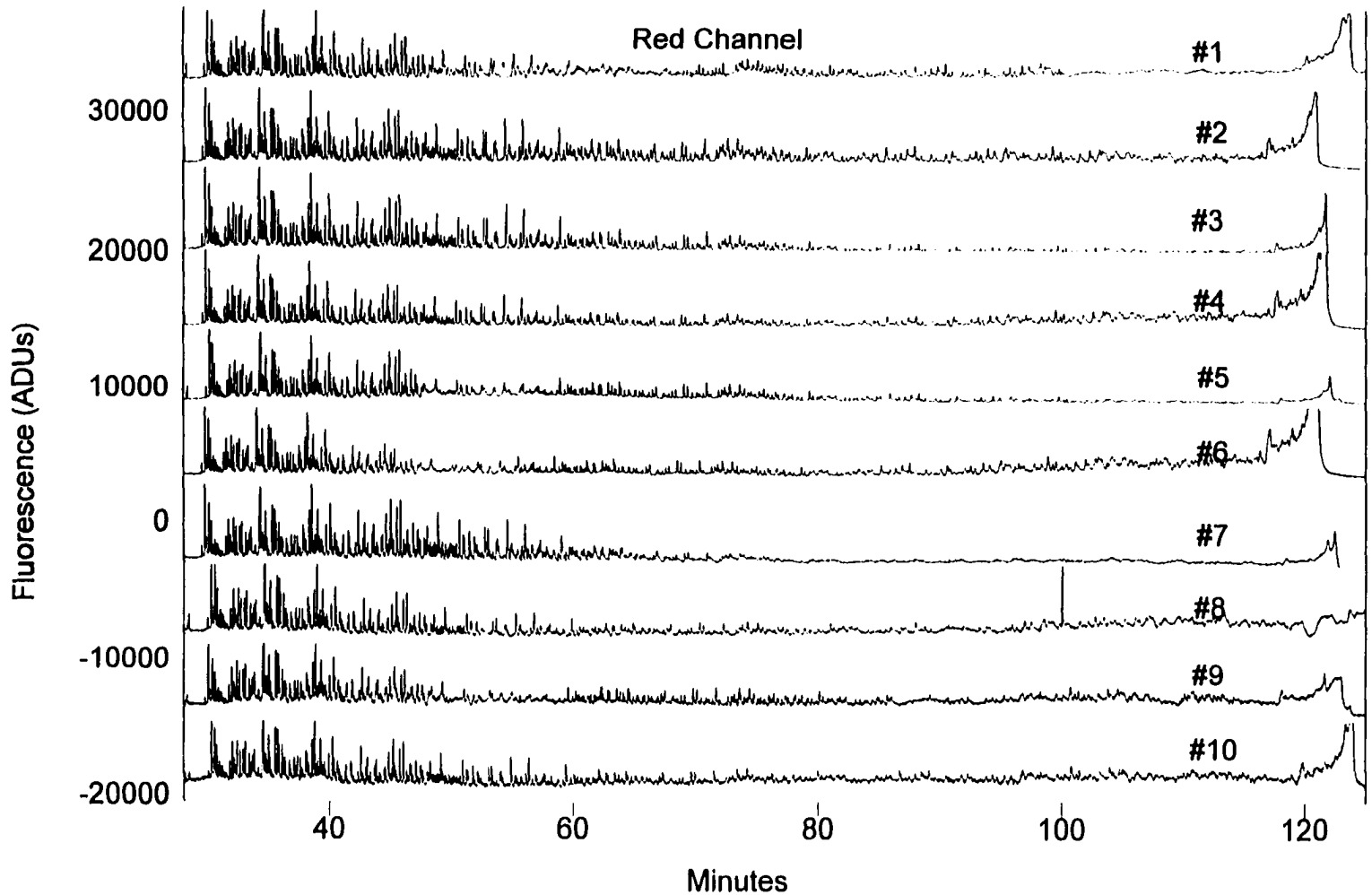


Figure 11 (continued): (b)

Figure 11 shows the results of a 10 capillary array DNA sequencing experiment through 10 external electrode injection. All the parameters are the same as in Figure 9 except for the different injection design. This scheme is good for the small number of capillaries and not as convenient as the needle plates for large array capillaries.

Conclusion

The optimizations have been done on the focused line excitation and detection scheme from the point of view of reducing cross-talk and enchancing excitation efficiency. The individual injection schemes have been proposed and successfully used for by doing two channel DNA sequencing.

References

[1] K. Ueno and E. S. Yeung, *Anal. Chem.*, *66*, 1424 (1994).

[2] J. A. Taylor and E. S. Yeung, *Anal. Chem*, *65*, 956 (1993).

[3] X. Lu and E. S. Yeung, *Appl. Spectrosc.*, *49*, 605 (1995).

[4] Private communications with *Polymicro Technologies Incorporated.*

[5] Q. Li and E. S. Yeung, *Appl. Spectrosc.*, *49*, 1528 (1995).

[6] Q. Li Ph.D. dissertation.

CHAPTER 2

OPTIMIZATION OF EXCITATION AND DETECTION GEOMETRY FOR MULTIPLEXED CAPILLARY ARRAY ELECTROPHORESIS IN DNA FRAGMENTS

A paper published in Applied Spectroscopy[1]

Xiandan Lu and Edward S. Yeung*

ABSTRACT

A new on-column excitation and detection geometry for multiplexed array capillary electrophoresis is evaluated. The detection limit is improved 100-fold for the same laser power due to more efficient coupling of the laser beam. The illumination is also more uniform over the entire array. Stray light and cross-talk are effectively suppressed by immersing the capillaries in water to roughly match the refractive index. A CCD camera combined with cutoff filters provides adequate sensitivity and speed for high speed DNA sequencing based on sieving in a polymer-matrix solution.

INTRODUCTION

The success of the Human Genome Initiative [1] is dependent on the development of new technologies in several fields [2,3]. In particular, new DNA sequencing schemes must be developed to fulfill the goals of automation, high speed, high accuracy and low cost. Polyacrylamide gel-filled capillaries have been demonstrated to enhance the separation speed of DNA fragments by 25 fold compared to slab gels [4-7]. Not only can the speed per lane be improved, but it is also easy to increase the overall throughput by using a large number of capillaries in parallel in a single instrument. In the last few years, our group [8,9] and several others [10,11] have demonstrated different setups and detection systems to accommodate parallel arrays in capillary electrophoresis (CE). These have the potential for high-sensitivity, high-speed and high-throughput DNA analysis.

In one approach [10-12], a two-color confocal fluorescence scanner was employed for 25 capillaries. The laser is focused onto one capillary by a microscope objective. And the fluorescence is collected and focused onto a single photomultiplier tube by the same objective. Detection is one capillary at a time, and the capillary array is translated across the excitation and detection region at the speed of 20 mm/s, which is 1 Hz for each capillary, by a mechanical stage. From the results, observation time is long enough for each capillary to obtain enough signal-to-noise ratio for base calling without loss of temporal resolution. Since data acquisition is sequential and not truly parallel, it may limit its use for very large

numbers (hundreds) of capillaries. The sequencing high throughput will be determined by the exposure time required for each capillary to get enough signal-to-noise ratio. Because of the utilization of the microscope for confocal detection, the focus is very shallow (about 25 μm). This requires the translational movement to be very flat or else the objective will miss the image.

In another setup [13-14], multiple sheath-flow and four-color detection have been developed. Twenty gel-filled capillaries were lined up at a 0.35-mm pitch, and were placed into a flow cell to couple with 20 large i.d., gel-free capillaries. Two laser beams are combined into one to cross the flow streams in a line for excitation, and a charge-coupled device (CCD) was used for simultaneous detection perpendicular to the excitation beam. Superior stray-light rejection was achieved [15-16]. The main challenges in scaling up to hundreds of capillaries are alignment of the individual sheath flows, possible turbulence in the flow paths, and improper matching of the laser beam waist over a long distance with the core diameters containing the eluted fragments. And also after each run, the sheath-flow cell need to be dismantled to avoid cross-contamination. So there is good reason to perform on-column detection by trading sensitivity for convenience.

Recent work in our group focused on the improvement in sensitivity, signal uniformity and efficiency of the system. Since the DNA fragments from a Sanger reaction are in the amol and not the zmol range per capillary, we decided to employ on-column detection to simplify the instrumentation. The first arrangement [8] involves

axial-beam laser-excited fluorescence detection. Optical fibers were inserted into the ends of the capillaries for excitation and simultaneous detection was achieved with a CCD camera perpendicular to the capillary array. This is an easy and efficient approach to couple the laser into each capillary without moving parts and without critical alignment. However, the intrusion of the optical fibers into the separation capillaries affected the electroosmotic flow and increased the possibility for contamination and clogging. The second arrangement [9] involves a beam expander and a cylindrical lens to distribute the laser light into a thin line which intersects the axes of the capillaries. Grooves on an aluminum block were constructed to optically isolate each capillary from the others to reduce cross-talk. A CCD camera was used to detect fluorescence at 45° relative to the excitation beam. Compared to the first arrangement, this provides a lower detection limit and a more uniform distribution of excitation intensities. Because of the Gaussian intensity distribution, a longer laser line compared to the array width has to be used. Furthermore, half of the laser light in the array region is wasted due to the presence of the spacer grooves. Cross-talk is acceptable, but is still in the 10% range.

In most laser-excited fluorescence detection schemes, the incident laser beam and the collected fluorescence light are perpendicular to each other in order to reduce background from scattering. Here we present a new functional system for monitoring a large array of capillaries simultaneously undergoing electrophoresis with high sensitivity and efficiency. By immersing the capillaries in an index-

matching fluid, scattering and refraction from the capillary walls are minimized and the laser beam is able to pass through all capillaries with only minor losses.

EXPERIMENTAL SECTION

Side-Entry Excitation. The schematic diagram of the experimental arrangement is shown in Fig. 1. An air-cooled argon-ion laser (Uniphase, San Jose, CA, model 2213-150ML) operating simultaneously at several visible lines was used for excitation. The wavelength of the laser was chosen by an interference filter or a glass prism. When using a glass prism, the setup allows multiple-color excitation. The laser beam was focused by a 10-cm focal length lens (Edmund Scientific Co., Barrington, NJ). The range over which the beam remains smaller than 75 μm was 1.5 cm. This was confirmed by translating a photodiode (Hamamatsu Corp., Middlesex, NJ) across the beam in conjunction with a 5-μm pinhole. The output of the photodiode was monitored by a digital multimeter (Keithley Instruments, Cleveland, OH). Two mirrors (Melles Griot, Irvine, CA) were used to adjust the direction of the laser beam to become parallel to the plane of the capillary array and to pass through the centers of the capillaries.

Fused-silica capillaries (Polymicro Technologies, Phoenix, AZ, Model 75 μm i.d., 150 μm o.d., total length 50 cm, effective length 35 cm) were mounted by adhesive tape (3M water-proof Scotch Brand Tape, St. Paul, MN) on a polished aluminum surface in a close-packed configuration. The 1-cm detection windows

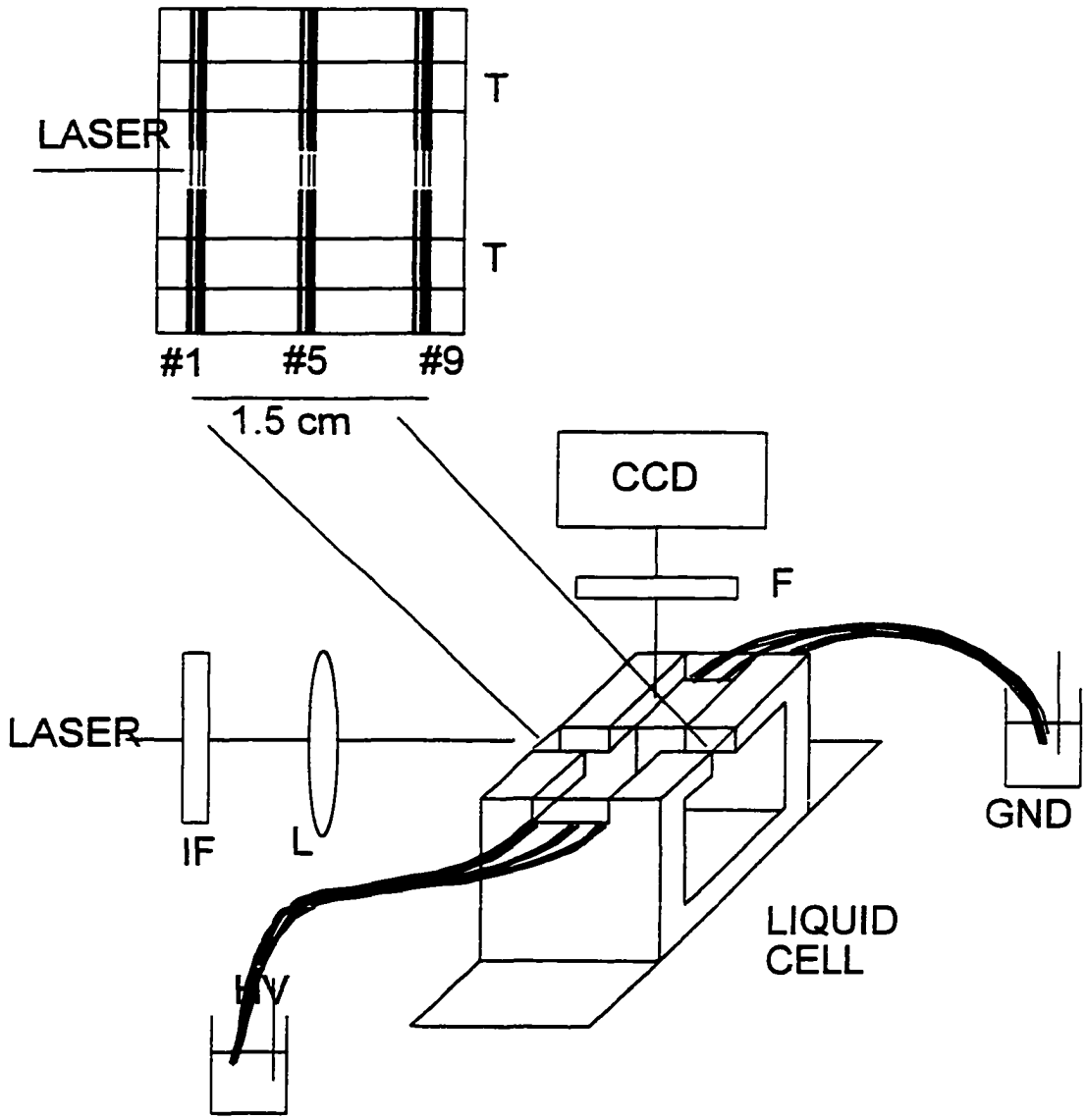


Figure 1. Schematic diagram of the laser excitation and CCD detection system. **F**, cut-off filters; **L**, 10-cm f.l. lens; **IF**, laser-line filter; and **T**, tape for holding the capillaries in place.

were created by removing the polyimide coating with boiling sulfuric acid. All the capillaries were immersed into water in the liquid cell. A translational stage was used for fine adjustment of the position of the cell in order to place the focal point of the laser at the center of the array and obtain the right optical coupling arrangement.

CCD detection system. The camera head (CH-220 thermo-electrically cooled/liquid-circulation heat dissipation, Photometrics TH7883-PM) was cooled down to –40° C. It was mounted on top of the array, facing downward. There are 384x576 pixels, each being square with 23-mm edges. The camera electronics unit (CE-200) contains an analog-to-digital converter (ADC) providing 14 bits of precision, at a conversion gain that is software controllable.

The camera controller (CC-200) contains a 68000 processor, image-frame RAM (12 MB), firmware ROM, an IEEE-488 interface to a host personal computer, a "mouse" port, and interface circuitry for a video monitor (RS-170). The RS-170 subsystem and the mouse provide feedback for the experimenter during both equipment alignment and data acquisition. The capillary array in the detection region was imaged onto the CCD sensor through a 24-mm wide-angle lens (Canon, Tokyo, Japan, Model FD 24mm F1.4L). Different sets of filters were placed in front of the lens in different experiments.

The software for CCD image data acquisition is offered by Photometrics. Data analysis is performed off-line, using software written in Turbo Basic (Borland). During data acquisition, all images are stored in the cache memory, which are later

transferred to the hard drive. The maximum image number is determined by the amount of RAM and the size of each frame. For 100 capillaries, by binning 3 consecutive pixels along each capillary, the pixel number for each frame is only 200.

Focus adjustment. To provide the most efficient imaging, each capillary was imaged to a width of 2 CCD pixels. With 75 mm i.d. and 150 mm o.d. capillaries tightly packed side-by-side in an array, alternate pixels will trace out the liquid cores of the capillaries. The adjacent capillary walls, which represent useless information, will be imaged onto the spaces in between. So, it is critical to simultaneously achieve the correct magnification, the good packing quality, the correct alignment, and the correct focus in manipulating the camera lens.

Achieving the best focus is simplified by the use of the RS-170 video monitor. A series of images were obtained and displayed on the video monitor to allow slight adjustments of the focus each time. For final adjustment, we used two bare columns filled with 10^{-10} M fluorescein on both sides of the array. The best achievable focus should result in the least number of pixels and the strongest signal for each capillary.

Sensitivity test. The 488-nm laser line (about 25 mW) was selected by a glass prism for excitation. Five capillaries were packed side-by-side in the liquid side-entry cell. The running buffer was 10 mM phosphate at pH 9.5. Electrophoretic separation was driven at –20 kV using a high voltage power supply (Spellman, Plainview, NY, Model UHR50PN50). Samples were introduced by electrokinetic

injection for 3 s. Two cut-off filters (Melles Griot, Model OG550) were used in front of the CCD camera to reject stray light.

Fluorescence detection of DNA size markers. The laser line at 514nm (3 mW) was chosen to illuminate the analytes. A RG610 cut-off filter and a 630nm interference filter were used to reduce scattered light. Signals were detected by the CCD camera at the rate of 1 Hz with 0.8-s exposure time. Nine capillaries divided into 3 groups of 3 were lined up at three positions along the laser beam (Fig. 1). The center 3 capillaries were at the beam waist of the laser. The other two groups were at the edges 0.75 cm each from the center. The capillaries were coated by the method of Hjerten [12]. The total lengths of the capillaries were 50 cm and the effective lengths were 35 cm or 40 cm. Among each group, the capillary in the middle was made longer than the other two to produce a different set of migration times. The high-voltage power supply was operated at –10 kV.

The running buffer was 1xTBE (89 mM Tris., 89 mM boric acid, 2 mM EDTA) with 1 mg/ml of ethidium bromide (Sigma Chemical Co., St. Louis, MO). The polymer matrix [13] was a solution of 0.6% each polyethylene oxide (PEO) of MW 300,000, 600,000, 1,000,000, 2,000,000, 5,000,000 and 8,000,000 respectively (Aldrich Chemical Co., Milwaukee, WI). pBR 322 Hae III DNA (Boehringer Mannheim Biochemicals, Indianapolis, IN) was diluted with deionized water and injected at 10 kV from the negative high voltage side for 3 s. Before injection, the capillaries were equilibrated at 10 kV for 10 min. After each run, the old polymer

matrices must be flushed out from the capillaries. The capillaries are reusable by filling with a fresh polymer solution.

Separation of DNA sequencing ladder. The 488-nm laser output at 10 mW was used for excitation. A 488-nm interference filter was used to eliminate the plasma lines. Two OG515 cut-off filters were used to discriminate against the laser line. One capillary from the center group in the arrangement above was selected for the separation, which was driven by –10 kV.

Buffer and polymer matrices were the same as above except 5 M urea was added. PGEM/U DNA sample was obtained from the Nucleic Acid Center of Iowa State University (Ames, IA). A 1.7-ml denaturing solution (the ratio of EDTA to formamide = 1:5) was added to the sample vial. This was immersed in a water bath (90-100 °C) for 3 min to denature the DNA. The sample was injected at –3 kV for 12 s.

RESULTS AND DISCUSSION

Figure 2 shows the results of the sensitivity test of the system. A 5-capillary array was used. The fluorescence signals from an injection of 9.0×10^{-11} M fluorescein into each capillary were easily detected. The detection limit is at the 10^{-12} M level. Compared to the results with the previous setup [9], there is at least 100x improvement. In the previous setup, the laser was distributed into a thin line to cover 100 capillaries. Taking into account the portions of the laser beam wasted on the

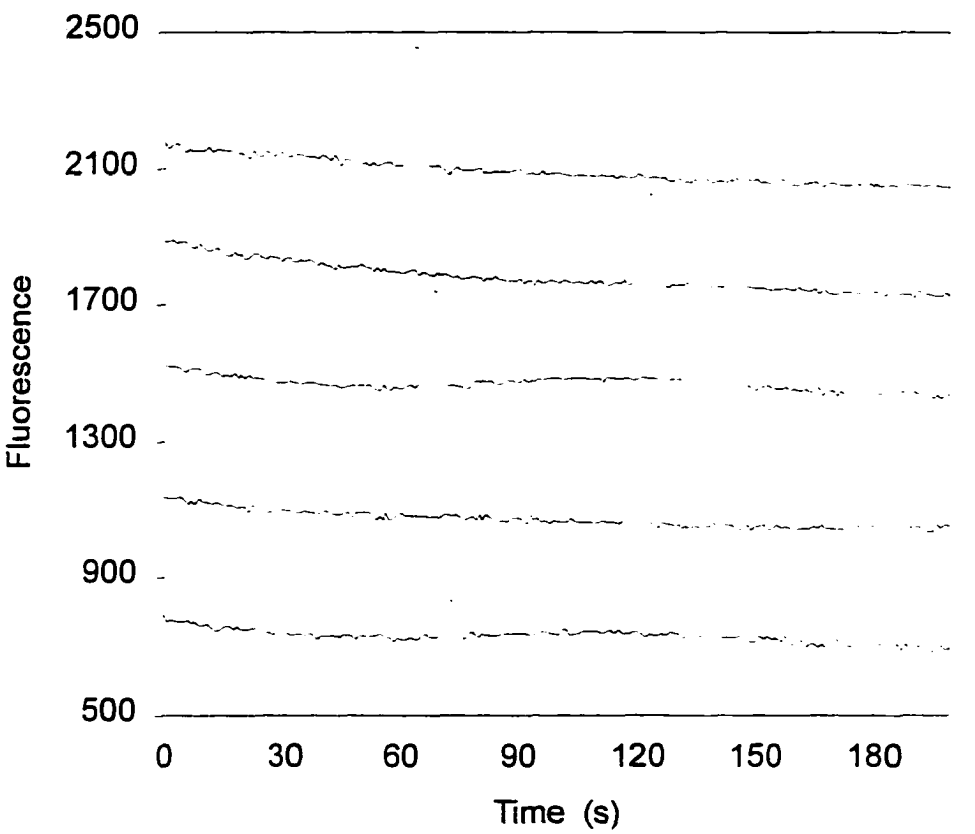
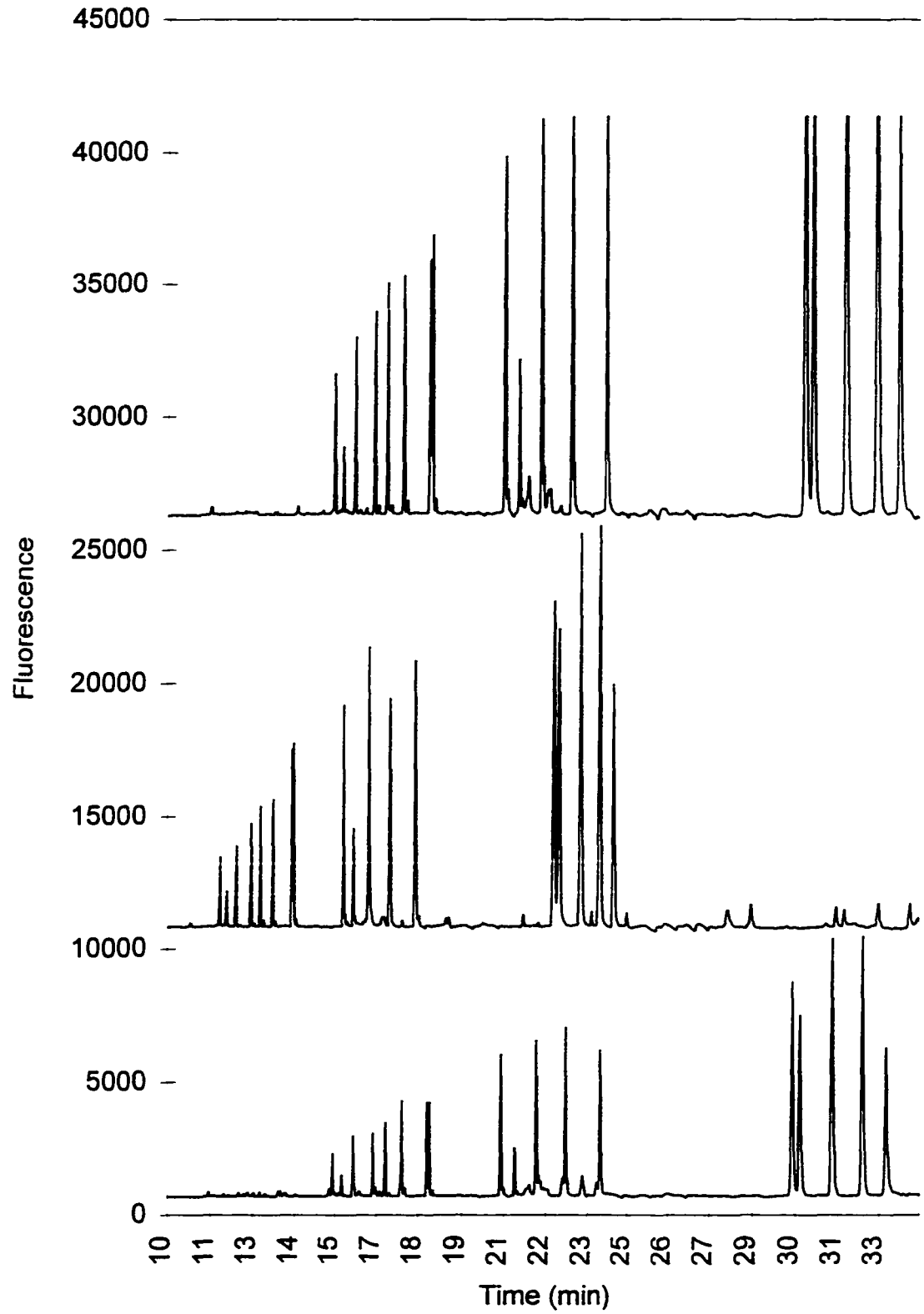


Figure 2. Detection of 9.0×10^{-11} M fluorescein in a five-capillary array. Laser, 488 nm, 25 mW; and injection, 3 s electrokinetic at –20 kV. Data acquisition was initiated 1 min after the start of electrophoresis.

spacer grooves and those falling outside the array, less than 0.5% of the laser power actually irradiated each capillary. In the present setup, the laser sequentially passes through all the capillaries. Because of the low concentration of the DNA samples (10^{-10} to 10^{-12} M), even though dyes with very high absorption coefficients (10^5) are used as tags, the laser power after traversing one capillary will be reduced by less than 0.0001%. So the laser power available for all the capillaries will not decrease much due to absorption of the DNA sample.

