1995

Laser-based detection system for high-throughput DNA sequencing with multiplexed capillary electrophoresis

Qingbo Li
Iowa State University

Laser-based detection system for high-throughput DNA sequencing with multiplexed capillary electrophoresis

by

Qingbo Li

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Chemistry
Major: Analytical Chemistry

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1995

TABLE OF CONTENTS

GENERAL INTRODUCTION

Dissertation Organization

This general introduction is followed by five chapters and the general conclusion section. Chapters 1, 3 and 5 represent journal manuscripts. The references for the numbered chapters follow the text of the chapter in which they are cited. The references for the General Introduction are listed following the General Conclusion. Listed in the appendices are the "C" codes of the programs used to run the SCM5000E charge injection device in this work.

Capillary Electrophoresis

Historical overview Electrophoresis has long been a separation technique for biological applications since its first practical demonstration shown by Tiselius in 1937 (1). This is a method by which charged molecules are separated based upon differential migration in an applied potential field. The introduction of polyacrylamide in the 1950's (2) as a separation medium for biomolecules revolutionized the electrophoresis techniques. Polyacrylamide remains one of the most common gel matrices used for biological separations today.

In 1974, Virtenen demonstrated the advantage of small inner diameter (i. d.) columns for electrophoresis (3). In 1979, Mikkers *et al* carried out

electrophoresis in polymer capillaries (4). Modern advances in capillary electrophoresis (CE) is ascribed to Jorgenson and Lukacs (5) who established the theoretical and experimental approaches to high resolution electrophoresis in thin glass capillaries in 1981. The capillary is filled with solution. Solutes can be separated based upon their differences in charge to mass ratios. Due to the small i.d. of the column and therefore the large surface-to-volume ratio, heat dissipation is very efficient. Higher voltages can be applied to achieve extremely high resolution. The technique is usually called capillary zone electrophoresis (CZE). Since then, CZE blossomed to be a powerful separation technique performing various functions with further developments. In 1983, Hjerten (6) adapted sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to capillary columns for capillary gel electrophoresis (CGE). This technique transferred the conventional slab-gel SDS-PAGE mechanism into small i.d. columns to take advantage of the efficient heat dissipation of small i.d. columns. Since CZE can only separate charged species, in 1984, Terabe et al (7) introduced micellar electrokinetic capillary chromatography (MEKC) to separate neutral compounds based on their differential partitioning into detergent micelles. In addition to the separation mechanism based upon size, charge, and hydrophobicity, Hjerten and Zhu (8) demonstrated in 1985 that the separation can also be achieved based upon differences in isoelectric points by using capillary isoelectric focusing (CIEF). Applications of CGE for protein and DNA

separations grew rapidly after Cohen and Karger demonstrated the separation power of polyacrylamide gel-filled capillary electrophoresis (9).

The typical setup of a CE system is more or less the same as the one shown in Fig. 1. A fused-silica capillary with i.d. of 10-100 μm and outer diameter (o.d.) of 150-360 μm is cut into a 30-100 cm length. The capillary is filled with the selected running buffer or gel matrix for separation. The two ends of the capillary are immersed in two buffer reservoirs. An electrode is inserted into each of the buffer reservoir. A high voltage from several kilovolts up to 40 kV is applied across the capillary by connecting the two electrodes to a high-voltage power supply. This provides the electric field necessary to separate the sample injected from one end of the capillary by either hydrodynamic or electrokinetic methods. The solutes running through the capillary can be monitored through a detection window usually located near or at the outlet end of the capillary by a variety of detectors (10). The signal is collected and analyzed with a computer.

Among the various CE separation techniques (11), CZE and CGE are the most widely used ones. The major features of them are discussed in the following text.