The important issue here is light scattering and refraction by the cylindrical capillary walls. Immersing the capillaries in water, which roughly matches the refractive index, is the answer. Figure 3 shows the results of the separation and detection of pBR 322 Hae III fragments from the center one in each of the groups of 3 capillaries (out of 9, see Figure 1). Refraction and scattering are greatly reduced compared to operation in air. Fairly even signals from all the capillaries are obtained. The variation in the detection limit obtained from the 3 groups are within a factor of 2. This difference is expected due to the different spot-size of the laser at each location and diverge effect of the capillary due to the imperfect matching of the immersing liquid, capillary wall and sieving metrix. Most of the rays in a 75-μm beam will not follow a diameter of the capillary for efficient excitation. Since the dimension of the capillary is at the micron level, maintaining the flatness of the array so that the laser can pass through the centers of the capillaries is very important for producing even signal levels for the array. Even better match in the refractive index is possible

Figure 3. Fluorescence of pBR 322 Hae III DNA from three groups of capillaries spaced 0.75 cm apart. Concentration, 0.125 mg/ml; laser, 514 nm, 3 mW; and injection, 3 s electrokinetic at −10 kV. The main features correspond to fragments of sizes (in bp) 51, 57, 64, 80, 89, 104, 123,124, 184, 192, 213, 234, 267, 434, 458, 504, 540 and 587.


45000
40000
35000
30000
25000
20000
15000
10000
5000
0
Fluorescence
10
11
13
14
15
17
18
19
21
22
23
25
26
27
29
30
31
33
Time (min)

by selecting the appropriate immersion fluid. This will further reduce the refraction, the stray light, thus improve the excitation efficiency (will be discussed in Chapter 3).

Cross-talk between the separation channels is another important issue, especially for a multiplexed detection system. Fluorescence light refracted from the walls of the adjacent capillaries causes cross-talk. We have observed such cross-talk in capillary arrays [8,9]. After repeated experiments, it was found that cross-talk can be reduced by good design and alignment, which have been discussed in Chapter 1. The other way to reduce the cross-talk is immersing the capillary array into index matching fluid.

Figure 4 shows the signals of 3 pixels corresponding to the core of 3 consective capillaries which were immersed in water. The optical coupling arrangement is one capillary covers 2 pixels. The liquid core matches one pixel. The average of 2% crosstalk can be observed compared to 6% in air (Figure 4 in Chapter 1). Figure 5 represents the signals of 3 pixels corresponding to the 3 consective capillaries in water by 3:1 optical coupling arrangement. The average of less than 1% of cross-talk has been observed. Some capillaries are completely free from cross-talk. Once again, this is ascribed to the combination of the reduction of scattered light and refraction after refractive index matching, right optical coupling arrangement, and good focus for all the capillaries in the array.

Figure 6 depicts the separation of DNA fragments from the Sanger sequencing reaction obtained from one capillary in the center group. The sample was prepared following the commercial protocol with T7 polymerase and dye-labeled terminators.

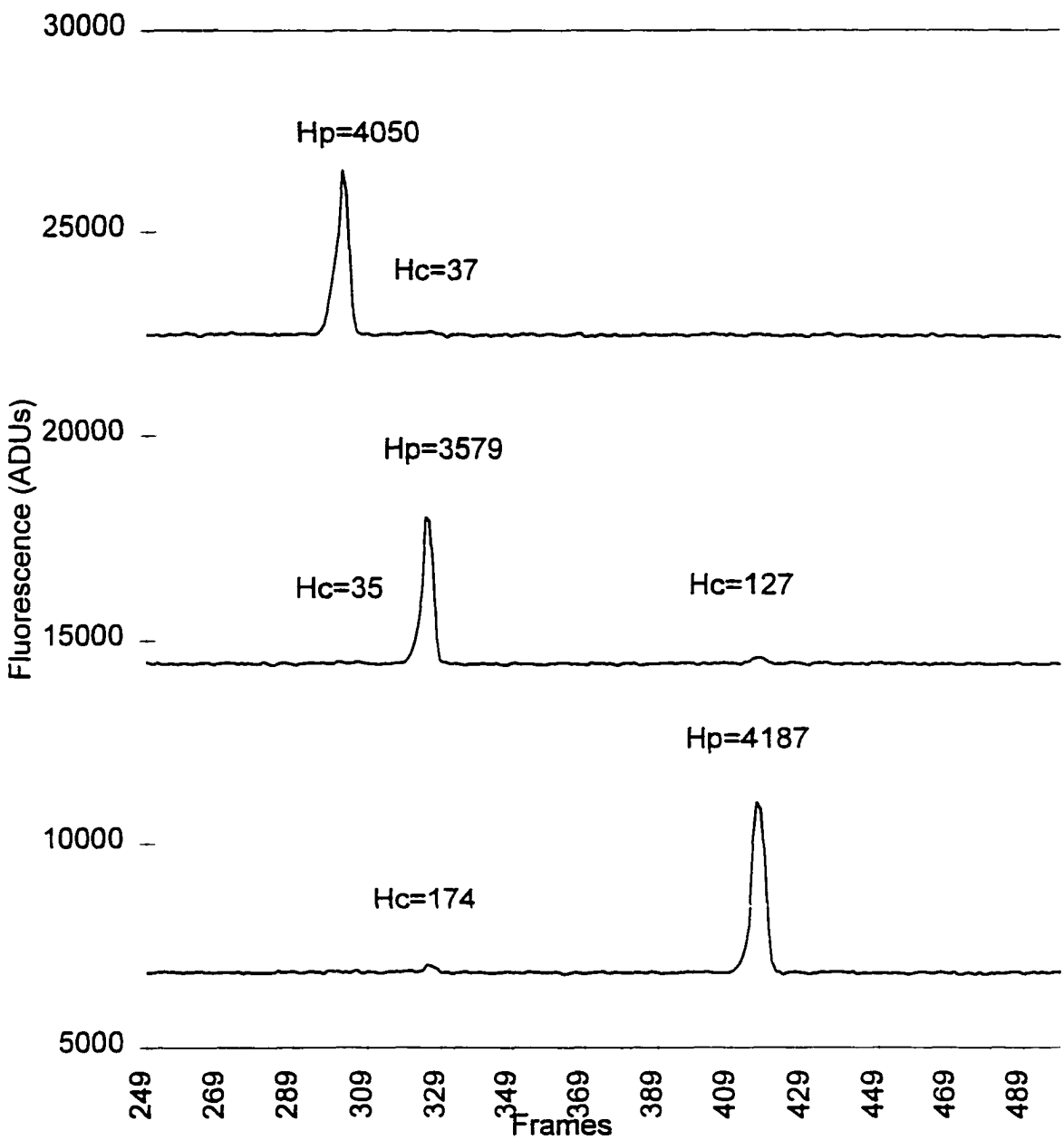


Figure 4. Fluorescence of fluorescein from 3 pixels of 3 consective capillaries in water by 2:1 optical arrangement.

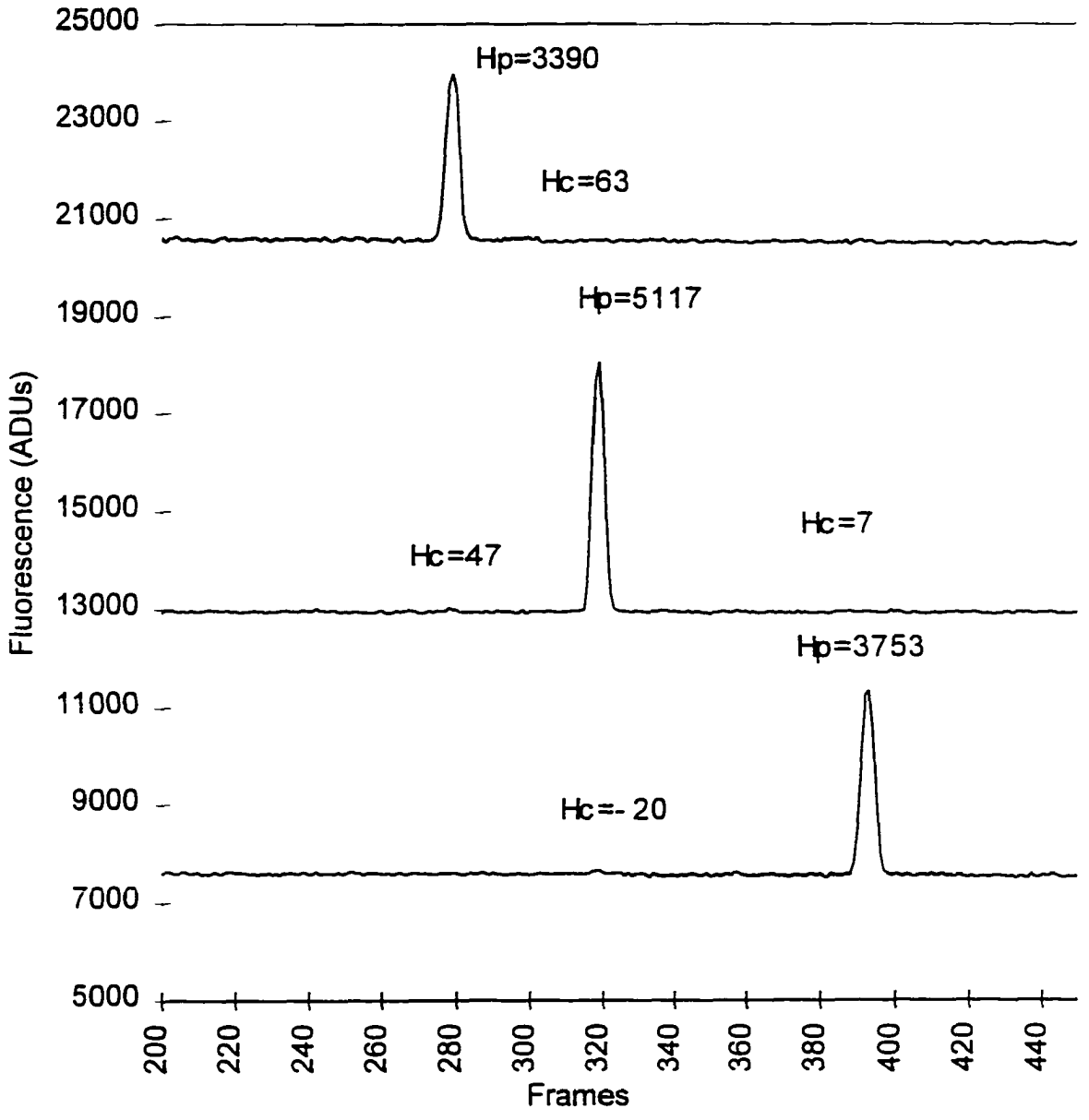
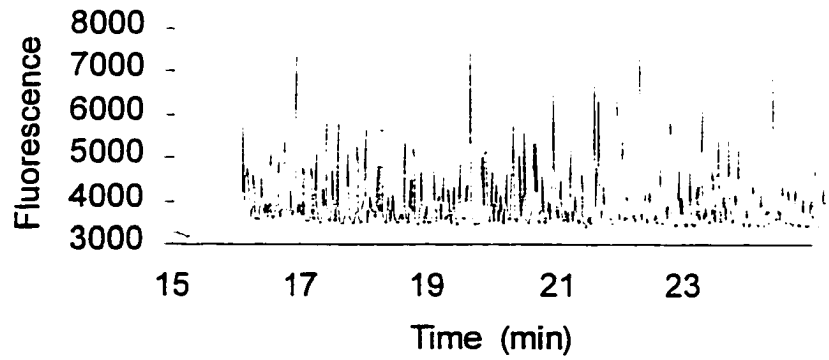
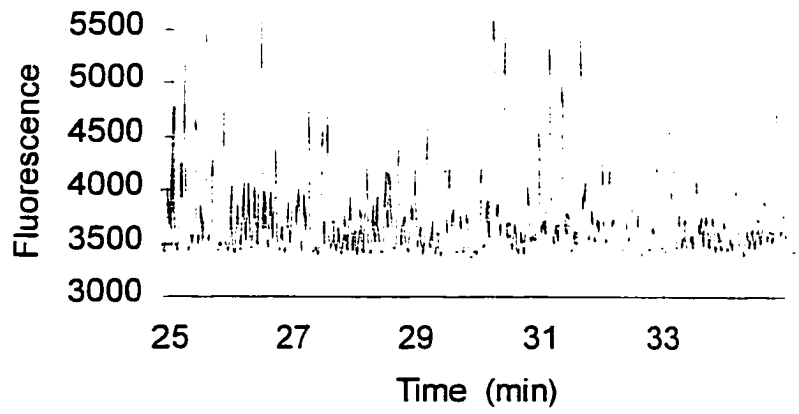
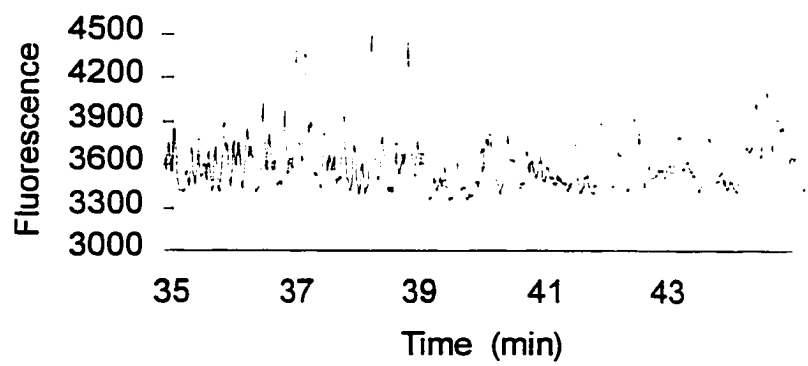


Figure 5. Fluorescence of fluorescein from 3 pixels of 3 consective capillaries in water by 3:1 optical arrangement.

Figure 6. Separation of PEGM/U Sanger sequencing fragments. Laser, 488 nm, 10mW; and injection, 12 s electrokinetic at –3 kV. The scale expansions for **A, B** and **C** are 1x, 2x and 3.3x, respectively, to enhance visualization of the larger fragments.


Fluorescence
8000
7000
6000
5000
4000
3000
15
17
19
21
23
Time (min)


Fluorescence
5500
5000
4500
4000
3500
3000
25
27
29
31
33
Time (min)


Fluorescence
4500
4200
3900
3600
3300
3000
35
37
39
41
43
Time (min)

Less of the larger fragments were injected because of electrokinetic bias [13]. The frame rate for data acquisition was 1 Hz (0.8 s exposure). The data rate was limited by the nonlinear reaction of our particular CCD shutter when the overhead time goes down to 0.1 s. The time resolution is adequate for this particular DNA separation, as indicated by the shapes and spacings of the early-eluting peaks. However, for even faster separations [14], one would need to increase the data rate. Fig. 6 shows that with 0.8-s exposure, the sensitivity of the system is clearly adequate for actual DNA sequencing runs. In fact, sensitivity here is limited by background fluorescence from the polymer matrix and not from insufficient fluorescence intensity from the DNA fragments. The data rate for the new generation of CCD, even in the high-sensitivity, slow-scan mode, can be down to the millisecond per frame level. So, the use of a CCD detector is feasible even when the separation speed is further increased. We have confirmed that the CCD shutter can be kept open during the entire separation while the frames are read out at constant intervals. Even with our relatively slow data-read rate (50 kHz), there is no observable smearing of the information as the charges are shifted down the CCD columns. For 100 capillaries (200 pixels), the total time for shifting an entire CCD column is only 4 ms, which is short compared to the exposure time. The frame rate of our camera can thus be reduced to 0.6 s for 0.5-s exposure times. The only drawback is that one must read entire subarrays from the CCD. When the setup is modified to allow 2 or more laser lines in excitation

[9], it will be more advantageous to use a CID detector, which allows random access of the data [15].

In summary, the results obtained in this work proposed a side-entry scheme for a highly multiplexed CE system in biological, forensic and genetic studies. With the further improvements in excitation optics design, the side-entry scheme can be scalable up to hundreds of capillaries. As for the detection system, 4096 CCD arrays are commercially available to image 2048 capillaries. The array width will increase with the scaling up of the capillary array, which is still compatible with large-format wide-angle lenses. It will no longer be possible to maintain a 75-μm or narrower beam over this width. However, one can readily use higher laser powers in an unfocused beam or a long rayleigh range lens to compensate for the mismatch in size between the laser and the capillary cores. The present scheme promises to be an efficient system for high-throughput DNA sequencing.

ACKNOWLEDGMENT

The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Director of Energy Research, Office of Health and Environmental Research.

REFERENCES

1. L. M. Smith,*Genome 31, 929 (1989).*

2. F. Collins, *Science 262, 20 (1993).*

3. F. Collins and D. Galas, *Science 262, 43 (1993).*

4. A. S. Cohen, D. R. Najaran, A. Paulus, A. Guttman, J. A. Smith, and B. L. Karger, *Pro. Natl. Acad. Sci. U.S.A. 85, 9660 (1988).*

5. L. M. Smith, *Nature 349, 812 (1991).*

6. A. S. Cohen, D. R. Najaran, and B. L. Karger, *J. Chromatogr. 516, 49 (1990).*

7. A. E. Karger, J. M. Harris, and R. F. Gesteland, *Nucleic Acids Res. 19, 4955 (1991).*

8. J. A. Taylor and E. S. Yeung, *Anal. Chem. 65, 956 (1993).*

9. K. Ueno and E. S. Yeung, *Anal. Chem. 66, 1424 (1994).*

10. R. A. Mathies, X. C. Huang *Nature, 359, 167-169 (1992).*

11. X. C. Huang, M. A. Quesada, R. A. Mathies *Anal. Chem., 64, 967-972 (1992).*

12. X. C. Huang, M. A. Quesada, R. A. Mathies *Anal. Chem., 64, 2149-2154 (1992).*

13. H. Kambara, S. Takahashi *Nature, 361, 565-566 (1993).*

14. S. Takahashi, K. Murakami, T. Anazawa, and H. Kambara, *Anal. Chem. 66, 1021 (1994).*

15. N. J. Dovichi, J. C. Martin, J. H. Jett, m. trkula and R. A. Keller *Anal. Chem. 56, 348-354 (1984).*

16. N. J. Dovichi and F. Zarrin, *Anal. Chem. 57, 2690-2692 (1985).*

17. S. Hjerten, *J. Chromatogr. 347, 191 (1985).*

18. T. T. Lee and E. S. Yeung, *Anal. Chem. 64, 1226 (1992).*

19. H. T. Chang, *Ph.D. thesis, Iowa State University (1994).*

CHAPTER 3

SIDE-ENTRY EXCITATION AND DETECTION OF MULTIPLE CAPILLARY ARRAY BY GUIDED WAVE FOR HIGH THROUGHPUT DNA SEQUENCING

A manuscript prepared for submission to Applied Spectroscopy

Xiandan Lu and Edward S. Yeung*

ABSTRACT

In high throughput DNA sequencing, efficient coupling of the laser to each capillary is a challenge. Our group developed two multiple point irradiation systems [1,2]. The present work describes a more efficient excitation and detection method in which the laser light propagates through the capillary array without undergoing a serious reduction in power. A 100 capillary array was sandwiched between two quartz plates with an index-matching solution in between, thus forming a planar waveguide. The light was directed into the channel across the capillary array. Two types of capillary (round, 150μm o.d., 75μm i.d. capillary; square, 340μm outer edge, 75μm inner edge capillary) were used to test the optical configuration. DNA sequences of PGEM/U were obtained from both size capillaries. The waveguide cell design for a 100 capillary array is discussed.

INTRODUCTION

The challenge of the Human Genome Project is to increase the sequencing efficiency by 100 times over the existing state of the art in order to complete sequencing of the human and other model organisms' genome by the year 2000 [3]. Several technologies, such as multiple capillary gel electrophoresis (CGE) [4-10], single-molecule detection based sequencing [11-13], sequencing by hybridization [14-17], and sequencing by mass spectrometry [18-31] emerged as a result of this initiative. Among those, CGE technology is expected to play an integral role in achieving the designated goal of the Human Genome Project.

A number of properties of CGE, such as superior heat dissipation and short running time, make CGE a good choice over conventional slab gel methods. The sequencing chemistry, tagging chemistry and separation mechanism are analogous. Use of a new gel matrix [32-36] (linear polymer solution) has overcome problems originally encountered with crossed-linked gels. A capillary is typically replenished with a new matrix for each new DNA sequencing experiment. This new protocol eliminates the possibility of contamination from previous runs. Even though a 25 times increase in speed [37] has been obtained for a single capillary in gel electrophoresis compared to slab gel electrophoresis, running one capillary at a time will not have advantages over slab gel electrophoresis because as many as 24 lanes can be run simultaneously on the commercial slab gel instrument. Therefore, great efforts are focused on developing multiple capillary systems. Several types of

capillary array systems were proposed. In one configuration [4-6], a scanning confocal fluorescence excitation and detection system was developed and demonstrated on a 24 capillary array system. In another configuration [7,8], a multiple sheath flow system was used for a 20 capillary array. Other schemes include a multiple point excitation system [1,2] and a side-entry excitation system [9,10]. Another interesting format is the use of a machined channel [38]. In order to have enough resolution for DNA sequencing, the channels have to be 30-50 cm long. The glass or silicon surfaces of these channels need new surface treatment to avoid electroosmotic flow or other interactions with DNA. The scheme for sample injection and sieving matrix replacement need to be established first [39]. The capillary array system is presently the most mature format for DNA sequencing.

In a previous experiment [9], we developed a side-entry laser irradiation and detection system for multiple capillaries in which 9 capillaries were immersed into water to roughly match the refractive index of the capillaries. The reduction of the laser beam intensity along the laser beam path due to reflection is suppressed because the refractive index difference ($\Delta\eta$) between the boundaries is decreased by the water. Scattered light incident on the detector is reduced because the detection is performed perpendicular to the direction of propogation of the laser light and because the intercapillary refractive index is matched by water. The success of this excitation geometry for a 96 capillary array depends on the packing quality, the refractive index matching of the gel matrix, the fused silica and the surrounding

medium outside the capillaries.

Recently, Anazawa, Takahashi, and Kambara [10] reported a computer simulation for the side-entry excitation and detection system and evaluate of the transmittence as a function of the refractive index (η) (where η of the capillary wall = 1.47, η of the capillary inside = 1.36). For the case where the capillaries are placed in air, the laser beam passes through the cores of all capillaries because each capillary acts as a convex lens focussing the laser light repeatedly when $2<R/r<6$ (R: o.d. of the capillary; r: i.d. of the capillary).

When the capillaries are placed in air, however, the laser power reduction by reflection is significant. The incident light gradually decreases with increasing number of capillaries. The incident laser power is reduced by 7% by each capillary. Capillary 10 will receive about 50% of the laser power of capillary 1. At capillary number 32, the attenuation of the laser power is 10 fold. Higher laser powers are used to achieve good signal-to-noise for the last capillary. However, the dynamic range of the detector is then the limiting factor. The last capillary requires sufficient signal for base calling without signal saturation of the first several capillaries. The array has to be packed well so that the centers of the capillaries lie along one axis.

Here, we report the results for a side entry excitation and detection system where the 100 capillary array is uniformly excited. Two types of capillary (round and square) were tested and sequencing data was obtained. The concept for large capillary array DNA sequencing is introduced.

PRINCIPLE

A computer simulation of the ray transmission through the immersion fluid and capillary array is shown in Figure 1. The whole assembly is considered as a refractive index lens. Since the refractive indices of the capillary wall and the gel are fixed, only the refractive index of the immersion fluid is manipulatable. When the immersion fluid (n_1) has a refractive index lower than that of the capillary wall (n_2), (Figure 1a) the wall of a capillary acts as a convex lens and vice versa (Figure 1c). There is no lens effect from a capillary wall when $n_1 = n_2$ (Figure 1b). The inside of the capillary (n_3) will always act as a concave lens because $n_3 < n_2$. Since capillaries with 150 μm o.d. and 75 μm i.d. (round) are beneficial to throughput and gel filling, only this format is considered. According to the equation. of ref. 10:

$$\Delta\theta = 2\left\{-\sin^{-1}\left(\frac{x}{R}\right) + \sin^{-1}\left(\frac{n_1 x}{n_2 R}\right) - \sin^{-1}\left(\frac{n_1 x}{n_2 r}\right) + \sin^{-1}\left(\frac{n_1 x}{n_3 r}\right)\right\} \qquad (1)$$

The refraction angle ($\Delta\theta$) changes with x and n_1 with fixed R/r, n_2, and n_3, R and r are the outer and the inner radii of the capillary respectively. The relationship between the refraction angle and x at 3 different n_1 values is shown in Figure 2. The x increases with increasing refraction angle. That is why $\Delta\theta$ for ray 2 is larger than that of ray 1 in all three cases of Figure 1. When $x < 1$, $\Delta\theta$ increases with n_1. When $x > 1$, rays will not enter the inside of the first capillary of the array. In the case where water ($n_1 = 1.333$) is used as the immersion fluid, rays which pass through the region where $r < x < R$, are focused onto the other capillaries. If those rays pass

through the cores of the capillaries afterwards, they still are refracted. In the case where glycerol (n_1 = 1.456) is used as the immersion fluid, the laser beam passes directly through without any obvious refraction. In the case where Cedarwood oil (n_1 = 1.513) is used as the immersion fluid, light diverges. In either of the three cases, the laser beam is not transmited to all of the cores of the capillary array because of refraction. In order to trap the laser power within the capillary array, two plates (glass or quartz) were used to provide a total internal reflection geometry whereby all of the light is trapped between the two plates.

Total Internal Reflection

To determine the direction of the propagation of light at the interface of two dielectric media, it is necessary to consider the refractive index difference between two media. The refractive index is defined as the ratio of the velocity of light in a vacuum to the velocity of light in that medium. A ray of light travels more slowly in an optically dense medium than in one that is less dense. The refractive index provides a measure of this effect. When a ray is incident on the interface between two dissimilar dielectrics media, refraction occurs, as illustrated in Figure 3-3. The angles of incidence ($\varnothing_2$) and refraction ($\varnothing_1$) are related to each other and to the respective refractive indices by Snell's law:

$$n_1 \sin \varnothing_1 = n_2 \sin \varnothing_2$$

The small amount of light reflected back into the originating dielectric medium depends on the difference in the two refractive indices, the angle of incidence and

Figure 1. Schematic diagram of the lens effect of a capillary. Refraction angle, $\Delta\theta$; R and r are the outer and inner radii of the capillary; x is the distance between the center of the capillary and the incident laser beam; n_1, n_2, n_3 are the refractive indices of immersion fluid, capillary wall and gel matrix, respectively.

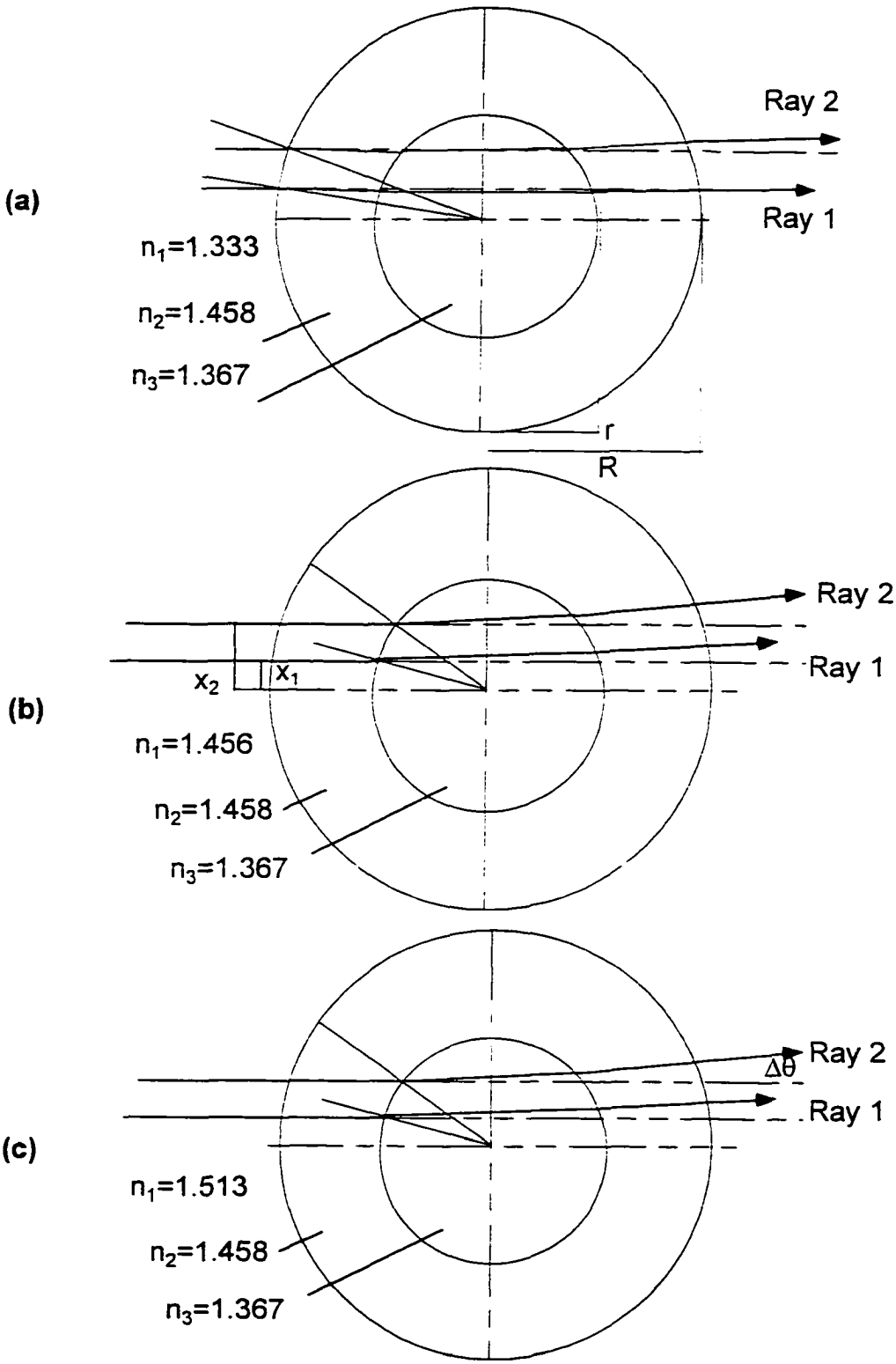
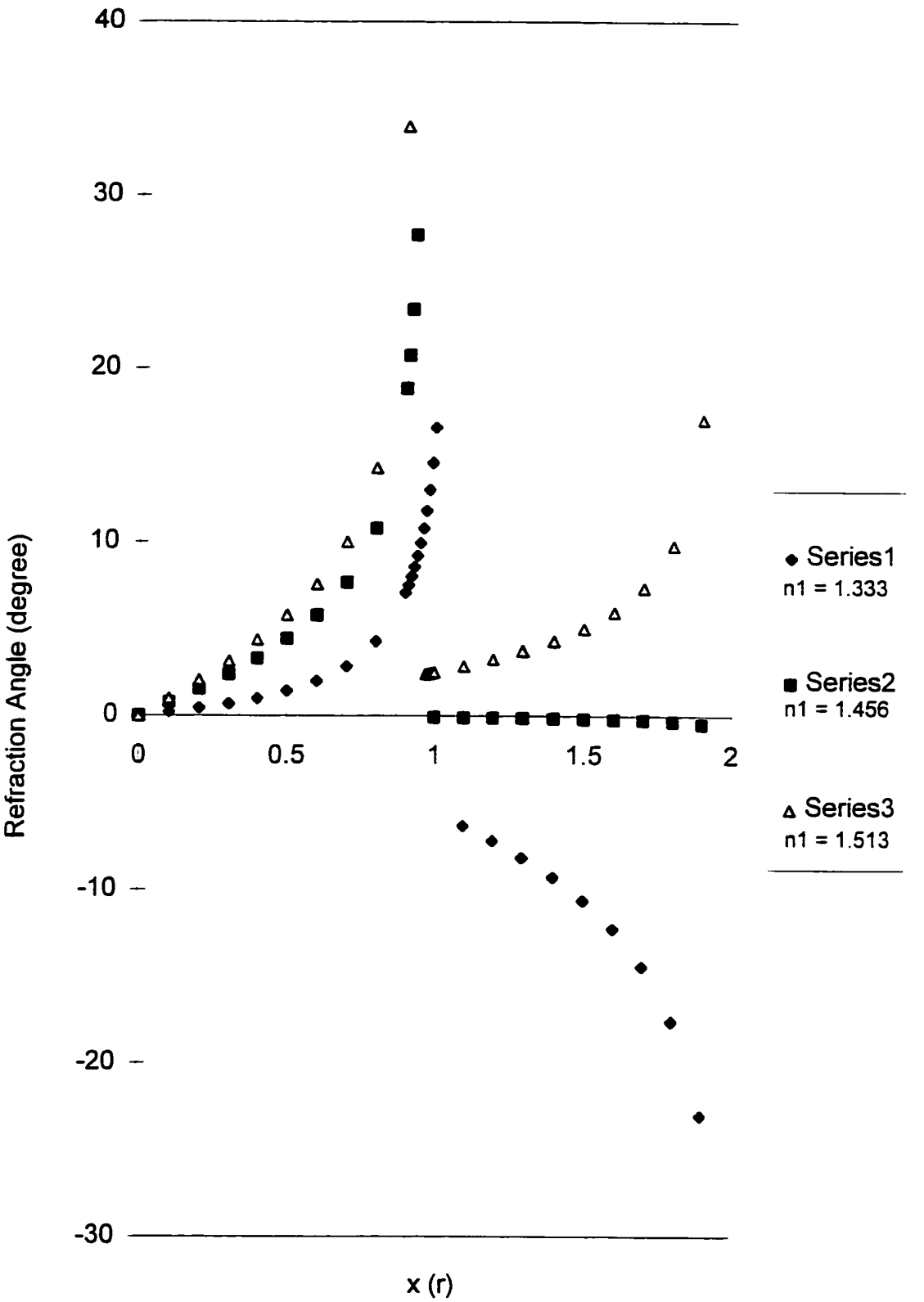

(a)
Ray 2
Ray 1
n_1=1.333
n_2=1.458
n_3=1.367
r
R
(b)
Ray 2
Ray 1
x_2
x_1
n_1=1.456
n_2=1.458
n_3=1.367
(c)
Ray 2
Ray 1
Δθ
n_1=1.513
n_2=1.458
n_3=1.367

Figure 2. Relationship between refraction angle and x (the distance from the center of the capillary to the incident laser beam) with three different immersion fluids, water, n = 1.333; glycerol, n = 1,456; Cedarwood Oil, n = 1.513.


40
30
20
10
0
-10
-20
-30
0
0.5
1
1.5
2
Refraction Angle (degree)
x (r)
Series1
n1 = 1.333
Series2
n1 = 1.456
Series3
n1 = 1.513

the polarization of the light (Figure 3a). Since $n_2 > n_1$, refraction angle is greater than incidence angle. When the sin $\varnothing_1$ reaches 1, $\varnothing_1$ is 90°. The refracted rays merge parallel to the interface of the two media (Figure 3b). This is the limiting case of the refraction and the angle of incidence is known as the critical angle $\varnothing_c$, and the value of the critical angle is given by:

$$\sin \varnothing_c = n_2 / n_1$$

At angles of incidence greater than the critical angle (Figure 3c), the light is reflected back into the originating medium with high efficiency (around 99.9%).

EXPERIMENTAL SECTION

Capillary Array Planar Waveguide Preparation

Capillaries for each experiment (Polymicro Technologies, Phoenix, AZ, model TSP075150, 75μm i.d., 150μm o.d.) were all cut from the same batch of capillary. The locations of the detection windows were marked. The capillaries were sandwiched between two thin cover slips side by side (Corning #1, Corning Inc., Corning, N.Y.) with the marks aligned and fixed on a capillary holder. The top and bottom cover slips were cut to form a window to expose the capillaries to an excimer laser (Questek, Boston, MA model, 2560νβ operating at 308nm wavelength). The coating of the capillaries was removed. Two pieces of 3M tape (3M Scotch Brand Tape, St. Paul, MN) were used to fix the orders of the capillaries. The array was transferred to the top of a 1" wide, 2" long and 1mm thick fused silica plate as the

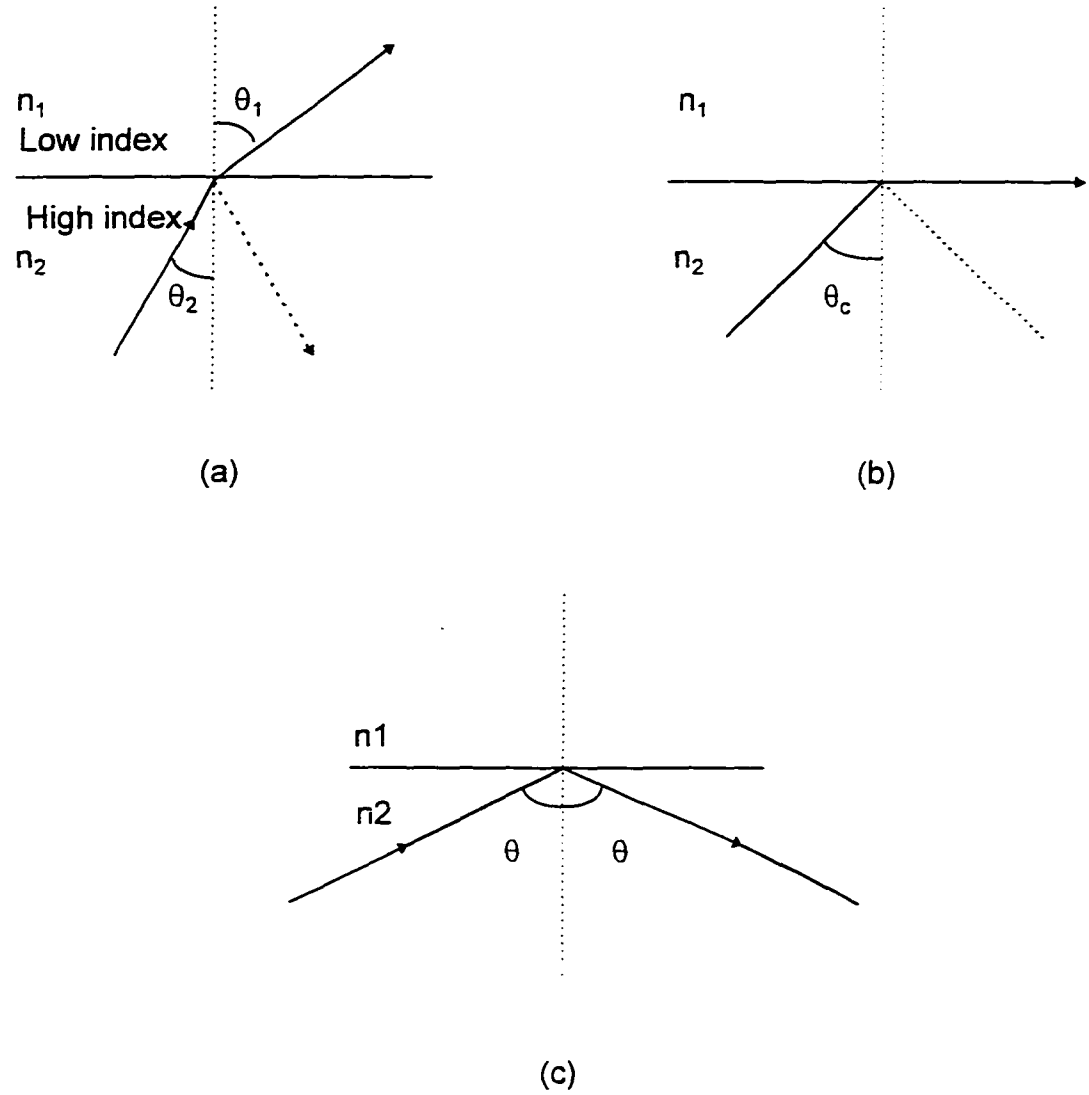


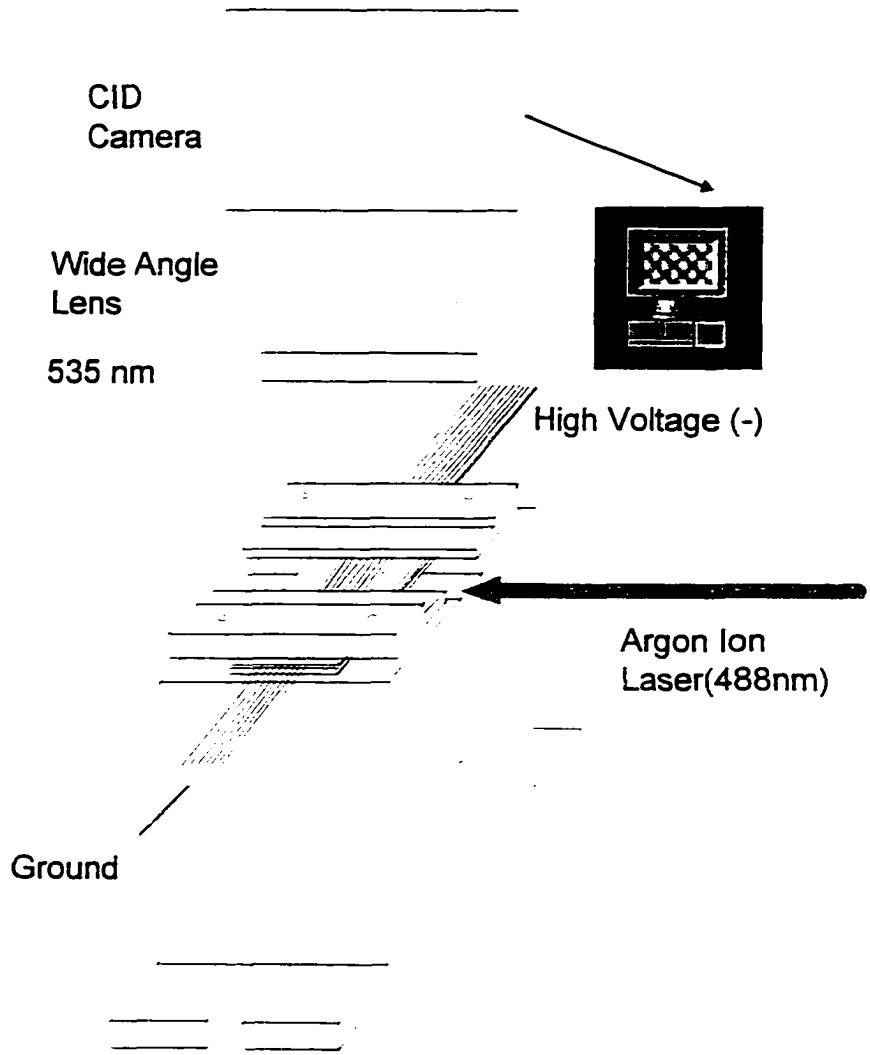
Figure 3. Light rays incident on high to low refractive index interface, $n_1<n_2$: (a) refraction; (b) the limiting case of refraction showing the critical angle; (c) total internal reflection where $\theta>\theta_c$.

bottom plate (ESCO Products, Oak Ridge, N. J.). Two pieces of 1" wide, 1 cm long and 1mm thick glass were placed above the capillary array, with the window region open. The space between the two plates and the capillary array at each side of the window were sealed with optical glue #61 (Edmund Scientific Co., Barrington, NJ). After exposure to UV-light for several hours, the UV cement cured and the array became fixed between the two plates (Cole-Parmer Instruments Co., Chicago, Illinois, model 9815-series lamps). Two small thin fused silica plates (1 mm) were attached to the bottom plate along the window region to provide an optical interface for laser entry and to form a shallow well to hold the immersion fluid. Four spacers were placed at the corners of the well to define the thickness of the waveguide. The window region was covered with a fused silica plate and entire cell was fixed by two small plexiglass clamps (Figure 4).

Optics

The schematic diagram of the fluorescence based multiple capillaries electrophoresis system is shown in Figure 4a. The details of the window region is shown in Figure 4b. An argon ion laser (Coherent, Santa Clara, CA, Model Innova 90) operating at 488 nm was used for excitation. Several different lens (Edmund Scientific Co., Barrington, NJ) combinations were used to launch the laser beam into the waveguide. Laser power was measured using a power meter (Melles Griot, Irvine, CA, Model 13PEM001). Fluorescence from the capillaries was detected by a charge injection device (CID) (CID Technologies Inc., Liverpool, NY) camera.

Figure 4. (a) Schematic diagram of the fluorescence based excitation and detection system for multiple capillary electrophoresis. (b) Detail diagram of the window region of the waveguide.


CID
Camera
Wide Angle
Lens
535 nm
High Voltage (-)
Argon Ion
Laser(488nm)
Ground

(a)

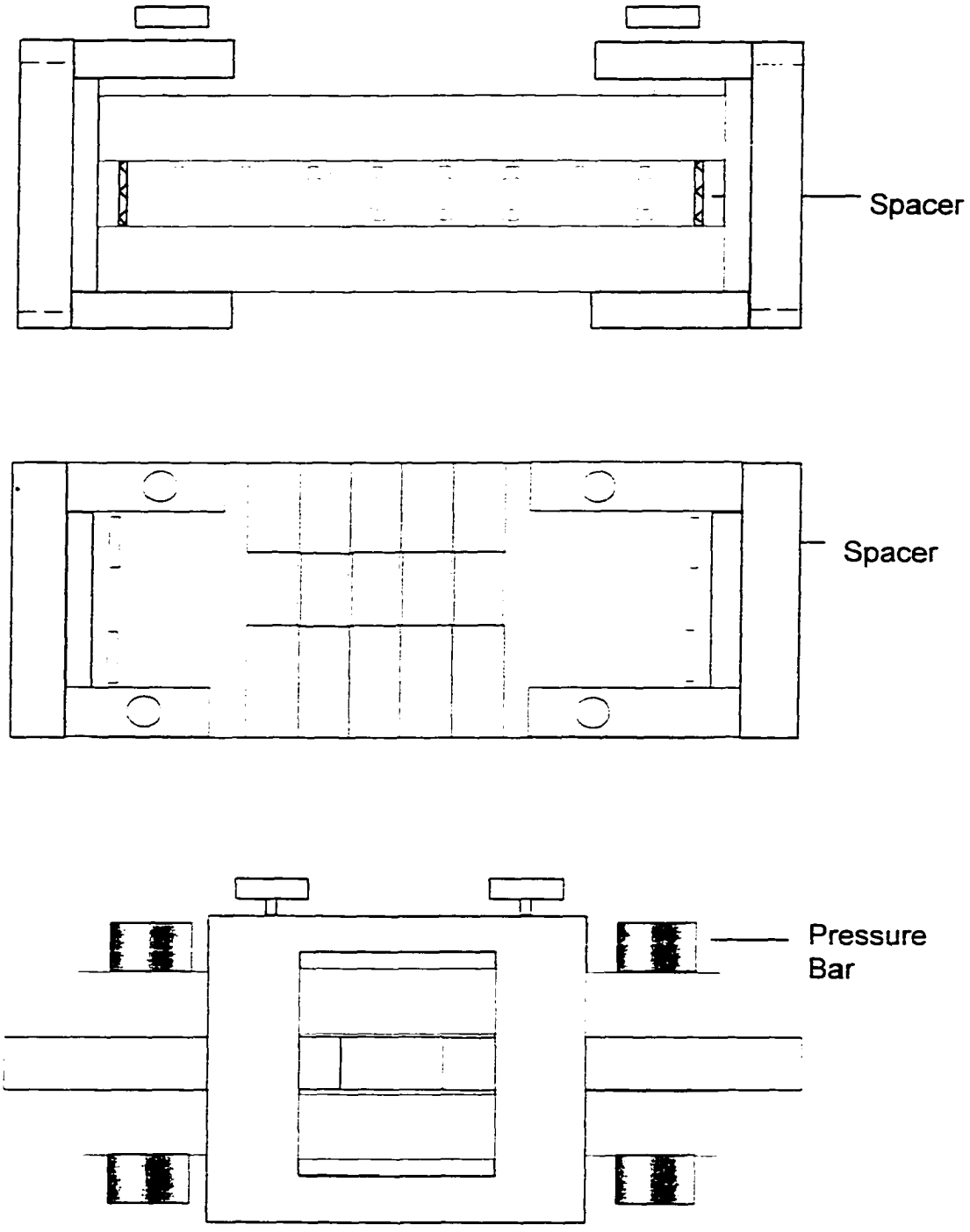


Figure 4 (continued): (b)

A wide-angle camera lens (Nikon, Japan, Model Nikkor 28mm f1.4D AF) was attached to the CID to provide a large solid angle for efficient detection. Cutoff filters (Edmund Scientific Co., Barrington, NJ) were attached in front of lens to eliminate the excitation light.