Capillary zone electrophoresis The fused-silica capillary wall has free silanol groups with a pK_a value of 3-4. Most running buffers have a pH value greater than 2. Therefore, when the capillary is filled with a running

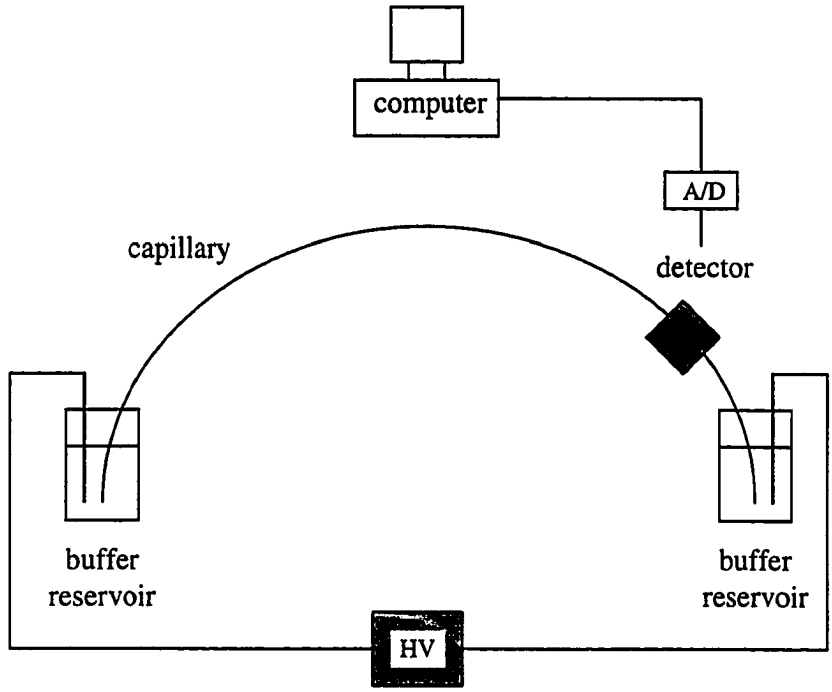


Figure 1. Schematic diagram of a capillary electrophoresis setup

buffer, the silanol groups ionize and form a negative charge layer. Positive ions in the running buffer are attracted to the surface and a double layer is formed. Under the influence of the electric field, the positively charged ayer carries the bulk of the running buffer and moves toward the cathode end of the capillary (Fig. 2). This movement of the bulk of the running buffer inside the capillary, so-called electroosmotic flow (EOF), is the important phenomenon occurring in CZE. The velocity of the electroosmotic flow, electroosmosis, can be given by

$$v_{eo} = \mu_{eo} E$$

where μ_{eo} is the electroosmotic mobility and E is the electric field strength. The electrophoretic velocity of a solute molecule is given by

$$v_{ep} = \mu_{ep} E$$

where μ_{ep} is the electrophoretic mobility of the solute. The net velocity of an ion in the capillary is the difference of its inherent electrophoretic velocity and the electroosmotic velocity.

$$v = v_{ep} - v_{eo} = (\mu_{ep} - \mu_{eo}) E$$

The separation efficiency, expressed as plate number, N, is given by

$$N = \frac{(\mu_{ep} - \mu_{eo}) V}{2D}$$

where V is the voltage applied across the capillary and D is the diffusion coefficient of the solute. The resolution between two solutes can be written as