Refractive Index Measurements

A refractometer BAUSCH&LOMB (Fisher Scientific Corp., Pittsburgh, PA) was used to measure the refractive indices of the different concentrations of gels and to determine the proper composition of the buffer system such that it has the same refractive index as the gel used for DNA sequencing. Refractive indices vary slightly with the temperature. Thus, temperature is maintained at 25°C with a water bath for all experiments. For refractive index measurements, gels or solutions were sandwiched between two glass plates. The two glass plates were cleaned with ethanol between each measurement. Distilled water was used to calibrate the refractometer. A refractive index of 1.3335 was found, which compares well with the literature value of 1.3330 for room-temperature water [40]. Three different gel compositions, 1.5% 8,000,000, 1.4% 600,000 PEO in 1xTBE with 3.5 M urea (gel #1), 1.8% 8,000,000, 1.4% 600,000 PEO in 1xTBE without urea (gel #2) and 2.0% 8,000,000, 1.4% 600,000 PEO in 1 TBE with 3.5 M urea (gel #3) were tested. Two batches of each gel were tested twice. The results are shown in Table 1. The results show little variation within the same batch of gel, but variations occurred between batches. This may be due to system errors (balance, pipette) or different stirring

Table 1. Refractive Indexes of Three Different Gels

	#1		#2		#3	
Batch 1	1.3665	1.3660	1.3378	1.3375	1.3682	1.3680
Batch 2	1.3658	1.3655	1.3372	1.3374	1.3672	1.3674

periods and speeds during the gel preparation process, which can cause changes on the composition of the gel. Urea gel has a higher refractive index than nonurea gel. The refractive index increases with the gel concentration.

In order to simulate all the parameters in a real sequencing case, the buffer for the dye solution was prepared to match the refractive index of the gel that is used for sequencing. The results reported here were obtained with gel #1. The buffer solution used was 0.05 TBE with 30% formamide (Table 2).

Uniformity Measurements of the Signals from the 100 Capillary Array

Aqueous 10^{-3} M fluorescein (Sigma Chemical Co., St. Louis, MO) was used as the stock solution. Buffer was prepared from a 20-times dilution of the stock solution on 1xTBE (89 mM Tris., 89 mM boric acid, 2 mM EDTA) by 30% formamide in water (pH=7.8). The fluorescein solution was diluted with buffer to the desired concentration. The capillary array was separated into bundles which were jacketed with suitably-sized PEEK tubing and the spaces between capillaries and PEEK tubing were sealed by 5 minute epoxy (True Value, Chicago, IL Model TruBond).

Table 2. Adjustment of the Refractive Index of Buffer with Formamide

	0.05 TBE
0% formamide	1.3330
25% formamide	1.3615
30% formamide	1.3675
50% formamide	1.3900
100% formamide	1.4449

Bundled capillaries were sequentially flushed with buffer, fluorescein solution, and buffer. The fluorescence signal from each capillary was obtained by subtracting the background from the signal. A 488nm Raman edge filter and a 535nm long pass filter were used to block the laser light.

Detection of PGEM/U DNA Fragments from the Capillary Array

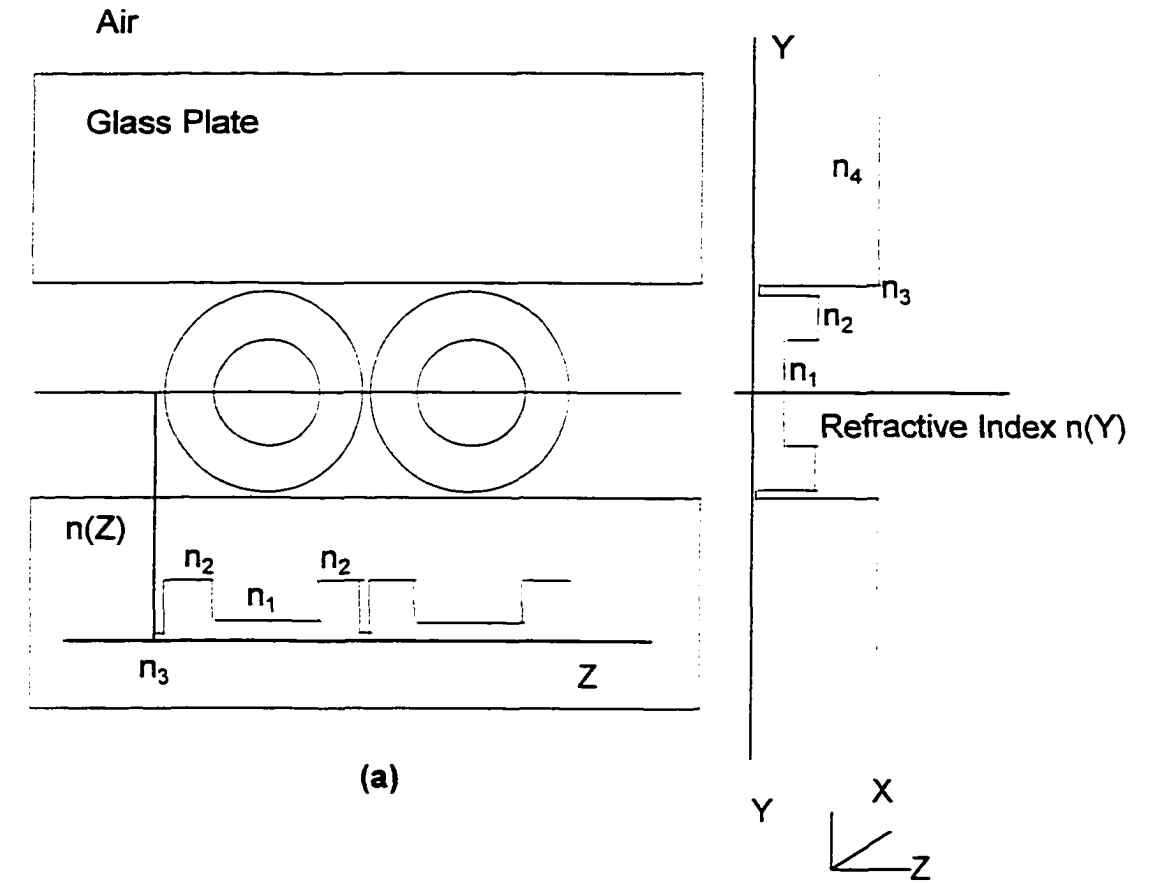
PGEM/U DNA was used as the standard DNA sample to test the integration system of excitation, detection, sample handling and injection. A 2.0μl denaturing solution (the mixture of EDTA and formamide) was added to the sample vial. The DNA was denatured by heating at 95°C for 3-4 minutes. The running buffer was 1xTBE (89 mM Tris., 89 mM boric acid, 2 mM EDTA) with 3.5 M urea. The polymer matrix was a solution of 1.5% 8,000,000 and 1.4% 600,000 PEO. The parameters

such as laser power, exposure time, aperture size of the camera lens, electric field strength for injection and separation, and the number of integrated channels used for detection is described for each experiment.

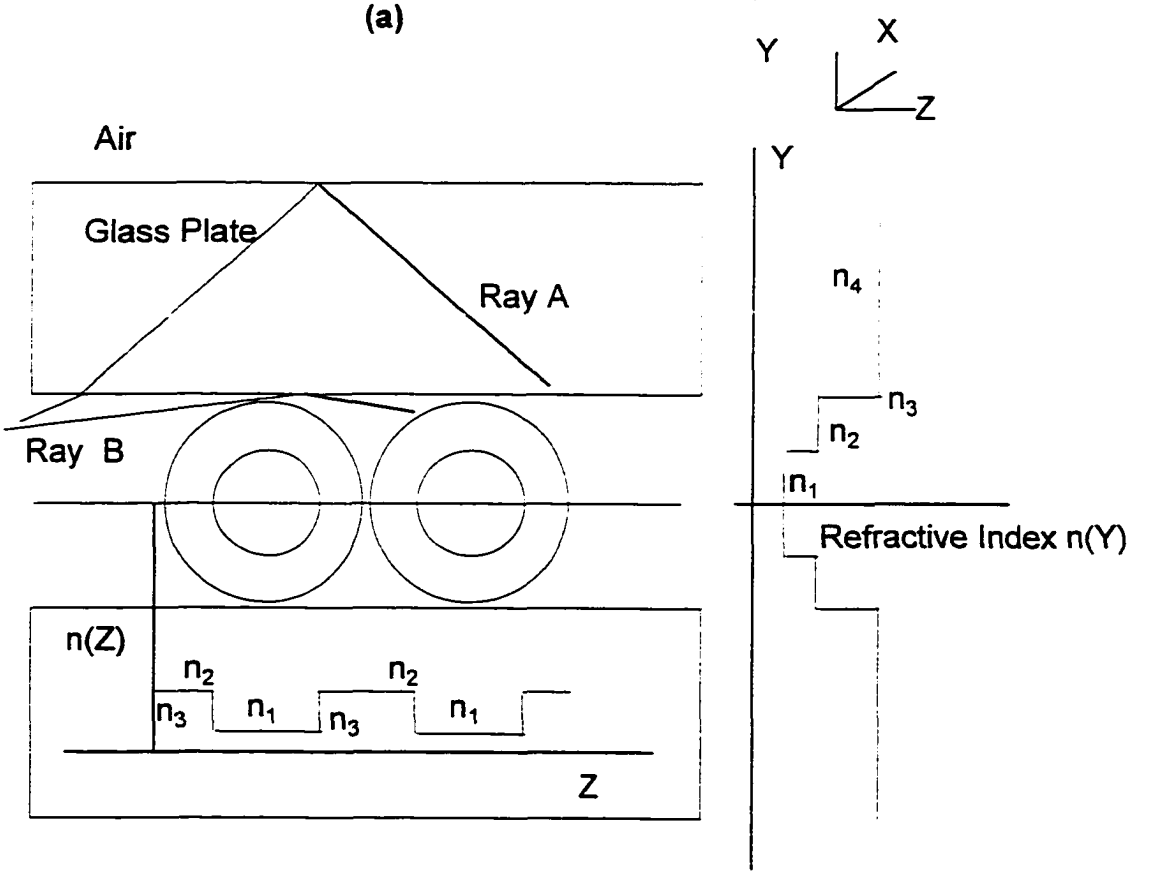
RESULTS AND DISCUSSION

Two waveguide modes for the capillary arrays were tested. In the first mode, the capillary array was sandwiched between two thin cover slips (number 1 Corning cover glass, 150 μm thick) with immersion fluid in between. Two different immersion fluids, water and immersion oil (Zeiss, Germany, n=1.456) were tested. The periodic change of refractive index along the laser transmission (Z-direction) and in the Y-direction when water is used as the immersion fluid is shown in Figure 5a. The signal decreases along the array was observed by eye through a 535nm cutoff filter while $5x10^{-8}$ M fluorescein was flushed through the 100 capillary array with 6 mW laser excitation. The refractive index distribution and the possible ways to trap rays when glycerol is used as the immersion oil are shown in Figure 5b. Lens combinations were tested to observe the effect of laser power density distribution on the capillary array. The beam profile for a 10cm lens [9] shows that the beam diameter is maintained below 150 μm over a range of 1.5 cm. Therefore, a 10 cm lens was first focussed at the center of the array. A 7 times attenuation was produced (Figure 6) even after excluding the first several capillaries. By focusing the laser beam onto the end of the array (Figure 7), the attenuation decreased to

Figure 5. Refractive index profile in the first waveguide mode. $n_1 = 1.365$, $n_2 = 1.458$, $n_4 = 1.523$. The critical angle of an air-glass interface is 41°. (a) capillaries in water, $n_3 = 1.333$; (b) capillaries in glycerol, $n_3 = 1.456$. Ray A, total internal reflection at the glass-air interface; Ray B, grazing angle reflection.


Air
Glass Plate
n(Z)
n2
n1
n2
n3
Z
Y
n4
n3
n2
n1
Refractive Index n(Y)

(a)


Y
X
Z
Air
Glass Plate
Ray A
Ray B
n(Z)
n2
n3
n1
n2
n3
n1
Z
Y
n4
n3
n2
n1
Refractive Index n(Y)

(b)

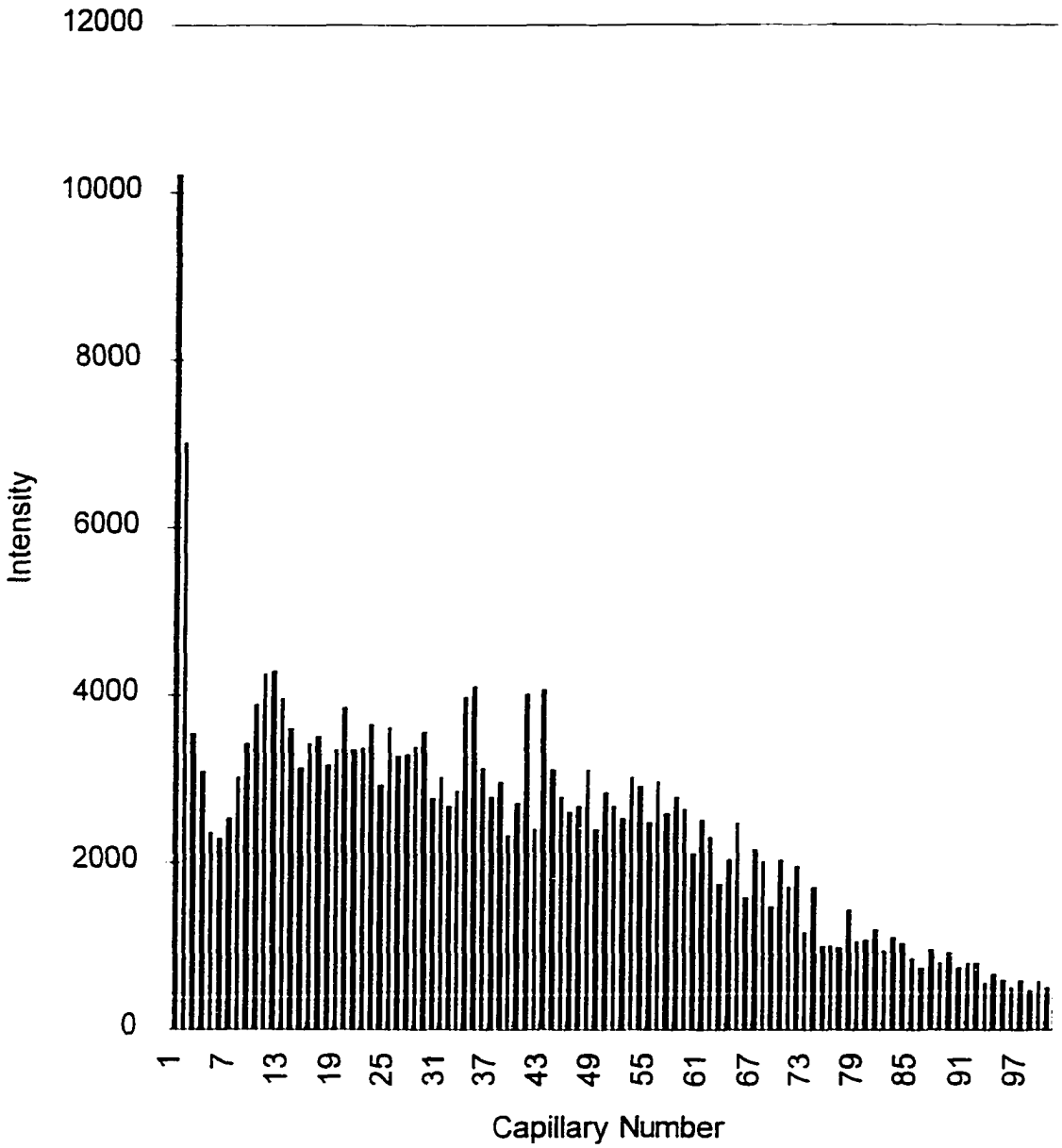


Figure 6. Fluorescence signal distribution on a 100 capillary array for $5x10^{-8}$ M fluorescein in 0.05xTBE. Laser, 6 mW with a 10 cm lens focusing at the center capillary of the array; both top cover glass plate and bottom glass plate (n=1.523), 150 μm thick; exposure time, 0.2 s; lens aperture, 2.8; immersion oil, n_D=1.456.

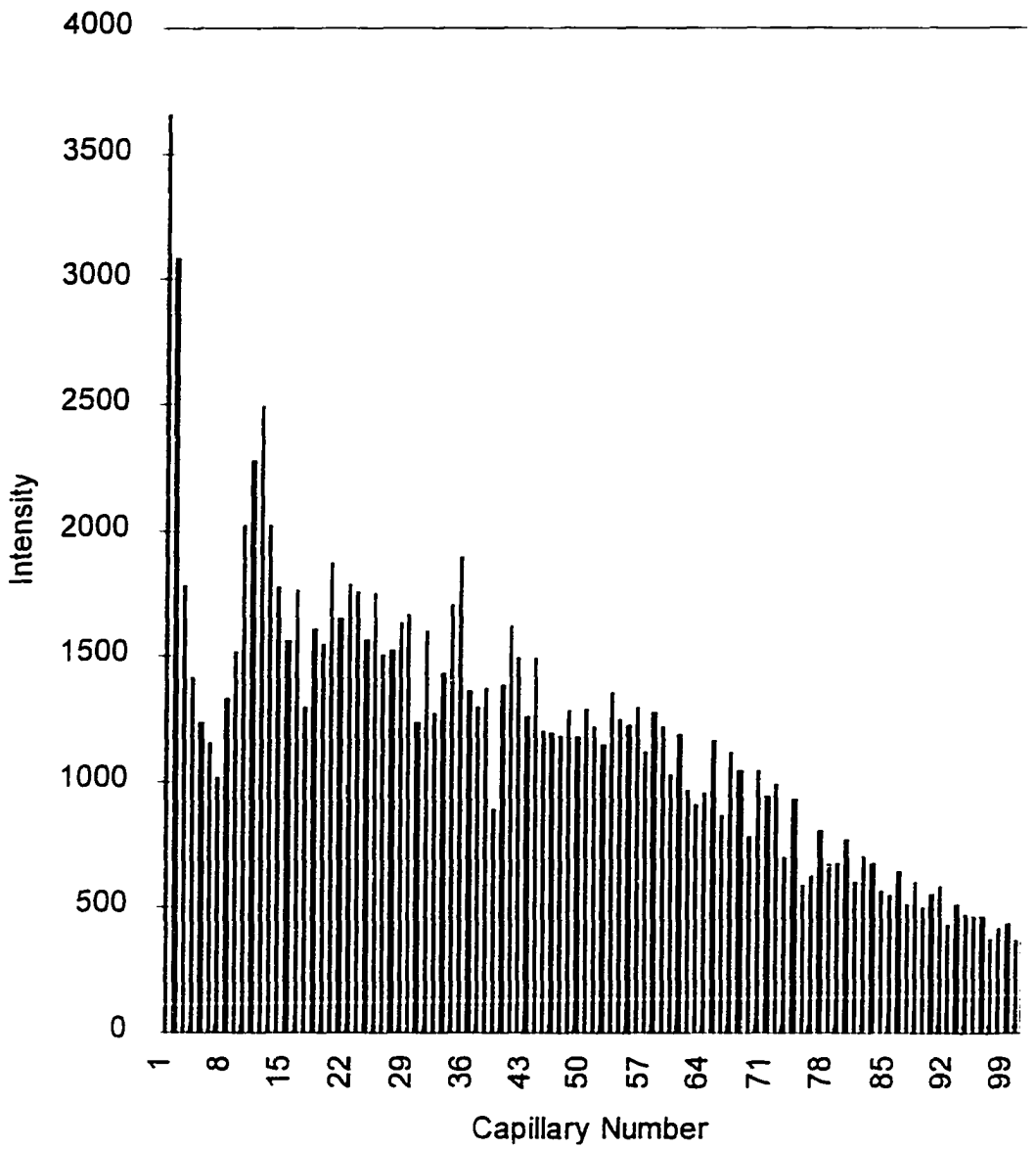


Figure 7. Fluorescence signal distribution on a 100 capillary array for $5x10^{-8}$ M fluorescein in 0.05 TBE. Laser, 6 mw with a 10cm lens focusing at the last capillary of array; other conditions are the same as for Figure 6.

less than 5 times. The capillaries functioning as concave cylindrical lenses change the focusing only along the y-direction (Figure 5b). The distribution benefits from having a higher laser power density incident on the last capillary in the x-direction than on the first capillary by focusing the beam at the end of the capillary array. The signal level decreased because the beam size is larger than the entrance to the waveguide.

In the planar waveguide system, the aperture is limited along the y-direction. In order to eliminate these losses, the laser beam is focused separately by two different cylindrical lenses. A 10cm focal length cylindrical lens is favorable for focusing at the center of the 100 capillary array along the x-direction. This leads to a uniform laser power density along the array. The second lens effect for signal distribution which controls the y-direction has been tested. The result of using a 5cm cylindrical lens in which the laser beam was focused at the entrance of the waveguide is shown in Figure 8. The result of a 30cm focal length cylindrical lens in which the laser beam was focused at the entrance of the waveguide is shown in Figure 9. The signal distributions are similar.

In the waveguide mode, when the beam is incident on the first capillary, it diverges. Only a part of the power passes through the following capillaries while the remainder reaches the air-glass interface. If the incident angle of light reaching the interface is larger than the critical angle, the beam is reflected back (like ray A in Figure 5b) to excite other capillaries. That is why the signal level keeps decreasing

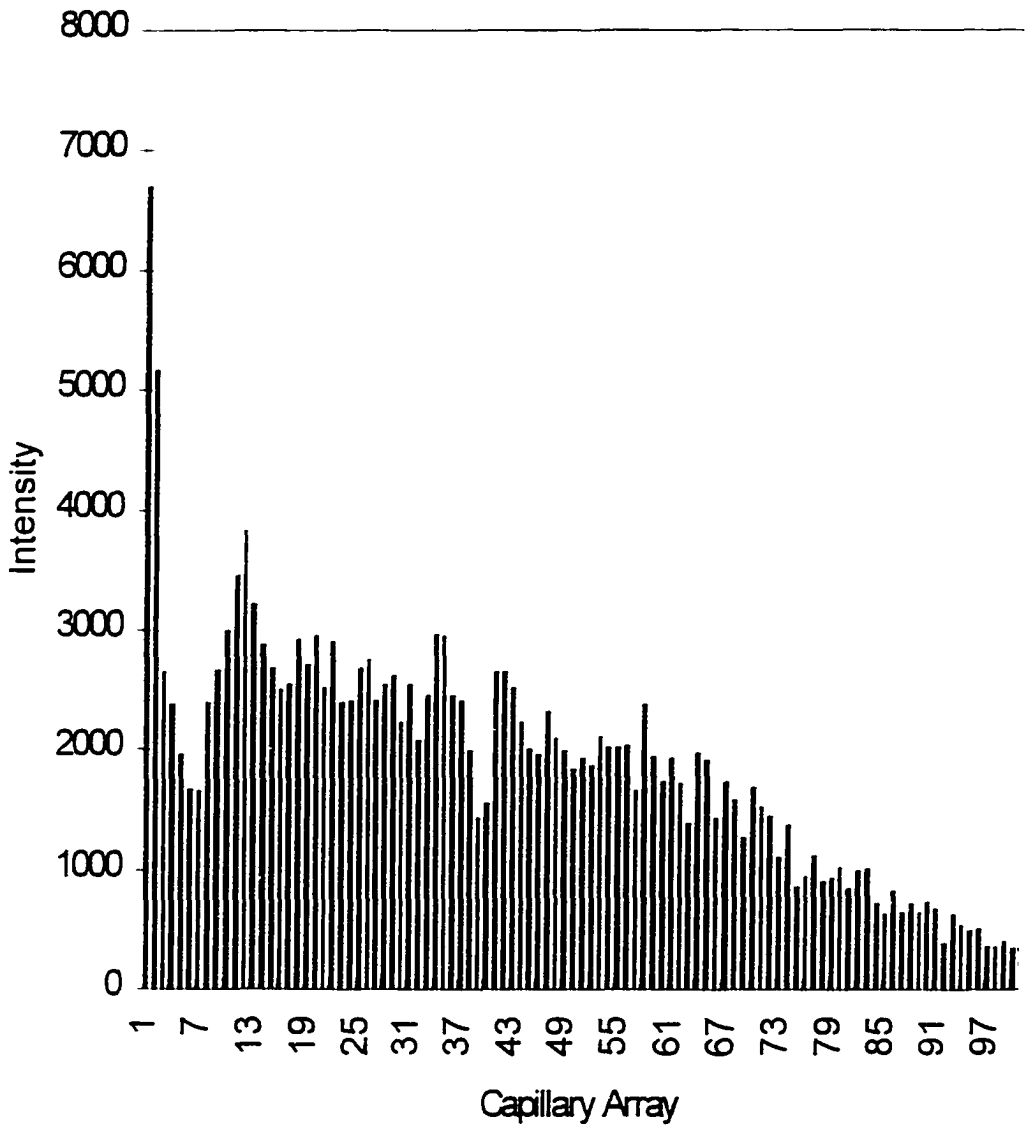


Figure 8. Fluorescence signal distribution on a 100 capillary array for $5x10^{-8}$ M fluorescein in 0.05 TBE. Laser, 6 mw with a 10 cm cylindrical lens focusing at the center of the capillary array along the x-direction, and another 5 cm cylindrical lens focus at the entrance of the waveguide along the y-direction; other conditions are the same as for Figure 6.

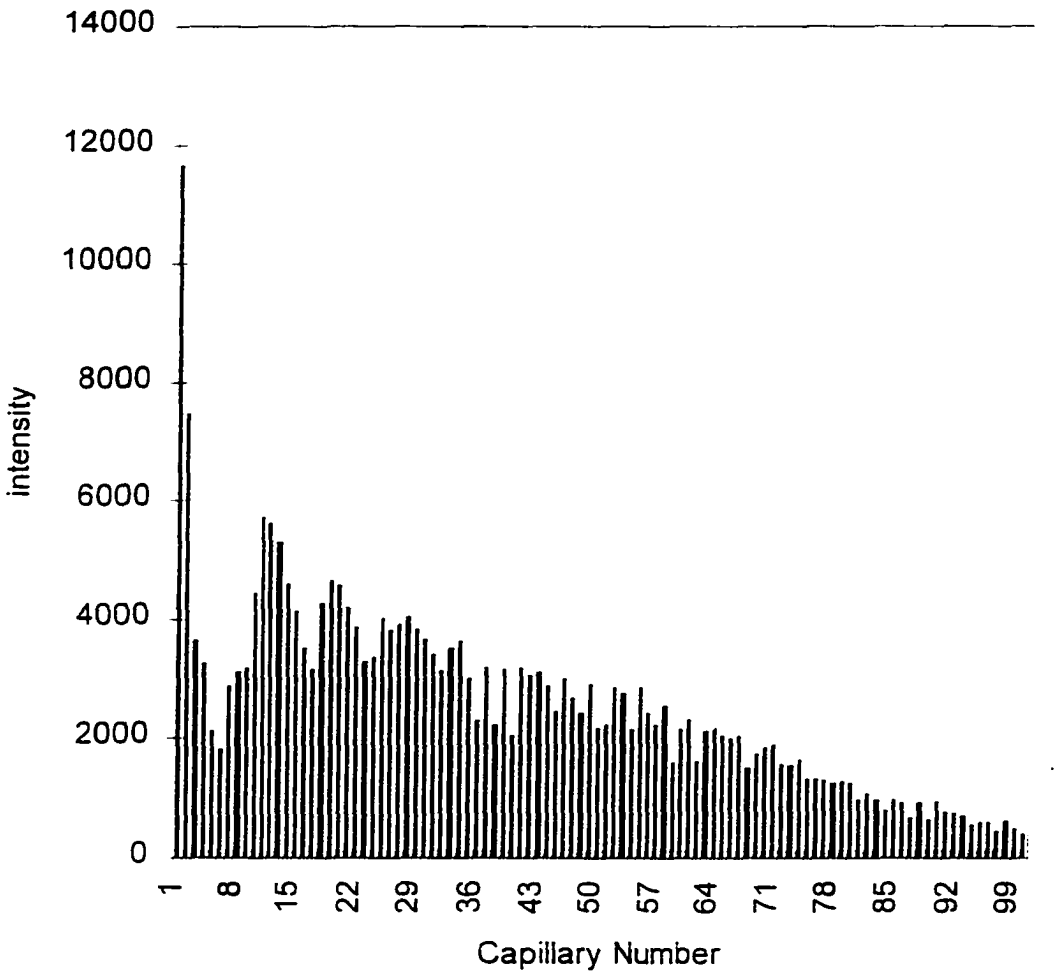


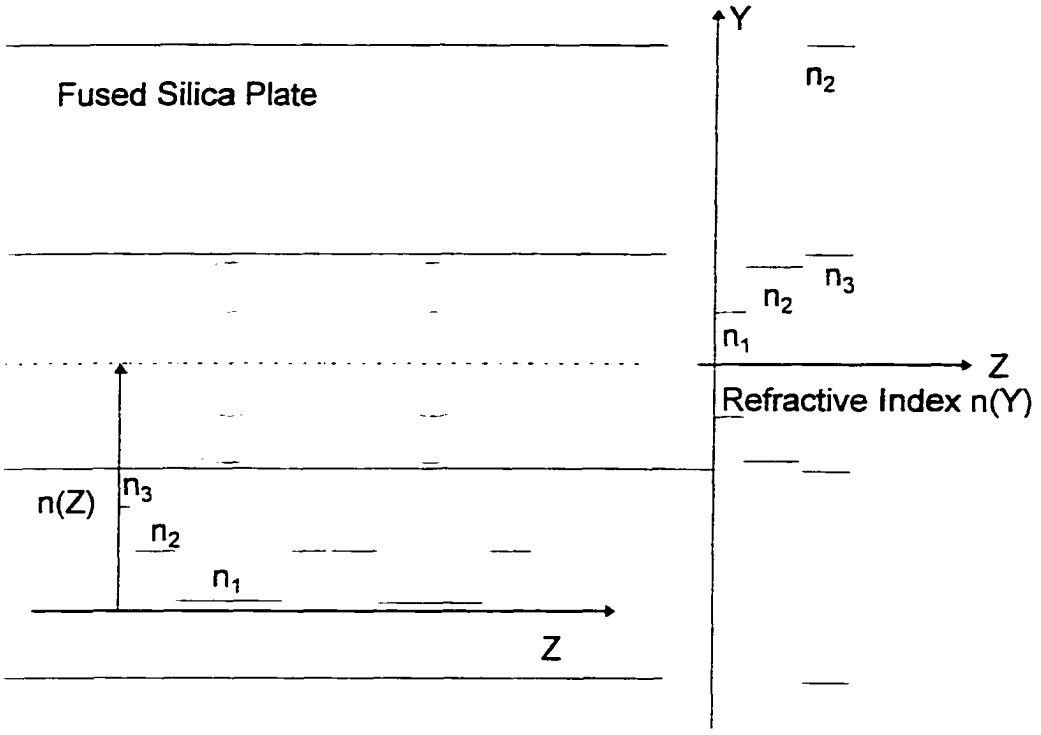
Figure 9. Fluorescence signal distribution on a 100 capillary array for $5x10^{-8}$ M fluorescein in 0.05 TBE. Laser, 6 mw with a 10 cm cylindrical lens focused at the center of the capillary array along the x-direction, and another 30cm cylindrical lens focused at the entrance of the waveguide along the y-direction; other conditions are the same as for Figure 6.

in the following several capillaries and comes back from the total internal reflection. The larger the divergence of the beam, the larger part of the beam will transmit through the glass without coming back.

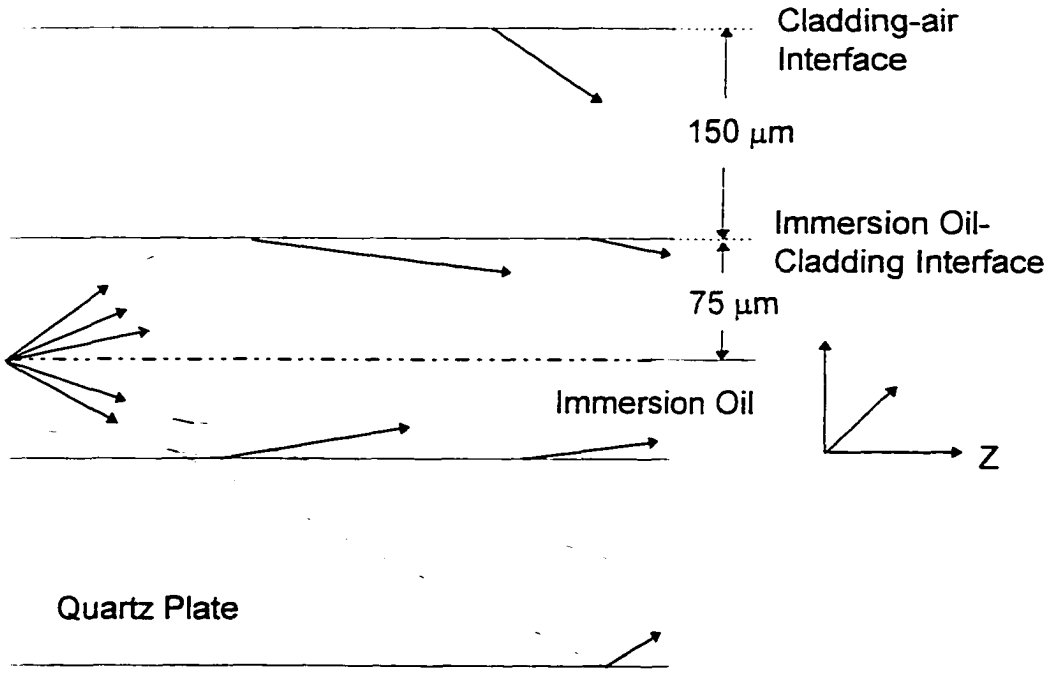
Regardless of which lens combination is used, the attenuation of the laser power along the array is large (5-8 times). The structure of the waveguide itself plays the dominant role. In this mode, there is no solid cladding protection for the waveguide. Foreign matter [41] accumulated on the guide wall degrades the reflection process by interaction with the evanescent field. Waveguides have large losses with air as cladding. The other major sources for attenuation in this mode are material absorption and refractive index variation along the ray pathway. The core of the capillary has a lower refractive index. The structure is similar to the planar waveguide with periodic bubbles. In addition, conventional melting techniques show the homogeneity of one piece of glass is not guaranteed. The batch to batch, piece to piece variation is unpredictable. The absorption of transition metal element impurities, such as Fe (peak wavelength: 400nm) and Mn (peak wavelength: 460nm) in glass, have strong absorption at 488nm wavelength, provide a large background, which limits the laser power available for a excitation and degrades the signal to noise ratio. All these effects eliminate glass as a waveguide material. Instead, quartz plates or fused silica plates, which have lower fluorescence background, are chosen for the second mode.

In the second waveguide mode (Figure 10a), an immersion oil with higher

Figure 10. (a) The Refractive index profile in the second waveguide mode. n_1=1.367, n_2=1.458, n_3=1.513, n_4=1.458. The critical angle, at the immersion oil-cladding interface, 74.5°; at the cladding-air interface, 43.3°. (b) multimode ray transmission.


Fused Silica Plate
Y
n_2
n_3
n_2
n_1
Z
Refractive Index n(Y)
n(Z)
n_3
n_2
n_1
Z

(a)


Cladding-air
Interface
150 μm
Immersion Oil-
Cladding Interface
75 μm
Immersion Oil
Z
Quartz Plate

(b)

refractive index (Fryer, Huntley, Illinois n_D=1.513 at 25°C) was used. A quartz plate which has a lower refractive index (n_D=1.458 at 25°C) than the immersion oil functions as the cladding. The thinnest plate available is 150μm. This cladding thickness is more than sufficient to allow the evanescent field to decay to a minimum value and to prevent losses from the penetrating energy. We use the waveguide setup, which is different from any of the traditional waveguide systems due to the fact that the refractive index periodically changes along the ray transmission. The light transmission process along our waveguide is similar to light propagation through a multimode fiber and is illustrated below (Figure 10b). The rays that pass immersion oil-quartz interface and reflect back from quartz-air interface within 15 cm range for a 1000 capillary or a 1.5 cm range for 100 capillary array will still count as the transmission modes.

The fluorescence signal distribution experiment was performed on a 100 capillary array using the second waveguide system. The distribution (Figure 11) is much more uniform than that of the first type of waveguide system. This suggests that the light trapping efficiency of the second waveguide system is higher than that of the first type. The variation of signal from the capillary array is less than 2 times.

Two experiments were performed to test the feasibility of this excitation scheme for a large array DNA separation. The separation and detection of PEGM/U in 10 consecutive capillaries is shown in Figure 12. The results of PEGM/U fragments separation in 20 random capillaries among the 100 capillaries of the array

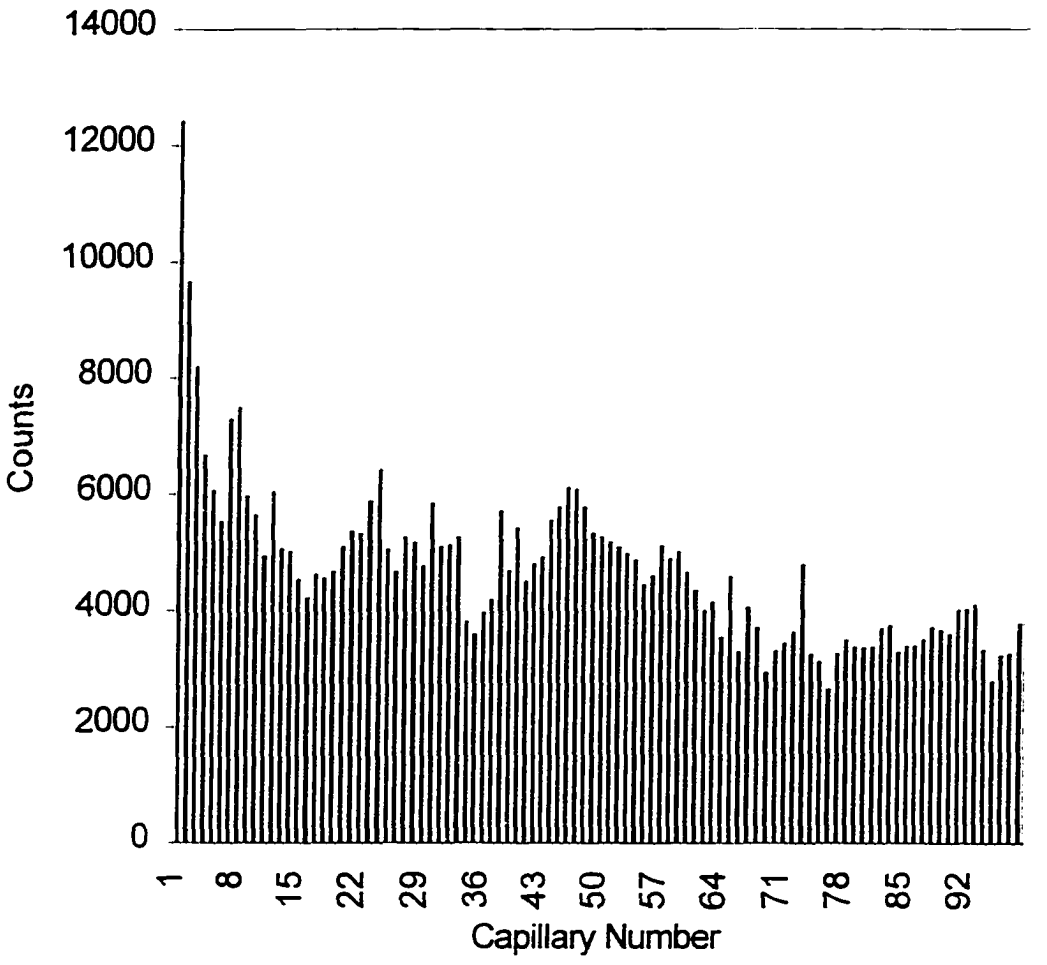
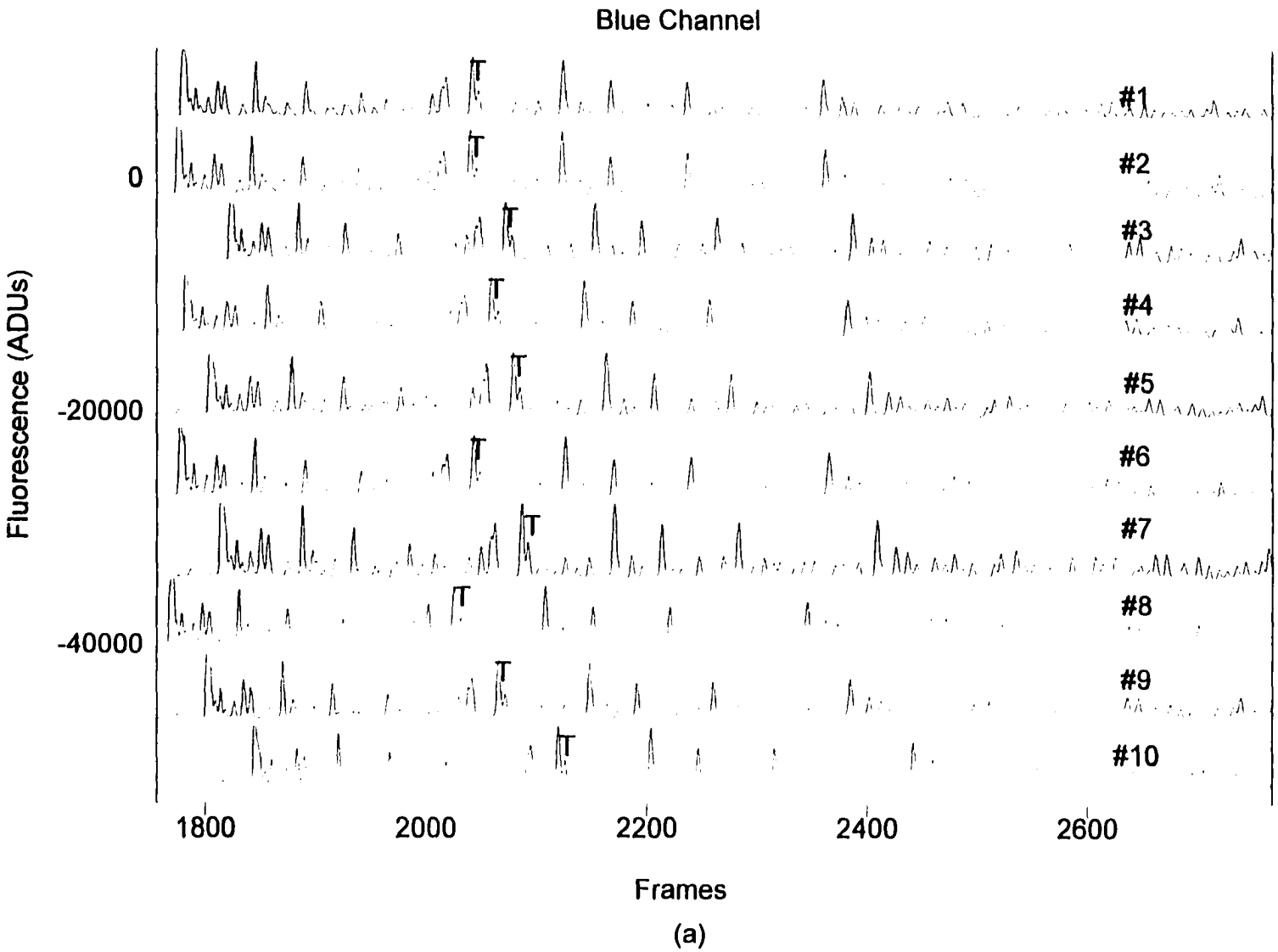


Figure 11. Fluorescence signal distribution on a 100 capillary array for $5x10^{-8}$ M fluorescein in 0.05 TBE with 30% formamide. Laser, 6 mw with a 10 cm lens focused at the center of the capillary array; top fused silica cover plate, 150 μm thick; bottom fused silica plate, 1 mm; exposure time, 0.2 s; lens aperture, 2.8; immersion oil, n_D=1.513.

Figure 12. The background subtracted electropherograms of DNA in 10 closely packed capillaries. The 488-nm laser output at 125 mW was used for excitation. A 488nm Raman edge filter (Kaiser Optical Systems, Ann Arbor, MI) was attached to the camera lens (Nikon F1.4/28mm, aperture 2.0). Another RG610 glass filter was tilted to split the image into two channels. The difference in the optical pathlengths for two channels was balanced with a quartz plate. Ten capillaries (65cm total with 50cm efficient length) were bundled with one piece of PEEK tubing for injection. A 2μl denaturing solution (the ratio of EDTA to formamide=1:5) was added to each sample vial. Three sample tubes were combined into one for the 10 capillary bundle sample introduction. Samples were introduced by electrokinetic injection for 35s at 250 V/cm. A 10 minutes resting period was allowed before data acquisition was initiated. The separation was performed at 250 V/cm. The exposure time gradient began 10 minutes after the initiation of data collection. The initial exposure time was 200 ms and then increased by 300 μs per frame until the finish of the run. The final exposure was 1000 ms. (a) blue channel; (b) red channel.


Blue Channel
Fluorescence (ADUs)
0
-20000
-40000
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
1800
2000
2200
2400
2600
Frames
(a)

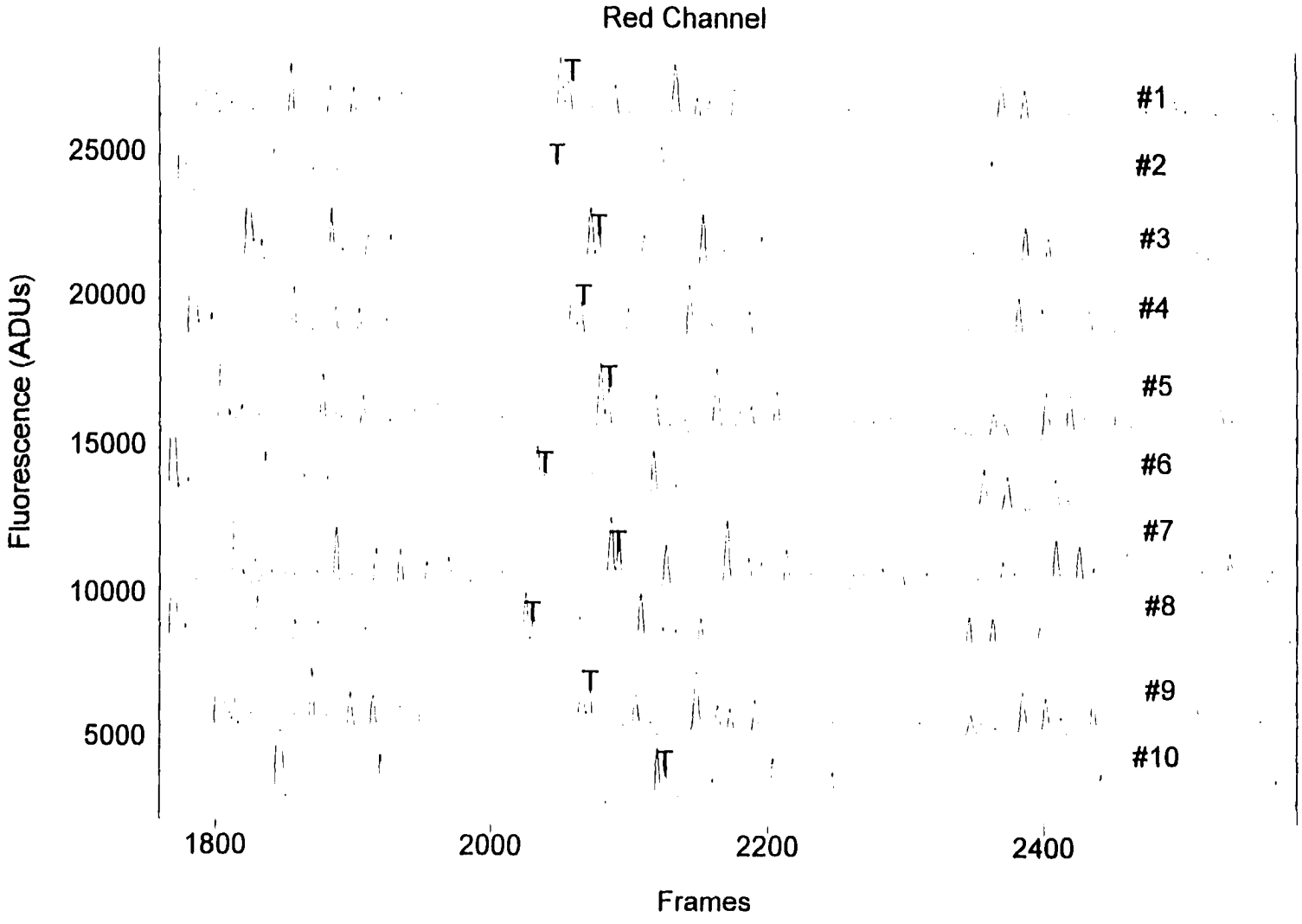


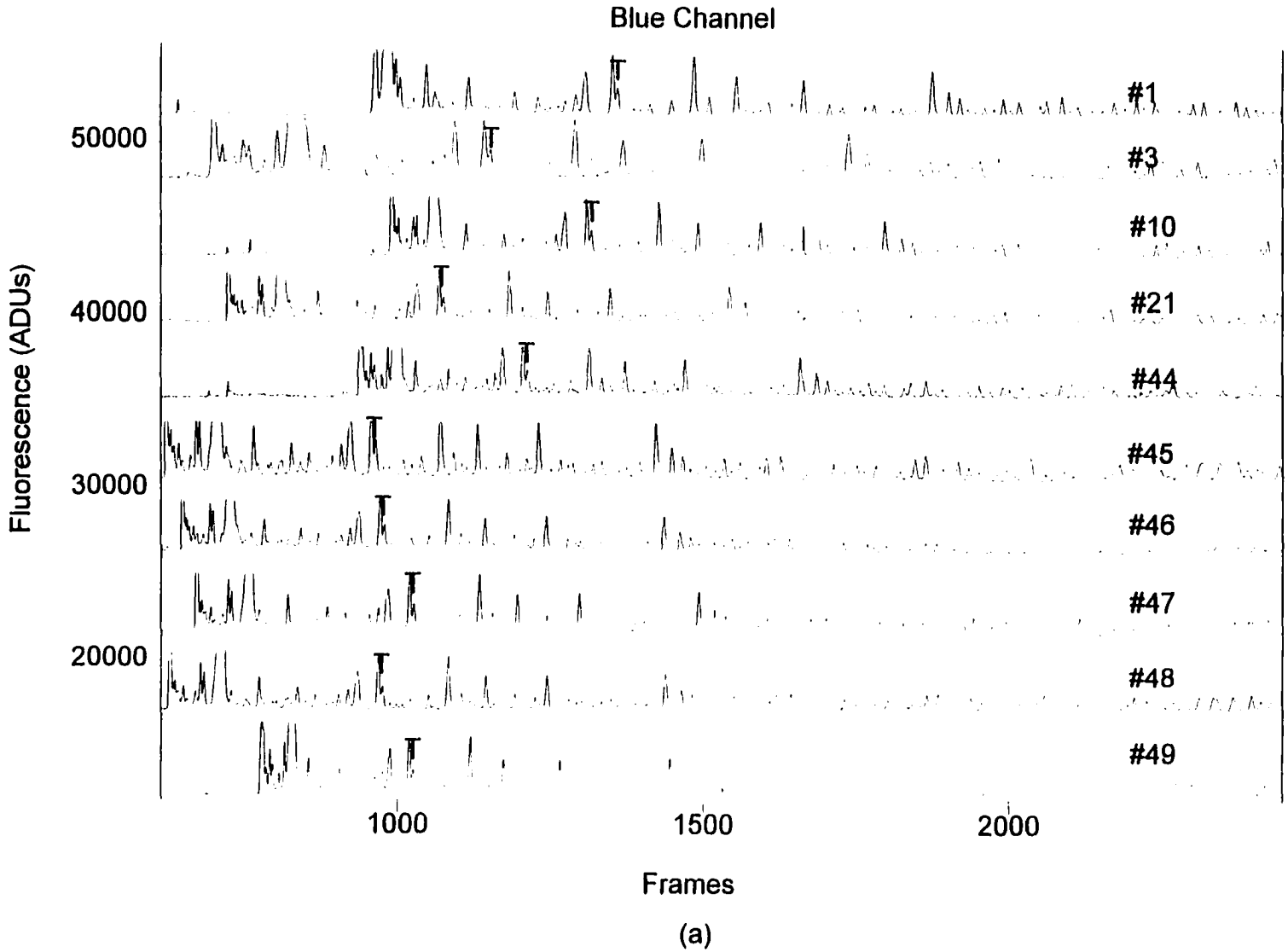
Figure 12 (continued): (b)

is shown in Figure 13. In both runs, peak T of the 45th base is used to evaluate the uniformity of the signal levels from the array capillaries. The results are shown in Table 3 and Table 4 for 10 capillary and 20 capillary runs respectively.