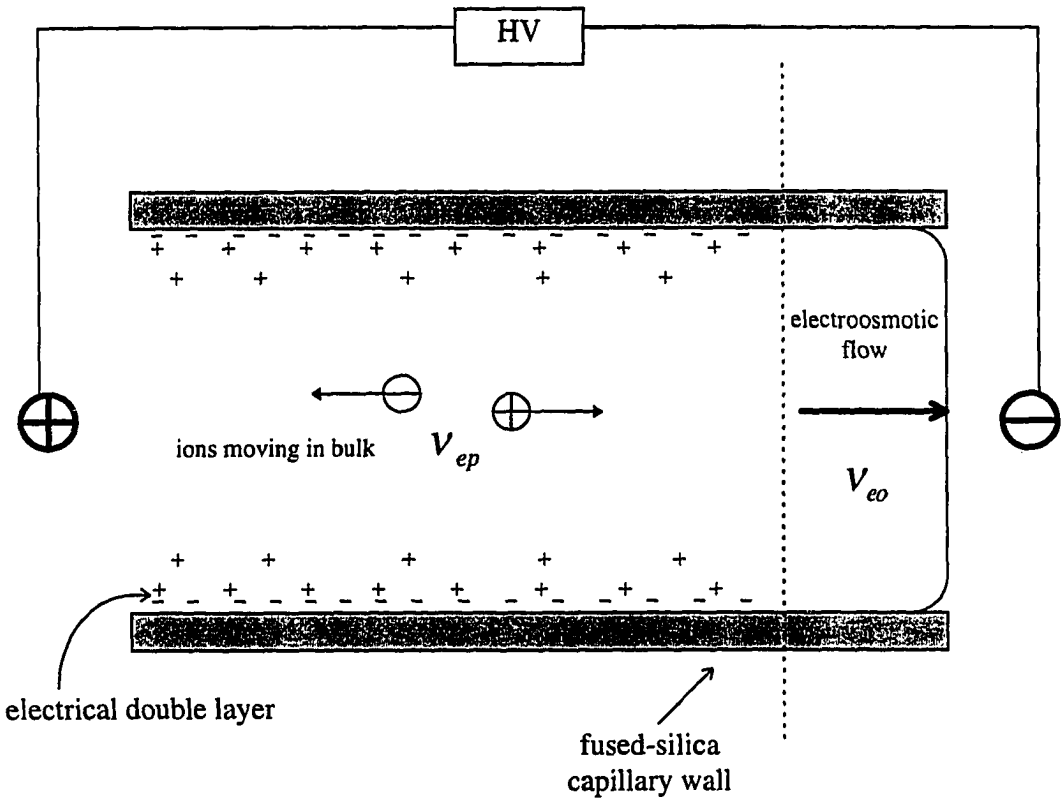


Figure 2. Diagrammatic representation of capillary zone electrophoresis.

$$R = \frac{(\mu_{ep,1} - \mu_{ep,2})\sqrt{V}}{4\sqrt{2D(\overline{\mu}_{ep} + \mu_{eo})}}$$

where $\mu_{ep,1}$ and $\mu_{ep,2}$ are the electrophoretic mobilities of the two solutes and $\overline{\mu}_{ep}$ is the average electrophoretic mobility. It can be seen that when $\mu_{eo} \equiv -\overline{\mu}_{ep}$, the resolution is maximized. The tradeoff is longer migration time. In fact, many state-of-the-art separations are achieved by controlling the electroosmotic flow or modifying the capillary surface property by adding electroosmotic flow modifier (12, 13), adjusting pH of the running buffer (14), coating the capillary(15), or using an external electric field (16).

Capillary gel electrophoresis Large biological molecules such as proteins and nucleic acids usually cannot be separated efficiently in CZE due to two aspects. First, the charge to mass ratios of these molecules are basically the same. Second, these macromolecules have higher hydrophobicity than smaller molecules or ions. They have a high tendency to adsorb onto the bare capillary wall. Separation of these molecules in CZE usually results in poor resolution, poor reproducibility and severe degradation of the performance of a capillary, if the capillary surface is not modified.

CGE has been developed to separate proteins and nucleic acids. Because electroosmotic flow would destroy the gel structure in a capillary, before filling a

capillary with gel, the silanol groups on the capillary wall need to be suppressed in order to eliminate the electroosmotic flow. This is conventionally achieved by coating the capillary with a thin layer of a polymer such as polyacrylamide (15). Most recently, Fung and Yeung (17) introduced a simple yet highly reliable column pretreatment method to suppress the electroosmotic flow in CGE separation of DNA. Before filling a capillary with polymer gel, 0.1M HCl is used to titrate the silanol groups to neutral form. The double layer on the capillary wall is eliminated and so is the electroosmotic flow. Because of the viscosity of the gel, the basic solution in the gel cannot reach the neutral silanol groups on the capillary wall within a period of time that is long enough for completing a DNA sequencing separation. The capillary can be regenerated and used many times. The advantage of this method over the conventional covalent polymer coating methods is that the capillary wall is not covalently modified. The pretreatment procedure is simple, fast, reversible, and very reproducible.

After the capillary is coated or pre-titrated, the capillary is filled with gel. A gel forms a sieving matrix that has characteristic pore sizes which are determined by the type, concentration and cross-linking of the gel. Since the electroosmotic flow has been suppressed, the movement of the macromolecules is only by electrophoretic mobility. Only charged species can be separated in CGE. Larger ions are obstructed more than the smaller ones and therefore move slower. Separation is thus based on the size difference of the ions (Fig. 3).