For the 10 capillary run, a 10cm circular lens was used to focus the laser at the center of the array. Two "dummy" capillaries were used to adjust the coupling of the laser beam into the waveguide cell. The space between two "dummy" capillaries is 1.5 cm. Uniform S/N from all 10 capillaries in both channels was obtained, as shown in Table 3. For the 20 capillary run, except for the first capillary, the variation in S/N from all of the capillaries was about 2 times, as shown in Table 4. In order to perform base calling using the software developed in our group, signal to noise ratio requires further improvement. Otherwise, extra peaks are incorrectly called as bases. An array of square capillaries was proposed for this application.

DNA Sequencing in Square Capillaries. The concept of application of rectangular capillaries for electrophoresis was introduced in 1937 [43]. But only recently the application of rectangular capillaries for electrophoresis was addressed [44-48]. These published works cite some advantages of rectangular capillary over cylindrical capillary: (1) better heat dissipation due to the larger surface to volume ratio. (2) a wider capillary increases sample capacity and provides better sensitivity for path-length-dependent detection systems (UV absorption, fluorescence). To our knowledge, none of the works were performed in capillary gel electrophoresis, especially for multiple capillary DNA sequencing.

Figure 13. The background subtracted electropherograms of PGEM/U DNA from 20 randomly picked capillaries of a 100 capillary array. The 488-nm laser output at 70 mW was used for excitation. A 488nm Raman edge filter and a 535nm cutoff filter (Kaiser Optical Systems, Ann Arbor, MI) were attached onto the camera lens (Nikon F1.4/28mm, aperture 2.0). Five sample tubes were combined into one for sample introduction. Samples were introduced by electrokinetic injection for 90s at 150 V/cm. After sample introduction, a 15 minutes resting period was allowed prior to data acquisition. The separation was performed at 230 V/cm. The exposure time gradient began on 10 minutes after the initiation of data collection. The initial exposure time was 200 ms. The final exposure time was 750 ms. Other parameters are identical to those of Figure 12.


Blue Channel
Fluorescence (ADUs)
50000
40000
30000
20000
#1
#3
#10
#21
#44
#45
#46
#47
#48
#49
1000
1500
2000
Frames

(a)

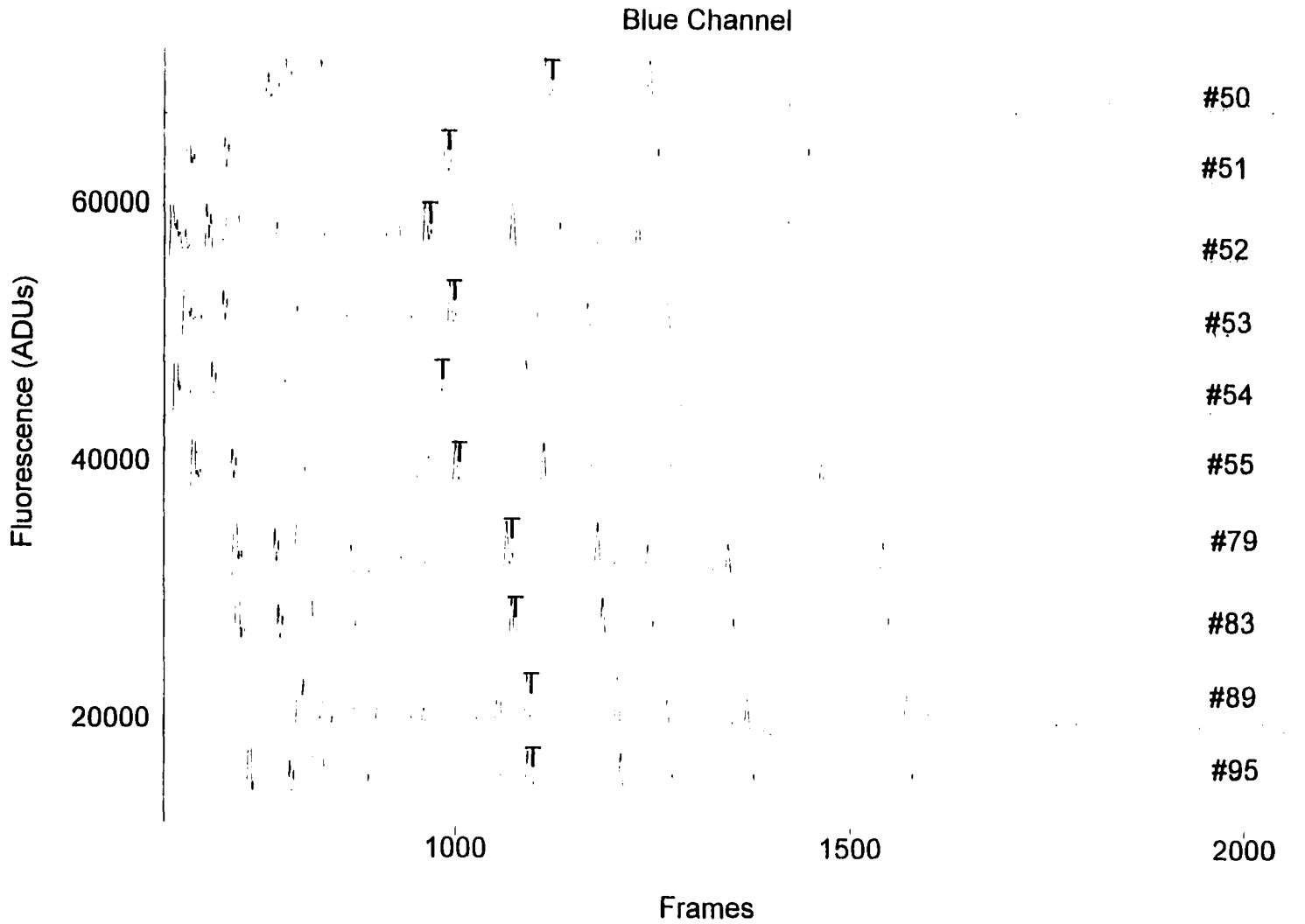


Figure 13 (continued): (b)

Table 3. Comparison of the detection limits of ten capillary array in waveguide system by 45th T base of PGEM/U (peak height of T, Hp; peak to peak noise, Hn; relative detection limit, Hp/Hn)

	Blue Channel			Red Channel		
Capillary	Hp	Hn	Hp/Hn	Hp	Hn	Hp/Hn
#1	5095	33	154	2513	37	68
#2	4051	37	109	1839	33	56
#3	4646	35	133	1892	33	57
#4	3595	32	112	2457	33	74
#5	4492	36	125	2883	36	80
#6	3901	36	108	2333	32	73
#7	5240	31	169	2833	34	83
#8	3413	31	110	2497	33	76
#9	3642	32	114	1726	33	52
#10	4157	35	119	1813	34	53

Table 4. Comparison of the detection limits of 20 among 100 capillary array in waveguide system by 45th T base of PGEM/U (peak height of T, Hp; peak to peak noise, Hn; relative detection limit, Hp/Hn)

Capillary	Hp	Hn	Hp/Hn	Capillary	Hp	Hn	Hp/Hn
#1	7924	37	214	#50	2080	40	52
#3	2245	41	55	#51	3680	34	108
#10	2770	34	81	#52	3487	36	99
#21	2991	36	83	#53	3061	38	81
#44	1538	28	55	#54	2025	32	63
#45	2738	40	68	#55	2598	36	72
#46	3270	35	93	#79	3353	36	93
#47	2924	38	77	#83	3459	36	96
#48	2657	37	72	#89	2289	32	72
#49	3034	41	74	#95	4268	36	118

Improved excitation efficiency is expected using square capillaries for the side-entry excitation scheme (Figure 14). The lens effect does not exist in the square capillary system. Capillaries are immersed into fused silica index matching fluid. This further reduces the stray light due to reflection that occurs at the interface of the two media. At the same time, laser power loss was minimized. The refractive index for gel of 1.5% 8,000,000 and 1.4% 600,000 PEO is 1.366 as shown in Table 1. The refractive index of a fused silica plate is 1.458. By using the side entry excitation scheme, the reduction of the laser power after each capillary is:

$$P(\lambda)=2x(\eta_2-\eta_1)^2/(\eta_2+\eta_1)^2=0.00212 \quad (1)$$

The transmitted incident light after n capillaries is 0.99788^n. The trend of laser power reduction in Figure 14 system when each capillary is filled with PEO gel (1.5% 8,000,000, 1.4% 600,000 PEO) is shown in Table 5. The 500^{th} capillary still receives 34.6% of the laser power received by the first capillary. A thousand capillaries can be implemented in this system by coupling laser power from both sides of the cell. Theoretically, the square capillary is the ideal choice for large array DNA sequencing, especially from the excitation and detection points of view. Several successful runs of DNA sequencing results were obtained using a 10 square capillary array system.

The two-channel sequencing results by "pseudo single capillary injection scheme" is shown in Figure 15. All 10 capillaries have uniform signals in both channels.Instead of gluing 10 capillaries inside the PEEK tubing for bundle injection,

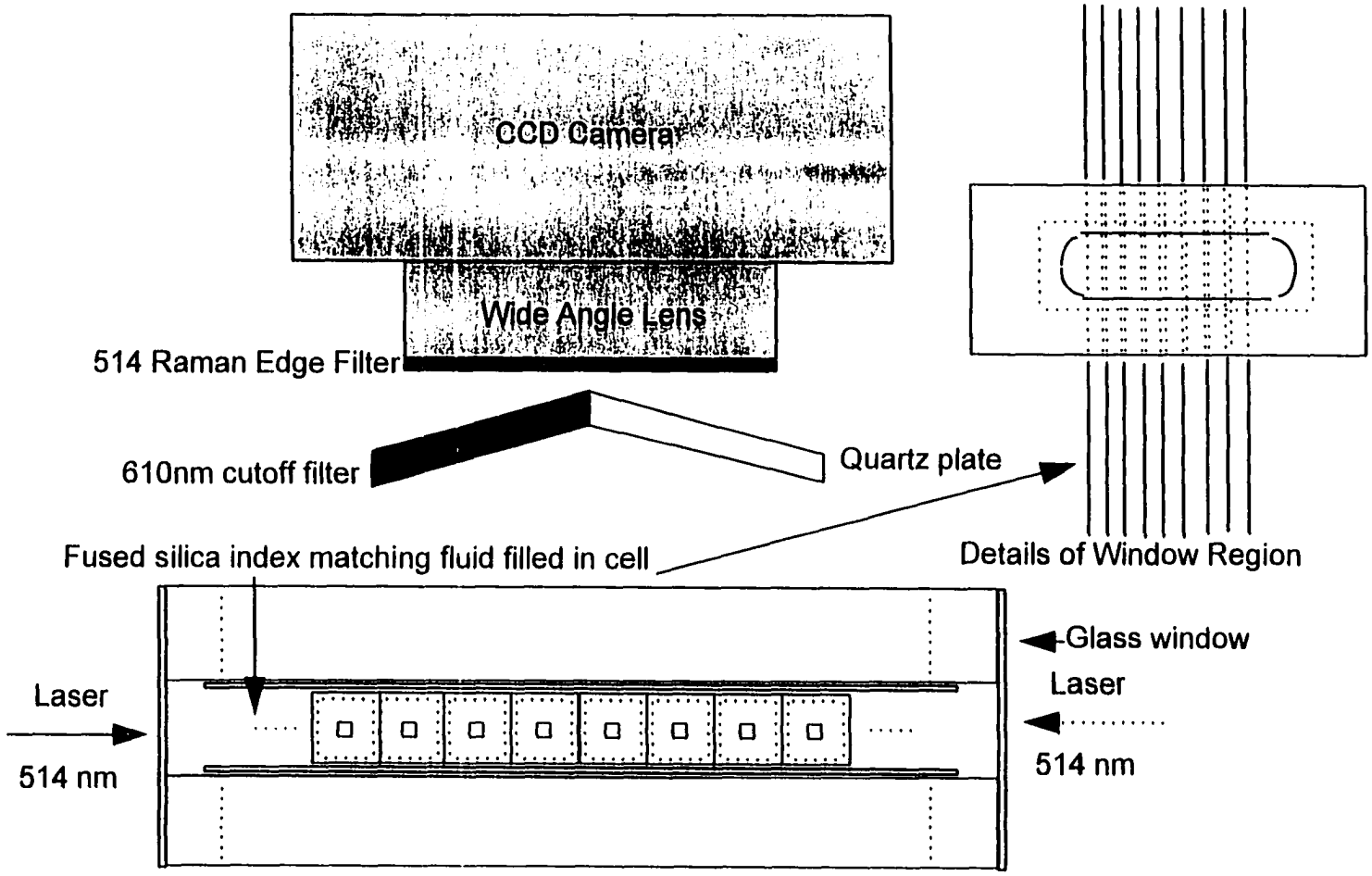


Figure 14. Schematic diagram of the laser excitation and CCD detection system for square capillary array DNA sequencing.

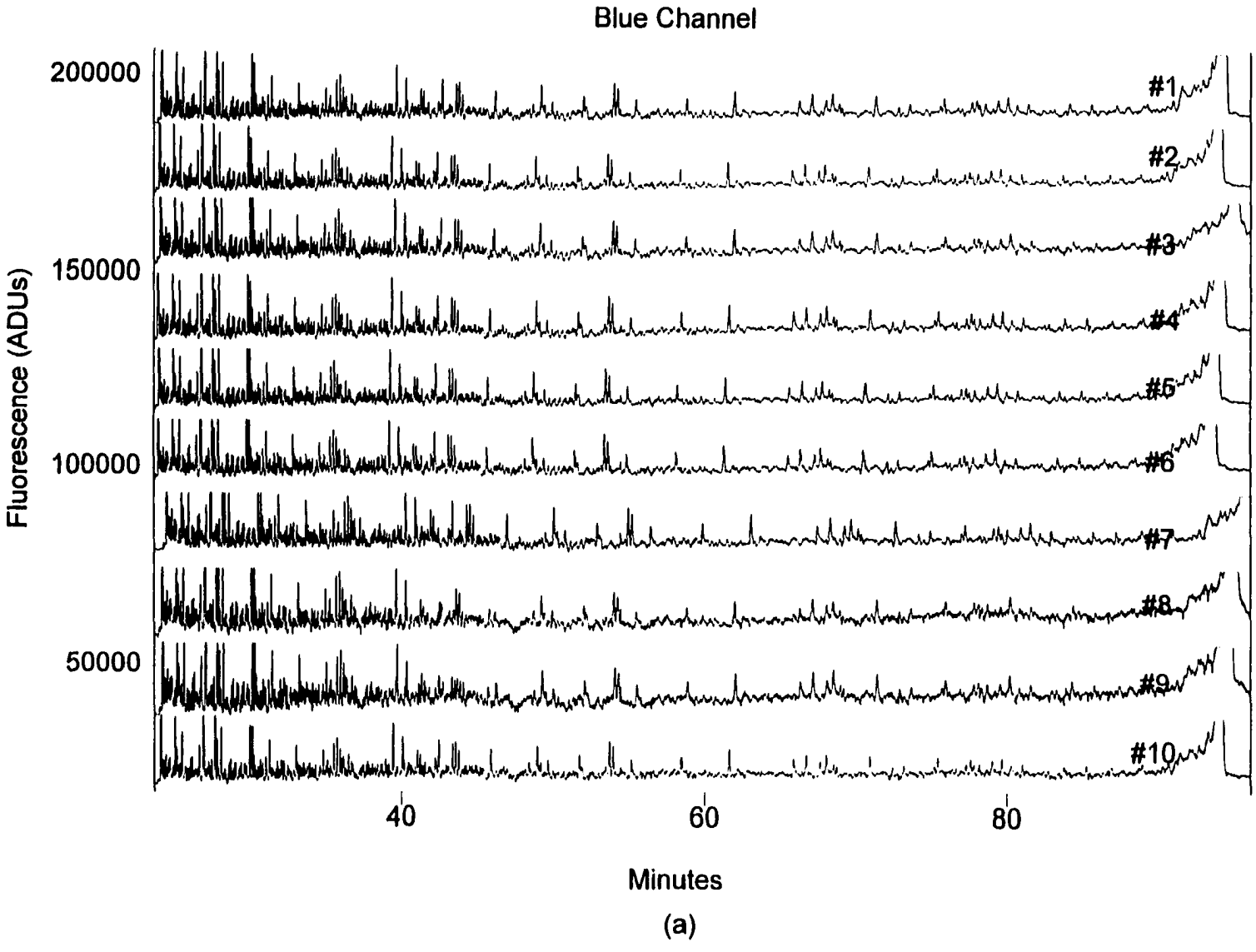
Table 5. Laser power vs. nth capillary in Figure 3-14 system when each capillary is filled with PEO gel (1.5% 8,000,000, 1.4% 600,000 PEO)

Nth capillary	Laser power (P)
#1	1P
#10	0.979P
#50	0.899P
#100	0.809P
#200	0.654P
#300	0.529P
#400	0.428P
#500	0.346P

capillaries are glued onto the outside surface of a 2cm 1/16" i.d., 1/8" o.d. PEEK tubing with equal spacing between capillaries. The injection electrode was inserted inside the PEEK tubing. The 10 capillaries and electrode extended past the tubing. The capillaries are glued to the external surface of the PEEK tubing with spacing between them. In this way, the PEEK tubing provides an environment for capillary array similar to a single capillary in one sample tube. Thus, simultaneous injection of sample into as many as 10 capillaries is afforded without problems such as competition for injection, capillary action or sample depletion problems that can happen when capillaries are bundled together.

The DNA sequencing results of 10 square capillary array using a 5 electrodes

Figure 15. Two-channel sequencing results by "pseudo single capillary injection scheme" into 10 square capillaries. Capillary, 50cm total, 39cm efficient; Injection, 60 s at 75 V/cm; Electrophoresis, 150 V/cm; Laser power, 15 mW; Exposure, 400 ms; Aperture, 2.0; Smaple, 3 tubes of samples were combined to yield 12μl total. (a) electropherograms of 10 capillary array from green channel (b) electropherograms of 10 capillary array from red channel.