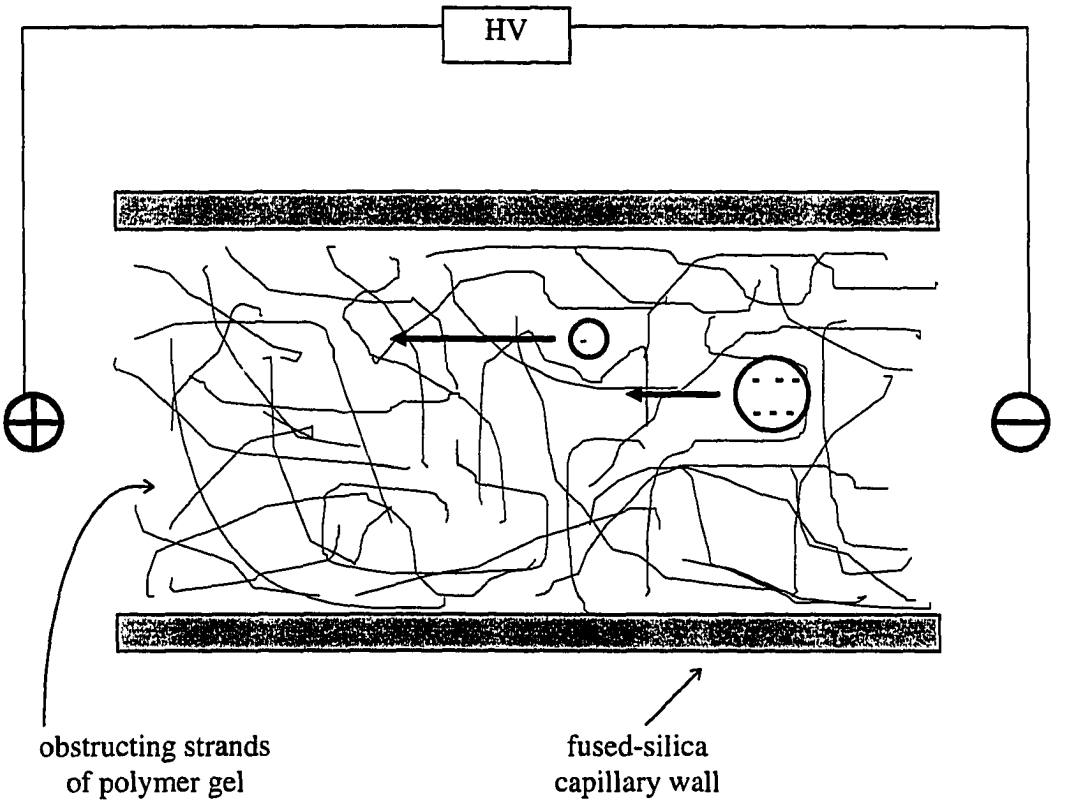


Figure 3. Diagrammatic representation of capillary gel electrophoresis. Larger ions move slower through the sieving matrix.

Cross-linked polyacrylamide gel has been used as the sieving matrix in CGE since its introduction. Although cross-linked polyacrylamide gel provides excellent separation efficiency (9), due to the difficulty of preparing cross-linked polyacrylamide gel-filled capillaries reproducibly and regenerating the capillary for multiple uses, application of the separation technique in automation and large scale routine operation is not practical. Replaceable non-cross-linked polyacrylamide gel for DNA sequencing (18-25) has attracted great attention in recent years, partly motivated by the Human Genome Project (HGP) (26). Although non-cross-linked polyacrylamide gel is less viscous and can be pumped more easily through a capillary than cross-linked polyacrylamide gel, the polymerization reaction in the gel preparation step remains a factor that is not easy to control in routine operations. Chang and Yeung (27) introduced a highly reproducible polymer matrix for DNA sequencing by mixing different sizes of polyethylene-oxide (PEO) polymer to optimize the separation for specific range of DNA fragments. The excellent reproducibility is mainly attributed to the simplicity of gel preparation and the stability of the gel. The low viscosity of the gel allows rapid replacement and fast separation while providing single-base resolution that enables correct base-calling for fragments at least as large as 340 base pair with 99% accuracy (28). Higher accuracy can be achieved by sacrifying some of the advantage of low viscosity of gel and high speed of separation.

CE Detectors

Overview The high separation power of CE could not be exploited fully if there were no compatible detectors. The small i.d. of the capillary contributes greatly to the separation efficiency in CE by allowing fast heat dissipation and thus higher voltage to be applied. Nevertheless, unlike in common HPLC, CE separation only handles much lower amount of analytes due to the limit of capillary size. The detectors used for HPLC cannot be adapted for CE without substantial modification and further developments.

The challenge of detecting very low amounts of analytes separated in CE incurs a lot of effort to develop state-of-the-art detection systems for CE separation. The development of CE detectors can be classified into three major categories: optical (10), electrochemical (29), and mass spectrometric (30). Among these three categories, optical detectors predominate in both commercial CE instruments and research labs, at least at the present stage. This is mainly due to the fact that optical methods are easier to miniaturize, more selective while covering wide range of applications, and more sensitive. Because of the limit of space and the relevance to the work in this dissertation, the following text only covers optical detectors, especially absorption detectors and laser-based fluorescence (LIF) detectors. A more comprehensive overview can be found in many fine reviews (10, 11).

Absorption detectors Absorption detectors have been widely used in commercial CE instruments, partly because they are readily adapted from their available LC versions, and partly because they are relatively universal, since most organic and biological molecules absorb light in the UV region. In absorption detectors, absorbance is measured by monitoring a small change on a large signal. The limit of detection (LOD) is limited by both the stability of the light source and the available dynamic range of the detector. In CE applications, two more challenges are imposed. Due to the small liquid core in a capillary, the light path length is much smaller, and light collimation is far from ideal for such a small liquid core if a ball lens is used. In addition, because much less light can be incorporated into a capillary with conventional lamps, shot noise becomes more significant in a CE absorption detector. Xue and Yeung (31) achieved 25 times improvement over the best absorption detector implemented in commercial CE instruments by using a laser light source and a noise-cancellation circuit with a double-beam absorption detection scheme. An LOD of 2×10^{-8} M of malachite green was achieved. But this is still much higher than the 10^{-13} M level reported in LIF detection (32).

Fluorescence detectors The signal in fluorescence measurements is much smaller than that in absorption detection. But since the background level in fluorescence measurements can be negligible under proper control, a very low LOD can be achieved because the LOD is determined by signal to noise ratio

(S/N) instead of the absolute signal strength. In addition to the advantage of low background, each atom or molecule can be excited thousands of times to produce thousands of photons provided that the excitation light source is intense enough. This inherent amplification capability renders the possibility of detecting a single atom or molecule with LIF detection (32). Rejecting stray light and suppressing background signals by proper optical and electronic design thus play the central role in pursuing the state-of-the-art LOD in LIF detection. Although the fluorescence intensity is still proportional to the probed volume, the short pathlength in CE is not the major problem for LIF detection, because a laser beam usually provides enough excitation power to produce so much fluorescence that the S/N is finally limited by background noise instead of the absolute fluorescence signal.

In a sheath flow cuvette detector, effluent from a capillary is confined within a very narrow stream. Excitation occurs at the capillary-free region where only liquid exists. Due to the elimination of stray light and reduction of Raleigh and Raman scattering from buffer solution, application of the detector to CE has produced some of the best LODs achieved in CE (32-34).

On-column detection is the most popular LIF detection mode used in CE because it does not involve complicated coupling design such as that in the sheath flow cuvette detector (32-34), although the later can achieve more impressive LOD by eliminating the stray light from the capillary. Three major

modes of on-column LIF detection have been demonstrated with impressive LODs (35-37).

A simple and rugged on-column LIF detector (35) with orthogonal geometry has been introduced and used successfully in analyzing chemical components in single red blood cells (38). Although the detector does not appear to be a complicated device, multiple fundamental principles are implemented in its elaborate design. Multiple optical components are involved to reject stray light, among which the most important ones are the laser beam focusing lens and the spatial filter in front of the photomultiplier tube (PMT). By using a biconcave 1-cm focal length lens, the laser beam can be focused smaller than the capillary inner diameter, which is usually above 50 μm, so that the laser beam passes through the liquid core of the capillary cleanly without causing tremendous scattered light from the capillary wall. If the capillary inner diameter is smaller than 50 μm, a diffraction limited lens is recommended. After the image of the detection window is focused clearly with a microscope objective, a spatial filter is used to reject the scattered light from the capillary wall by shielding the image of the capillary wall and only passing the image of the center liquid core in the capillary into the PMT. With this detector, a concentration of 3 pM fluorescein injected has been detected (35). At least four orders of linear dynamic range can be achieved.