Blue Channel
200000
150000
100000
50000
Fluorescence (ADUs)
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
40
60
80
Minutes
(a)

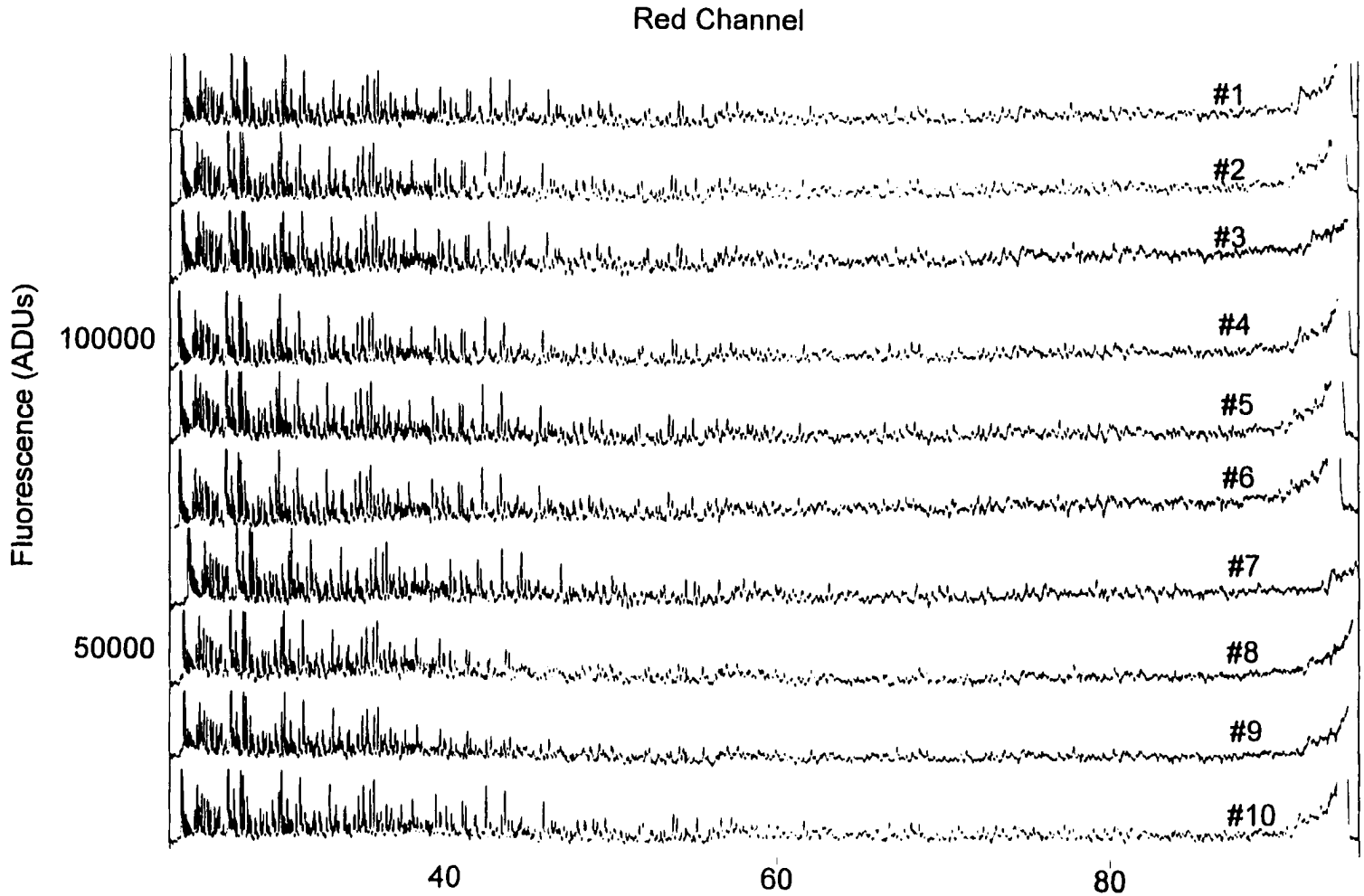
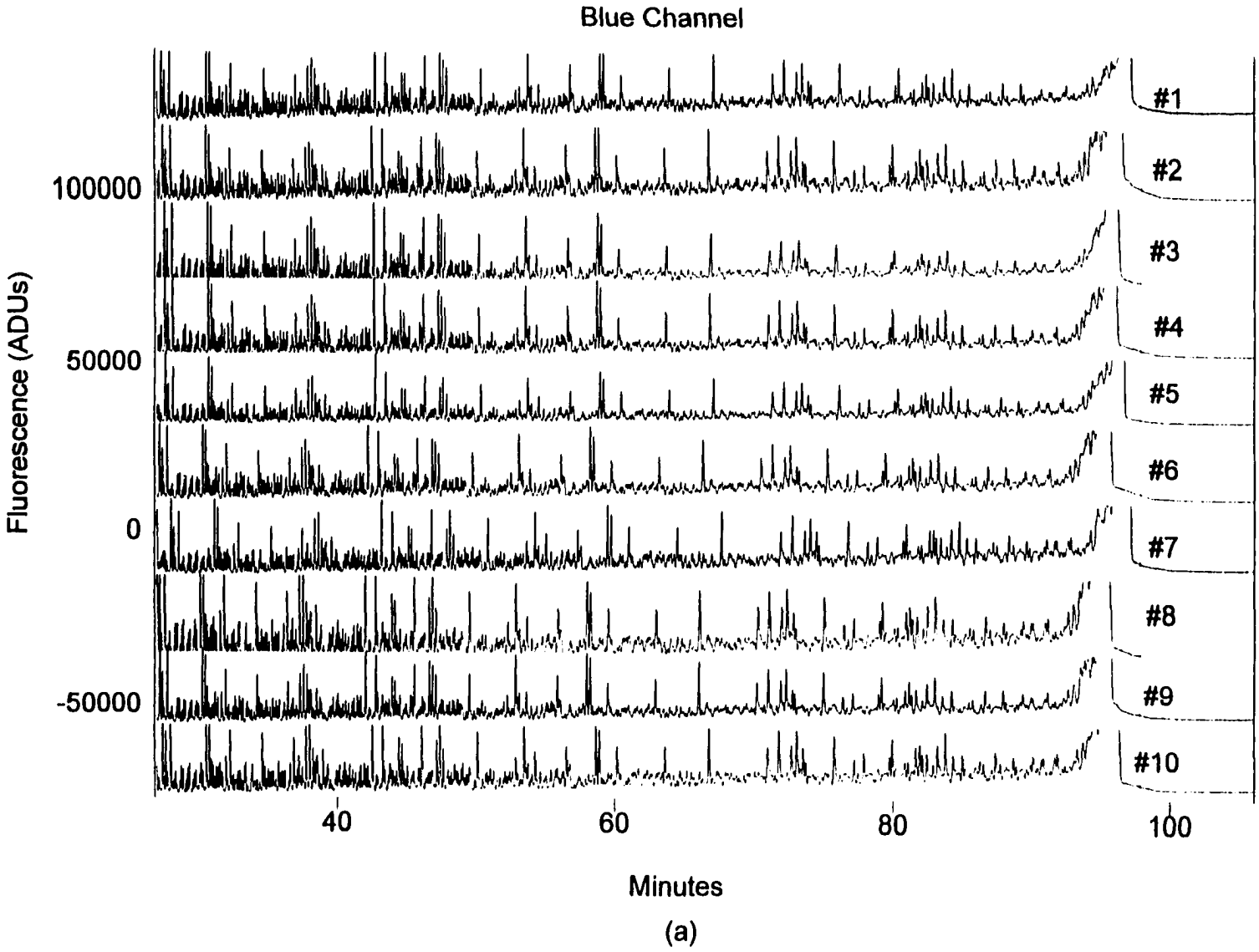


Figure 15 (continued): (b)

individual sample injection scheme (as shown in Chapter 1 Figure 10) is shown in Figure 16. The ten capillaries were divided into two groups. Each group was injected separatedly. Both S/N and resolution were excellent for this run of all 10 capillaries. Base calling was performed on capillary #3 using our group's software. More than 400 bases were called with the accuracy of 98% (Figure 17).

The DNA sequencing experiment was performed on a 100 square capillary array. The 100 square capillaries were divided into 10 groups, each containing 10 capillaries. Each group of 10 capillaries was bundled together at the detection end for gel filling. The injection ends are prepared to allow "pseudo single capillary injection" as in the experiment of Figure 15. The simultaneous injection of 100 capillaries was achieved with an electrode array. This electrode array consists of 10 individual electrodes that are connected to a common high voltage source. An electrode is inserted inside each PEEK tubing. The results for 12 out of 100 capillaries are shown in Figure 18. This 100 square capillary run was not totally successful. Signal to noise ratio is not good enough for base calling using our group's software. There are several possible reasons. Due to the fragility of the square capillary, several capillaries were broken during the gel filling process. This delayed the experiment for about four hours with the gel inside most of the capillaries. The electroosmotic flow is expected to increase and causes less sample introduced into the capillaries. It was found later [49] that signal level for the large capillary array simultaneou injection was lower than that of the small capillary array

Figure 16. Two-channel sequencing results by "five-electrode individual sample injection scheme (as shown in Chapter 1 Figure 10)" into 10 square capillaries. Capillary, 50cm total, 38cm efficient; Injection, 60 s at 80 V/cm; Electrophoresis, 150 V/cm; Laser power, 40mW; Exposure, 400 ms; Aperture, 2.0. (a) electropherograms of 10 capillary array from green channel (b) electropherograms of 10 capillary array from red channel.


Blue Channel
Fluorescence (ADUs)
100000
50000
0
-50000
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
40
60
80
100
Minutes

(a)

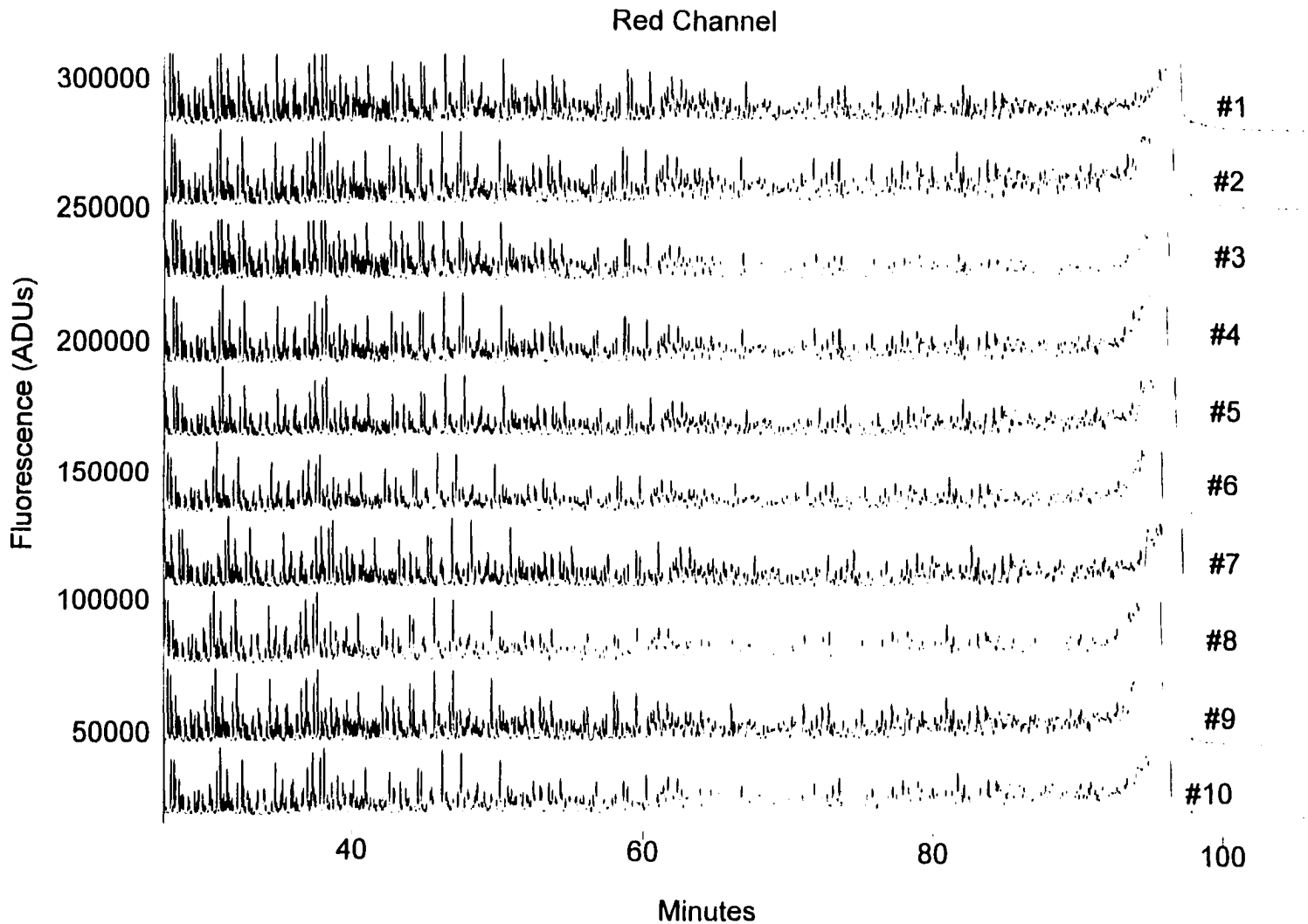


Figure 16 (continued): (b)

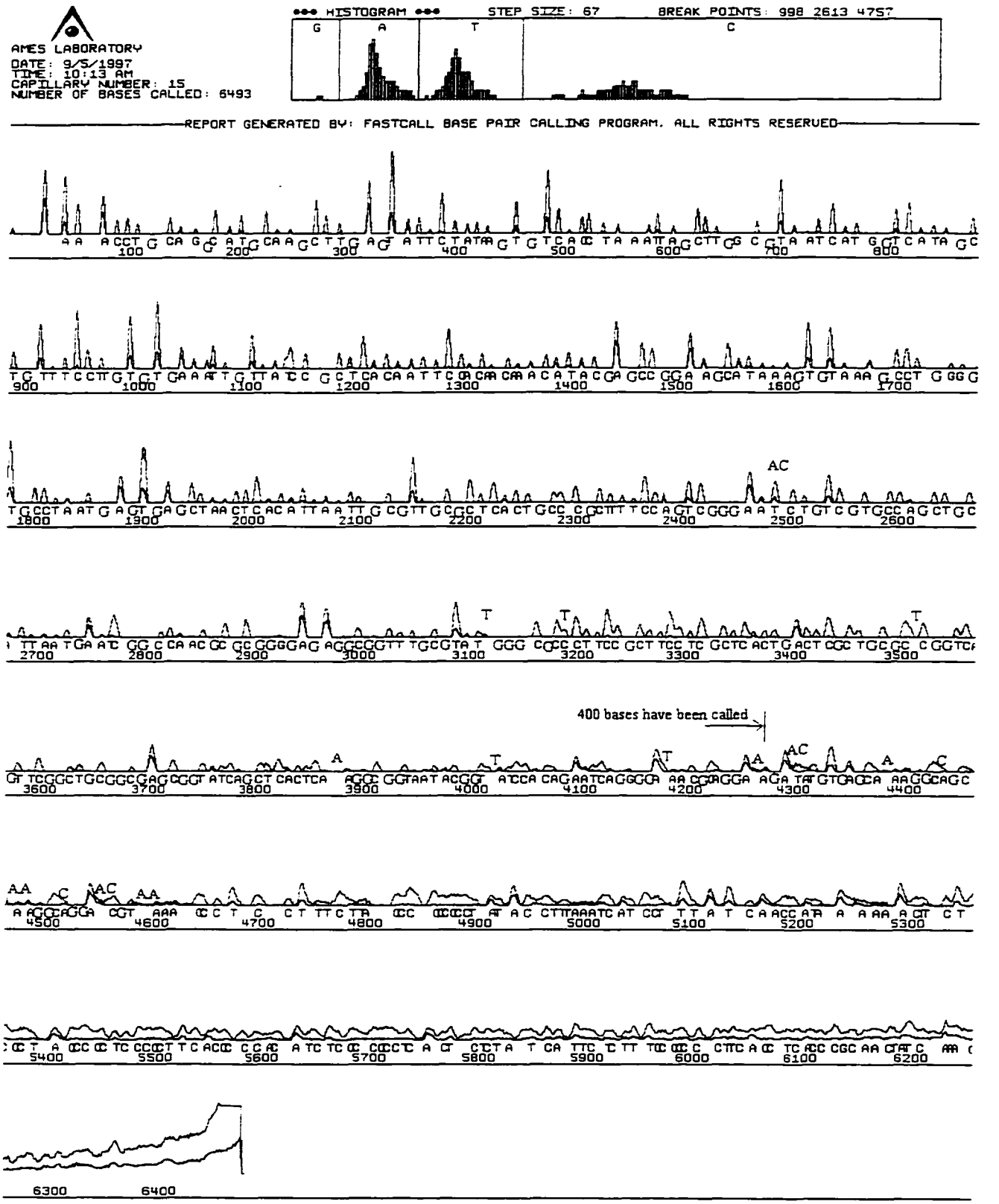
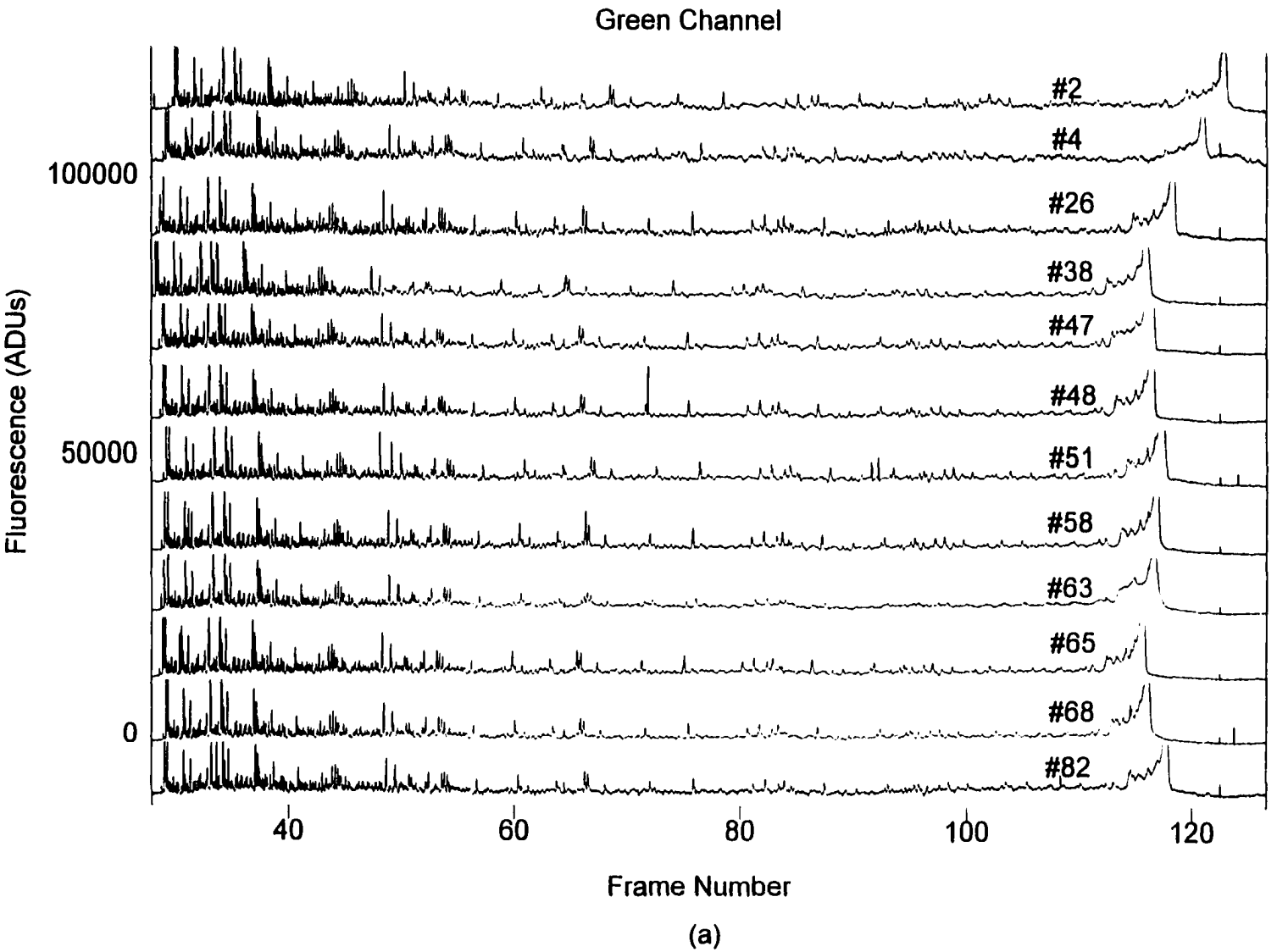


Figure 17. Base calling on capillary #3 of the 10 capillary run in Figure 16.

Figure 18. Detection of PGEM/U DNA from 100 square capillary array. Square capillary, 65cm total, 50cm efficient, 340μm outer edge, 75μm inner edge; Injection, 60 s at 75 V/cm; Electrophoresis, 150 V/cm; Laser power, 40mW; Exposure, 300 ms; Aperture, 2.0. (a) twelve electropherograms out of 100 capillary array from green channel (b) twelve electropherograms out of 100 capillary array from red channel.


Green Channel
Fluorescence (ADUs)
100000
50000
0
#2
#4
#26
#38
#47
#48
#51
#58
#63
#65
#68
#82
40
60
80
100
120
Frame Number

(a)

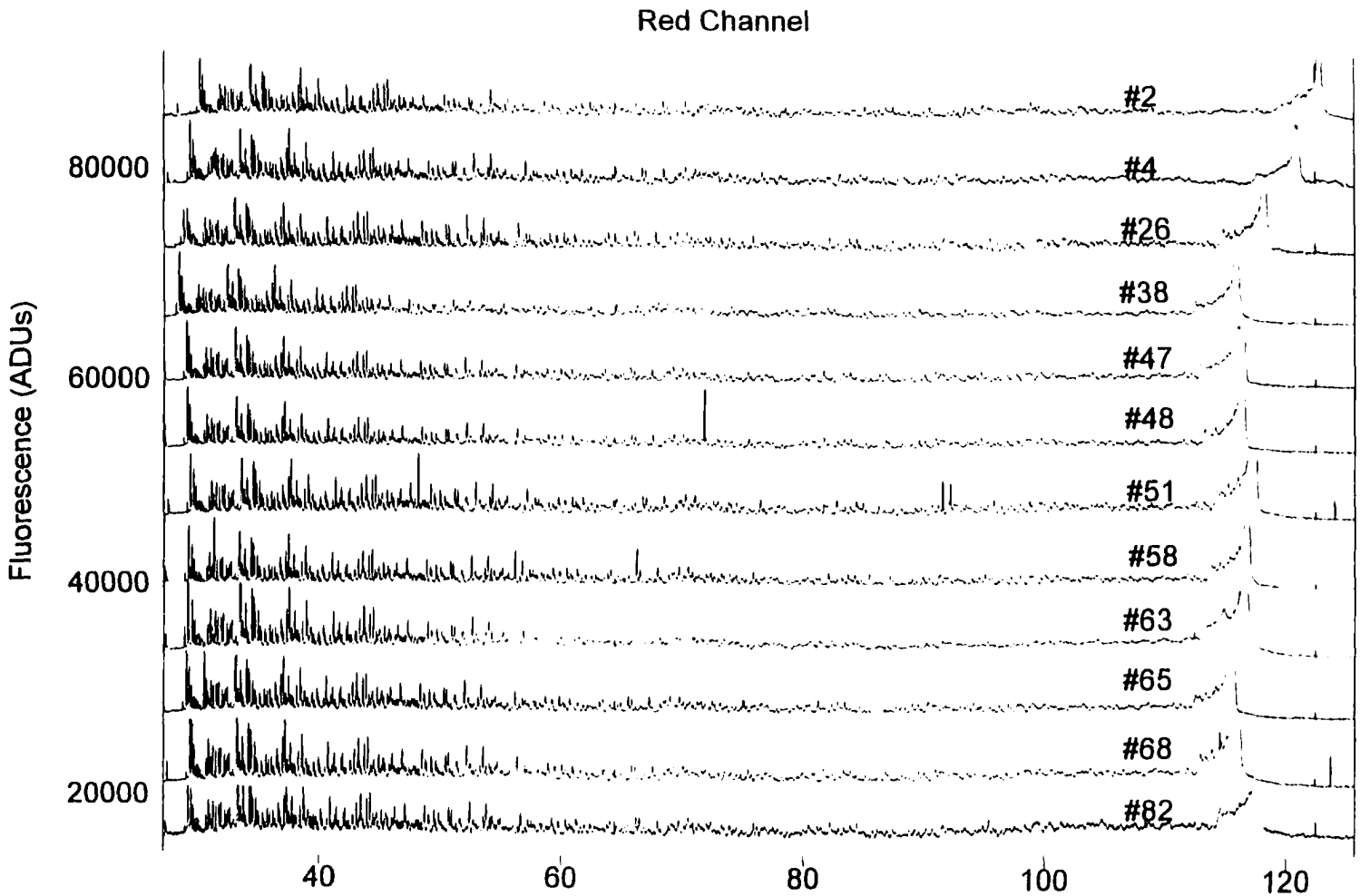


Figure 18 (continued): (b)

under the same sample introduction conditions. This interesting phenomena still can not be rationalized. By increasing the injection time, the signal level will come back.

CONCLUSION AND FUTURE DIRECTIONS

The development of the excitation scheme for large capillary array DNA sequencing has performed on two types of capillaries, circular and square. The excitation and detection efficiency is much higher using the square capillary. Capillary array DNA sequencing results have been obtained from both types of capillaries. There is no observable retention time difference for the DNA fragments between two types of capillaries. The base calling has performed on the square capillary array DNA sequencing data using our group's software. More than 400 bases can be called with the accuracy of 98%.

In order to get around the fragility of the square capillary, the following coupling scheme was proposed (Figure 19). The 150μm o.d. 75μm i.d. circular capillaries can be inserted into channels on a chip or into square capillaries with the edges of 150μm. This scheme offers the advantages of robustness and low cost associated with the circular capillary while retaining the high excitation efficiency of the square capillary.