Another type of on-column LIF detection employs an epi-illumination fluorescence microscope for both introducing excitation laser and collecting induced fluorescence (36). In this confocal geometry, a high-numerical-aperture microscope objective is used to focus the laser beam tightly to achieve high spatial resolution and thus excellent rejection of scattered light. The high numerical aperture of the objective also allows a large solid angle for light collection. Injected picomolar level fluorescein isothiocyanate-labled amino acids were detected. The linear dynamic range was over five orders.

Taylor and Yeung (37) introduced an axial-beam LIF on-column detection scheme by inserting an optical fiber into one end of a capillary to deliver laser light to excite solute bands. The induced fluorescence is collected with an optical fiber that is placed orthogonally to the capillary wall above the exit end of the excitation fiber. Collected fluorescence is delivered to a photomultiplier tube (PMT) for detection. 10^{-12} M of rhodamine 6G has been detected on the system.

Multiplexing CE detectors With the motivation of overcoming the throughput limitation in large scale DNA sequencing for the HGP, it is desirable to run multiple capillaries in parallel simultaneously. The primary barrier to multiplexed CE has been to develop a detector capable of detecting multiple capillaries without involving multiple detectors. In fact, all of the above LIF detection methods have been expanded to adapt simultaneous monitoring of multiple capillaries (39-42). At least at the present stage, each of these

multiplexing detection methods has certain limitations that keep it from efficiently involving in HGP. It is these barriers which the work in this dissertation aims to overcome. Further introduction on this topic remains until Chapter 4.

Single Cell Analysis

Single cell analysis has long been a disease diagnosis method in clinical applications. Traditionally, the diagnosis is based on direct observation of the shapes of cells or the reaction of cells to some specific staining reagents with the aid of a microscope. It has also been of great scientific significance to study biological events at cellular and subcellular levels. Modern biological studies of single cells involve much more complicated techniques. The common ones include flow cytometry (43), fluorescence microscopy (44), electronic microscopy (45), immunoassay (46), enzymatic radiolabelling (47), optical fiber sensors (48), and voltametric microelectrodes (49). Fluorescence and electronic microscopy provide excellent spatial resolution for studying the detail structures of cells. Combined with other biological and spectroscopic techniques, fluorescence microscopy can be used to monitor some specific species within a cell, either quantitatively or qualitatively. Flow cytometry is another well-established tool in cell biology. However, since a living cell contains hundreds or even thousands of different compounds, the methods mentioned above do not have the power to

screen multiple components within one single cell, especially when these compounds have quite similar structures and properties. It is thus desirable to separate these compounds before measurement. The quantitative information of a group of species and the correlation between these species within one single cell not only provide useful information for biological study, but also may lead to early disease diagnosis at single cell level which otherwise might be masked completely in bulk information of thousands of cells that is obtained by regular blood or tissue examinations.

In capillary liquid chromatography and capillary electrophoresis, the inner diameter can be as small as 2-10 μm. Single cells injected will not suffer unacceptable dilution. Excellent separation efficiency in CE and compatibility of aqueous separation buffer with biological molecules make CE suitable for analysis of wide range of intracellular species. Detection of these species is usually based on electrochemical method or LIF detection.

Separation of intracellular species at the single cell level first started from large cells such as snail neurons that have sample volumes of ~1 nL (50). Refinement of the procedure was applied to smaller cells such as bovine adrenal medullary cells with 1-2 pL sample volume (51,52).

Further miniaturization of single cell analysis methods leads to the analysis of human red blood cells (38). The red blood cells have the size of ~8 μm and the volume of ~90 fL. The small sample volume and small amount of

species present in these cells impose more stringent requirements on the sensitivity of the detection system. Yeung's group has successfully used LIF, particle-counting, and laser-generated microplasma emission detection methods to determine a series of species in individual red blood cells (38,53-62). Since the mass LOD is extremely low in these sensitive detection systems, contamination from the environment could interfere seriously with the determination of certain species that are also abundant in the environment, such as K and Na, if no precaution is taken (55). Methods developed for the analysis of single red blood cells can also be applied to analyze other mammalian cells (63).

The study of red blood cells indicates that there is large cell to cell variation of the amount of each intracellular species. The variation is related to the enzymes that regulate the species. Since enzyme activity within a cell is a function of the cell age, the amount of certain species can be used as age markers. It is desirable to perform multi-component measurements on each cell (57,58) or even different parameters of the same compound in one cell (59). The information thus obtained provides multi-dimensional analysis of each cell which could lead to the discovery of some important correlations that are useful for determination of the status of the cell or even disease diagnosis.