Preliminary tests on this arrangement employed a coupling of a 150μm o.d. 75μm i.d. circular capillary with a 360μm o.d. 152μm i.d. circular capillary. As is shown in Figure 20, the separation efficiency for the coupling scheme is lower

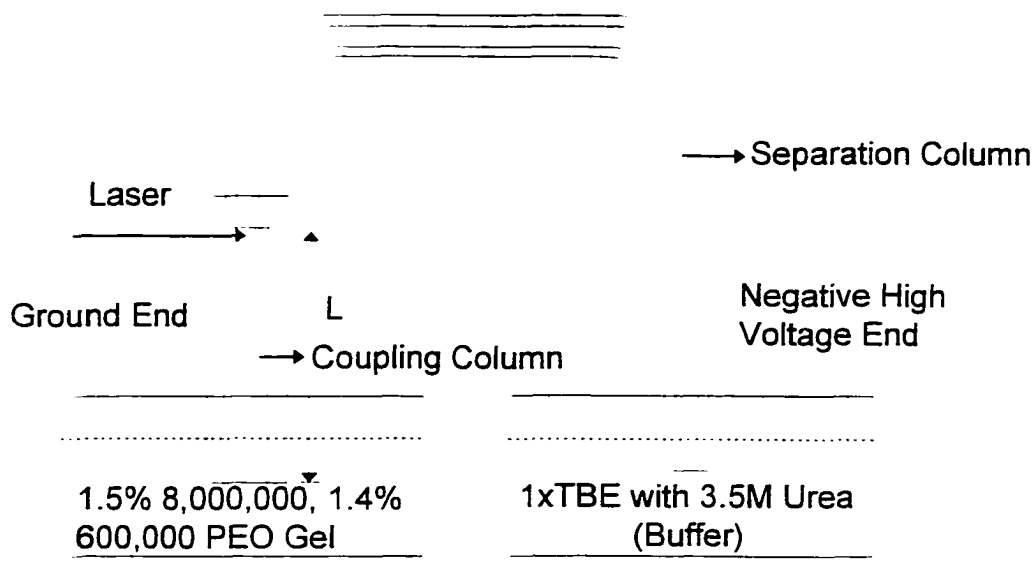
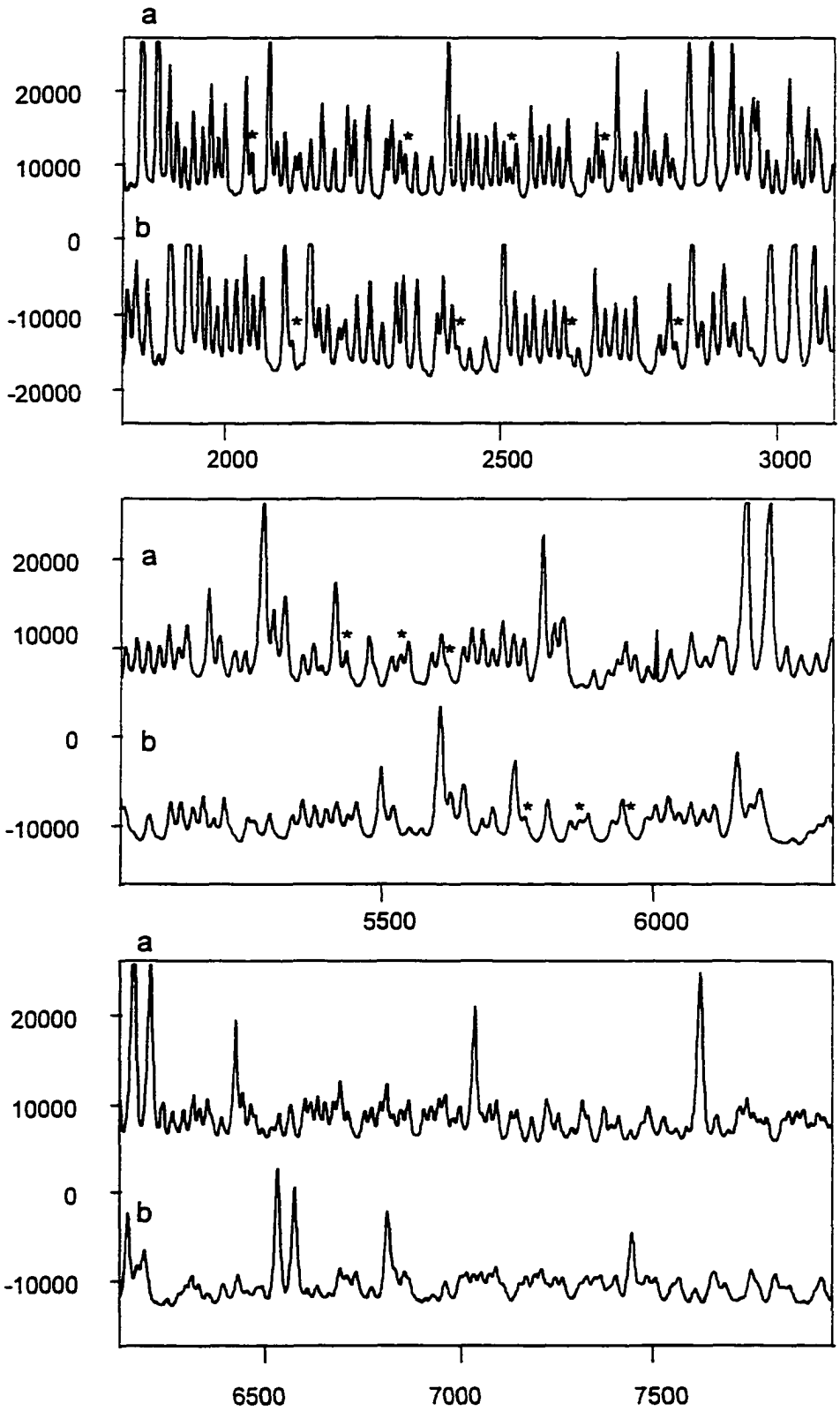


Figure 19: Schematic diagram of coupling scheme. Separation column, 55cm total, 150μm o.d., 75μm i.d.; coupling column, 360μm o.d., 152μm i.d; L, efficient length of the coupling column.

Figure 20. Comparison of the separation efficiency between the run from the coupling scheme and the standard run. Capillary, 55cm efficient; L, 4cm; Injection, 60s at 75 V/cm; Electrophoresis, 150 V/cm. (a) standard run. (b) run from the coupling scheme.


a
20000
10000
0
b
-10000
-20000
2000
2500
3000
a
20000
10000
0
b
-10000
5500
6000
a
20000
10000
0
b
-10000
6500
7000
7500

than that from the standard capillary run. There are several possible reasons for this.

In the scheme, there is an electrical field strength difference in the separation column and the coupling column. According to Ohm's Law:

$$R=kL/A \text{ (k; constant; L, length; A, area)} \quad (2)$$

The voltage drop (ΔV) on a resistor is proportional to resistance.

$$E \text{ (electric field intensity)} = \Delta V/L \quad (3)$$

The ratio of two field intensity is:

$$E_1 / E_2 = A_2 / A_1 \quad (4)$$

The area of the coupling column is 4 times of the separation column. The electrical field strength in the coupling column is 1/4 of the separation column. Hence, the DNA fragments slow down by a factor of 4 when they move into the coupling capillary. This leads to a four-fold compression in all widths of the DNA bands. This "stacking" process is good for the detection but could affect the separation efficiency because two bands can elute out together. This is particularly problematic for those bands without enough selectivity. The other possible reason may be the gel filling is not even through the separation column and the coupling column. The scheme needs to be further developed in order for large capillary array DNA sequencing to be implemented.

ACKNOWLEDGMENTS

The authors thank Mr. Xiaoyi Gong for the generous help in software development for data acquisition and extraction. The Ames Laboratory is operated for the U. S. Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Director of Energy Research, Office of Health and Environmental Research.

REFERENCE

1. J. A. Taylor and E. S. Yeung *Anal. Chem.* **1993**, *65*, 956-960.
2. K. Ueno and E. S. Yeung *Anal. Chem.* **1994**, *66*, 1424-1431.
3. The Human Genome Project: deciphering the blueprint of heredity, Mill Valley, CA **1994**.
4. R. A. Mathies and X. C. Huang *Nature* **1992**, *359*, 167-169.
5. X. C. Huang, M. A. Quesada, R. A. Mathies *Anal. Chem.* **1992**, *64*, 967-972.
6. X. C. Huang, M. A. Quesada, R. A. Mathies *Anal. Chem.* **1992**, *64*, 2149- 2154.
7. H. Kambara and S. Takahashi *Nature* **1993**, *361*, 565-566.
8. S. Takahashi, K. Murakami, T. Anazawa, H. Kambara *Anal. Chem.* **1994**, *66*, 1021-1026.
9. X. Lu and E. S. Yeung *Appl. Spectrosc.* **1995**, *49*, 605-609.
10. Takashi Anazawa, Satoshi Takahashi, and Hideki Kambara *Anal. Chem.* **1996**, *68*, 2699-2704.

11. J.H.Jett, R.A.Keller and etc., *Journal of Biomolecular Structure and Dynamics* **1989,***17,* 301-309.

12. J.D.Harding and R.A.Keller, *Trends in Biotechnology* **1992,***10,* 55 .

13. M. Nishioka, T. Tanizoe, A. Fujita, S. Katsura, A. Mizuno *Institute of Physics Conference Series* **1995,** *143,* 335-338.

14. R.Drmanac and R.Crkvenjakov, *Int.J. Genome Res.* **1992,***1,* 59.

15. R.Drmanac, S.Drmanac, Z.Strezoska, L.Hood, R.Crkvenjakov and etc., *Science* **1993,** *260,* 1649-52.

16. Mark D. Adams, Chris Fields and J. Craig Venter *Automated DNA Sequencing and Analysis,* Academic Press, San Diego, **1994**.

17. Deborah Nobel *Anal. Chem.* **1995,** *67,* 201A-204A .

18. Franz Hillenkamp, Michael Karas, Ronald C. Beavis and Brian T. Chait *Anal. Chem.* **1991,** *63,* 1193A.

19. L. Grotjahn, H. Blocker, R. Frank, *Biomed. Mass Spectrom.* **1985,***12,* 514.

20. C.R.Blakely, J.J.Carmody, M.L. Vestal, *Anal. Chem.* **1980,** *52,* 1636.

21. C.M.Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, *Anal. Chem.* **1985,** *57,* 675 .

22. R.J. Cotter, *Anal. Chim. Acta* **1987,***195,* 45-59.

23. M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp *Int. Mass Spectrom. Ion Proc.* **1987,** *78,* 53-68.

24. Daniel P. Little, Russell A. Chorush, J. Paul speir, Michael W. Senko, Neil L.

Kelleher, and Fred W. McLafferty *J. Am. Chem. Soc.*, **1994,***116,* 4893-4897.

25. Daniel P. Little and Fred W. McLafferty *J. Am. Chem. Soc.*, **1995,** *117,* 6783-6784.
26. K. Tang, S. L. Allman and C. H. Chen *Rapid Communications in Mass Spectrometry,* **1992,** *6,* 365-368.
27. Gary R. Parr, Michael C. Fitzgerald and Lloyd M. Smith *ibid,* **1992,** *6,* 369-372.
28. Kuang J. Wu, Anna Steding and Christopher H. Becker *ibid,* **1992,** *7,* 142-146.
29. Peter Juhasz, Mark T. Roskey, Igor P. Smirnov, Lawrence A. Haff, Marvin L. Vestal, and Stephen A. Martin *Anal. Chem.*, **1996,** *68,* 941-946.
30. Kuang J. Wu, T. Shaler and Christopher H. Becker *Anal. Chem.* **1994,** *66,* 1637-1645.
31. E. Nordhoff, R. Cramer, M. Karas, F. Hillenkamp, F. kirpekar, K. Kristiansen, P. Roepstoff *Nucleic Acids Res.* **1993,** *21,* 3347-3357.
32. A. E. Barron, W. M. Sunada, H. W. Blanch **1996,** *Electrophoresis, 17,* 744-757.
33. D. Figeys, N. J. Dovichi *Journal of Chromatography A* **1995,** *717,* 113-116.
34. J. Z. Zhang, Y. Fang, J. Y. Hou, H. J. Ren, R. Jiang, P. Roos, N. J. Dovichi *Anal. Chem.* **1995,** *67,* 4589-4593.
35. E. N. Fung and E. S. Yeung *Anal. Chem.* **1995,** *67,*1913-1919.
36. (a) Y. F. Pariat, J. Berka, D. N. Heiger, T. Schmitt, M. Vilenchik, A. S. Cohen, F. Foret, B. L. Karger *Journal of Chromatography A* **1993,** *652,* 57-66.

(b) Emanuel Carrilho, Marie C. Ruiz-Martinez, Jan Berka, Igor Smirnov,

Wolfgang Goetzinger, Arthur W. Miller, David Brady, and Barry L. Karger *Anal. Chem.* **1996**, *68*, 3305-3313.

37. H. Swerdlow, J. Z. Zhang, D. Chen, H. R. Harke, R. Grey, S. Wu, N. J. Dovichi, C. Fuller, *Anal. Chem.*, **1991**, *63*, 2835-2841.

38. (a) A.T. Woolley, R. A. Mathies, *Proceedings of The National Academy of Sciences of The United States of America*, **1994**, *91, iss 24*, 11348-11352. (b) Carlo S. Effenhauser, Aran Paulus, Andreas Manz, and H. Michael Widmer, *Anal. Chem.*, **1994**, *66*, 2949-2953.

39. Edward S. Yeung and Q. Li, *High Performance Capillary Electrophoresis*, Morteza G. Kbaledi, John Wiley & Sons, Inc, in press.

40. *CRC Handbook of Chemistry and Physics*, 68th ed.; R. C. West Ed.; CRC Press, Inc.; Boca Raton, FL, 1987-1988; p E-372.

41. *Optical Fiber Communications Principles and Practice, Second edition* John M. Senior, Prentice Hall , New York 1992.

42. The base-calling software developed by David Hall and Xiaohua Cheng.

43. A. Tiselius, *Trans. Faraday Soc.*, **1937**, *33*, 524.

44. M. Janson, A. Emmer and J. Roerade, *J. High Resolut. Chromatogr.*, **1989**, *12*, 797.

45. A. Cifuentes and H. Poppe, *Chromatographia*, **1994**, 39, 391.

46. T. Tsnda, J. V. Sweedler and N. Zare, *Anal. Chem.*, **1990**, *62*, 2149.

47. A. Cifuentes and H. Poppe, *Electrophoresis*, **1995**, *16*, 2051.

48. J. M. Mesaros, G. Luo, J. Roerade and A. G. Ewing, *Anal. Chem.*, **1993**, *65*,

3313.

49. Ho-ming Pang, unpublished data, 1997.

GENERAL SUMMARY

Capillary electrophoresis follows the separation mechanism of traditional electrophoresis with a capillary format. CE has many potential advantages to offer biochemists who use conventional slab-gel electrophoresis as well as analytical chemists who use various other chromatographic techniques. CE combines the ease and speed of HPLC with extremely high efficiency and the ability to handle very small samples. It is used for a wide variety of separations from peptides, proteins and DNA to inorganic ions and industrial polymers. The successful application of CE in DNA separation and sequencing will break the technology bottleneck of the Human Genome Project and will be widely used in gene finding, and disease diagnosis.

This dissertation is fully devoted to the development of multiple capillary electrophoresis systems for high throughput DNA sequencing. A simple optical setup was established. Two capillary array formats (round and square) were used for capillary array DNA sequencing. With improvements in the production quality of the square capillary manufacturing, the excitation and detection efficiency is further improved. This project reveals several challenges in instrumentation, CE separation, sample handling and injection schemes. Successful separations require intimate control of all experimental parameters. In order for this technology to be used for large-scale human DNA sequencing, automation of each step is important. The

cooperative efforts of scientists working in different disciplines (chemistry, physics, biochemistry and computer science) will smooth the journey toward that goal.

CITED REFERENCES

1. David Freifelder, George M. Malacinski. *Essentials of Molecular Biology; Jones and Bartlett Publishers, Boston London, 1993.*
2. U.S. Department of Energy and U.S. Department of Health and Human Services Understanding Our Genetic Inheritance, *The U.S. Human Genome Project: The First Five Years Report DOE/ER-0452P, April, 1990.*
3. Strege, M.; Lagu, A. *Anal. Chem. 63, 1223 (1991).*
4. McGregor, D.; Yeung, E. D. *J. Chromatogr. A 652, 31 (1993).*
5. Chiari, M.; Nesi, M.; Righetti, R. G. *J. Chromatogr. A 652, 31 (1993).*
6. J. Korenberg, etc. *Trends Biotechnol. 10 27-32 (1992).*
7. Olson, M., Hood, C., Cantor, C. and Botstein, D. *Science 245, 1434-1435 (1989).*
8. Gibbs, R. A. *Anal. Chem. 62, 1202-1214 (1990).*
9. Erlich, H.A.; Gelfand, D.; Sninsky, J. J. *Science 252, 1643-1651 (1991).*
10. Hunkapiller, T.; Kaiser, R. J.; Koop, B. F.; Hood, L. *Science 254, 59-67 (1991).*
11. *Program and Abstracts of AMS'95 'Third International Conference on Automation in Mapping and DNA Sequencing'* Lawrence Berkeley National Laboratory, Berkeley, California Nov. 5-8, 1995.
12. Abstracts of *DOE: Human Genome Program* Contractor-Grantee Workshop V, Santa Fe, New Mexico January 28-February 1, 1996.
13. C. T. Caskey *Science 236, 1223-1228 (1987).*

14. V. Hochgeschwender, *Trends Genet. 8, 41-44 (1992).*

15. a) M. Vilenchik, A. Belenky, A.S. Cohen *J. of Chromatography A 663, 105-113 (1994).*

b) A.S. Cohen, M. Vilenchik, J. L. Dudley, M. W. Gemborys, A. J. Bourque *J. of Chromatography 638, 293-301 (1993).*

16. Sanger, F., Nicklem, S. and Coulson, A. R. *Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977).*

17. Maxam, A. and Gilbert, W. *Proc. Natl Acad. Sci. USA 74, 560-564 (1977).*

18. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Hriner, C., Kent, S. B. and Hood, L. E. *Nature 321, 674-679 (1986).*

19. Lloyd M. Smith *Genome 31, 929-937 (1989).*

20. George L. Trainor *Anal. Chem. 62, 418-426 (1990).*

21. T. Hunkapiller, R. J. Kaiser, B. F. Koop, L. Hood *Science 254, 59-67 (1991).*

22. James M. Prober, George L. Trainor, Rudy J. Dam, Frank W. Hobbs, Charles W. Robertson, Robert J. Zagursky, anthony J. Cocuzza, Mark A. Jensen, Kirk Baumeister *Science 238, 336-341(1987).*

23. Operation mannual for ABI 373.

24. S. Carson, A.S. Cohen, A. Belenkii, M.C. Ruiz-martinez, J. Berka, B.L. Karger *Anal. Chem. 65, 3219-3226 (1993).*

25. S. Takahashi, K. Murakami, T. Anazawa and H. Kambara *Anal. Chem. 66, 1021-1026 (1994).*

26. Jingyue Ju, Indu Kheterpal, James R. Scherer, Chihchuan Ruan, Carl W. Fuller, Alexander N. Glazer, and Richard A. Mathies *Analytical Biochemistry 231, 131-140 (1995).*

27. Jingyue Ju, Chihchuan Ruan, Carl W. Fuller, Alexander N. Glazer, and Richard A. Mathies *Proc. Natl. Acad. Sci. USA 92, 4347-4351 (1995).*

28. Michael L. Metzker, Jing Lu, Richard A. Gibbs *Science 271, 1420-1422 (1996).*

29. Hideki Kambara, Tetsuo Nishikawa, Yoshiko Katayama and Tomoaki Yamaguchi *Bio/technology 6, 816-821 (1988).*

30. Milhelm Ansorge, Brian S. Sproat, Josef Stegemann and Christian Schwager *J. Biochem. and Biophy. Methods 13, 315-323 (1986).*

31. Mark D. Adams, Anthony R. Kerlavage, Jenny M. Kelley, Jeannine D. Gocayne, Chris Fields, Claire M. Fraser and J. Craig Venter *nature 368, 474-475 (1994).*

32. J.H.Jett, R.A.Keller and etc., *Journal of Biomolecular Structure and Dynamics 7, 301-309 (1989).*

33. J.D.Harding and R.A.Keller, *Trends in Biotechnology 10, 55 (1992).*

34. R.Drmanac and R.Crkvenjakov, *Int.J. Genome Res. 1, 59 (1992).*

35. R.Drmanac, S.Drmanac, Z.Strezoska, L.Hood, R.Crkvenjakov and etc., *Science 260, 1649-52 (1993).*

36. Mark D. Adams, Chris Fields and J. Craig Venter *Automated DNA Sequencing and Analysis,* Academic Press, San Diego, 1994.

37. Deborah Nobel *Anal. Chem. 67, 201A-204A (1995).*

38. Franz Hillenkamp, Michael Karas, Ronald C. Beavis and Brian T. Chait *Anal. Chem. 63, 1193A (1991).*

39. L. Grotjahn, H. Blocker, R. Frank, *Biomed. Mass Spectrom. 12, 514-24 (1985).*

40. C.R.Blakely, J.J.Carmody, M.L. Vestal, *Anal. Chem. 52, 1636, (1980).*

41. C.M.Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, *Anal. Chem. 57, 675 (1985).*

42. R.J. Cotter, *Anal. Chim. Acta 195, 45-59 (1987).*

43. M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp *Int. Mass Spectrom. Ion Proc. 78, 53-68 (1987).*

44. Daniel P. Little, Russell A. Chorush, J. Paul speir, Michael W. Senko, Neil L. Kelleher, and Fred W. McLafferty *J. Am. Chem. Soc., 116, 4893-4897 (1994).*

45. Daniel P. Little and Fred W. McLafferty *J. Am. Chem. Soc., 117, 6783-6784 (1995).*

46. K. Tang, S. L. Allman and C. H. Chen *Rapid Communications in Mass Spectrometry, 6, 365-368 (1992).*

47. Gary R. Parr, Michael C. Fitzgerald and Lloyd M. Smith *ibid, 6, 369-372 (1992).*

48. Kuang J. Wu, Anna Steding and Christopher H. Becker *ibid, 7, 142-146 (1992).*

49. Peter Juhasz, Mark T. Roskey, Igor P. Smirnov, Lawrence A. Haff, Marvin L. Vestal, and Stephen A. Martin *Anal. Chem., 68, 941-946 (1996).*

50. Kuang J. Wu, T. Shaler and Christopher H. Becker *Anal. Chem. 66, 1637-1645 (1994).*

51. E. Nordhoff, R. Cramer, M. Karas, F. Hillenkamp, F. kirpekar, K. Kristiansen, P. Roepstoff *Nucleic Acids Res. 21, 3347-3357 (1993).*

52. A. J. Kostichka; M. L. Marchbanks; R. L. Brumley; H. Drossman; L. M. Smith *Bio/Technology 10, 78-81 (1992).*

53. Andrew G. Ewing, Paula Beyer Hietpas, and Katherine M. Bullard *HPCE 96 lecture.*

54. Edward S. Yeung and Q. Li, *High Performance Capillary Electrophoresis,* Morteza G. Kbaledi, John Wiley & Sons, Inc., in press.

55. J. A. H. Drossman Luckey, A. J. Kostichka, J. D'Cunhua, L. M. Smith *Anal. Chem. 62, 900-903 (1990).*

56. X. Lu and E.S. Yeung *Appl. Spectrosc. 49, 605-609 (1995).*

57. E. N. Fung, E. S. Yeung *Anal. Chem. 67,1913-1919 (1995).*

58. Zhang J. Z., Fang Y., Hou J.Y., Ren H. J., Jiang R. Roos P., Dovichi N. *Anal. Chem. 67, 4589-4593 (1995).*

59. A. E. Barron, W. M. Sunada, H. W. Blanch *Electrophoresis 17, 744-757(1996).*

60. Emanuel Carrilho, Marie C. Ruiz-Martinez, Jan Berka, Igor Smirnov, Wolfgang Goetzinger, Arthur W. Miller, David Brady, and Barry L. Karger *Anal. Chem. 68, 3305-3313 (1996).*

ACKNOWLEDGMENTS

It has been said, and I agree, that "the hardest job in the world is to be an international graduate student". This is not just because of language problems, heavy course-loads, and the pressure of research progress, but also because of the necessary adjustment and re-evaluation of ones thinking philosophy to the western style. The achievement of this degree for me was a hard but precious journey. It was full of joy, anger, excitement, frustration, happiness and sorrow. There are so many people to whom I am indebted for helping me achieve this degree. I would like to extend my thanks and acknowledgments to them all.

First, I would like to thank my major professor, Edward S. Yeung, for his guidance throughout each step of this work. His expert guidance, excellence in research, creative ideas and patience during my graduate study will benefit my whole career. I appreciate the opportunity he provided me with work in his group and learn cutting-edge research.

I would like to thank Professor Xinyao Zhou, my Master's Thesis advisor, Professor Zibin Qin, my Bachelor's degree advisor, for the research experience they provided me with and for their guidance during the early stages of my career development.

I would also like to acknowledge my committee members, Professors Sam Houk, Cheng Sheng Lee, Pat Thiel, and Clark Fugier Ford for their time and input in

my oral examinations and this thesis.

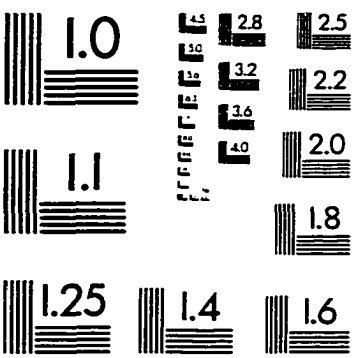
I thank all of Professor Yeung's group members, both past and present, for their friendship, support, cooperation, conversations about graduate school life and for helping me to keep a positive outlook on future. We had a great time together. All these friendships have made time in Ames a joyful experience.

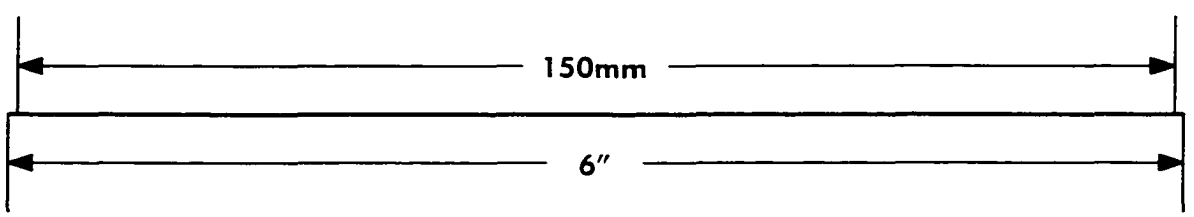
Finally, I dedicate this to my husband, to my son, to my mother, to the memory of my father, and to my in-laws. Their continuous encouragement and boundless support will always be the greatest motivation for me. My greatest regret is that my father, who cared for me the most, can not share this joyful moment with me.

To my husband, Jie Li, thank you for understanding and encouraging everything that I wanted to do for more than 10 years. Thank you for enjoying and sharing my every achievement since college, for accompanying me in the lab during nights, for backing me with all the love that I could ever ask for more, and for your commitment to grow old together. To my son, Brian J. Li, thank you for bring happiness into our whole family.

This work was performed at Ames Laboratory under Contract No. W-7405-Eng-82 with the U.S.Department of Energy. The United States government has assigned the DOE Report number IS-T1830 to this thesis.

IMAGE EVALUATION TEST TARGET (QA-3)





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989