While the large cell-to-cell variation reveals the age distribution of cells within an individual, this inherent nonuniformity makes it difficult to exclude outliers among a large population of normal cells which have been quite diverse

already. For chromatography-based single cell analysis to be a truly competitive technique for early disease diagnosis, the technique cannot rely on presorting cells by staining or physical measurement such as size, shape, and density of cells, which have been the methods used in flow cytometry and microscopy. The exclusion criteria can only be established on the basis of statistical analysis of a large population of cells. Except for microscopy, flow cytometry and laser-generated microplasma emission, the present chromatography-based single cell analysis methods do not have the high throughput to scan a large number of cells in order to establish an accurate population distribution of normal cells that can be relied on to pick out abnormal cells. As one example, for measurement of intracellular K and Na by single-column CE (55), it takes 9 hrs to analyze about 70 cells even though the separation of K and Na is probably the fastest one compared to those of other intracellular species. The number of cells analyzed during such a long period is still not enough for establishing an accurate distribution model. Recent advances in multiplexed CE, although currently driven by the HGP, will surely shed a light onto the puzzle in the future.

High-throughput DNA Sequencing

Motivation Scientists are not satisfied with understanding our lives just at the individual cell level. Their fervent enthusiasm and wild imagination

has led to the initialization of an enormous scientific venture that could finally end up with providing us the ultimate understanding and control of human disease, aging, and death -- the Human Genome Project (HGP) (26, 64).

The goal of this project is to determine the precise location and molecular details of all of the genes and interconnecting segments that make up the human chromosomes (26). Since genes are the small but complex DNA segments that control the activities of the cells and thus the entire life of an individual, such information would eventually become an extraordinary powerful tool to explore the mysteries of human development, life, disease, and death at the molecular level (64).

The HGP has been undertaken in the scientific community that involves biologists, chemists, physicists, computer engineers, and even sociologists. The organization of the HGP is complicated and changes at different periods. On the basis of functions, the HGP can generally be divided into three major functioning categories (26,65):

1. Mapping and sequencing the human genome

    a) Genetic mapping: Complete a genetic map with markers spaced around 1 million base pair (bp). Identify each marker by sequence tagged sites.

    b) Physical mapping: Assemble sequence tagged site maps of all chromosomes with the goal of having markers spaced at about 100,000

bp intervals. Generate overlapping sets of cloned DNA of closely spaced unambiguously ordered markers.

c) DNA sequencing: Actually sequence the cloned DNA fragments at the lowest cost with newly developed techniques of high-throughput capacity.

2. Data collection analysis -- informatics

Develop effective new software and database designs to support the large-scale mapping and sequencing projects. Develop algorithms and analytical tools that can be used to interpretate the genomic information.

3. Ethical, legal and social issues

Develop programs aimed at understanding the implications of the HGP. Identify and define the major issues and develop initial policy options to address them.

It can be seen that DNA sequencing itself is just one of the steps in the HGP. However, the speed and accuracy of the DNA sequencing step plays a key role in the HGP.

There is no doubt that the outcome of the awesome undertaking of HGP will be nothing less than a revolution against the way science has traditionally progressed (64). But the project is forbiddingly expensive and will take forever if it is based on currently available DNA sequencing techniques. Under current

technology, 10^5 - 10^6 bases can be sequenced per year by a competent lab. Since the human genome consists of 3 x 10^9 bases, it will take 1,000 to 10,000 years to sequence the whole human genome. It costs $3 to $5 per base with today's technology, which will result in a 9 to 15 billion dollar expense to complete the HGP.

It is clear that without significant improvement of current DNA sequencing technology, the HGP is just science fiction. It is this challenge that the work in this dissertation is oriented toward. The final goal is to sequence the human genome with the cost of no more than $0.50 per base within an acceptable length of time.

Most recently, "startling messages" have been delivered by the project leaders of several automated DNA sequencing labs funded by the National Institutes of Health's National Center for Human Genome Research (NCHGR) and the Department of Energy (DOE) (66, 67). The scientists assert that the whole sequence of the 3 billion base pairs in human DNA can be obtained by 2001, five years ahead of schedule of the HGP, by starting the sequencing of the Genome right now without waiting for further development of new sequencing technologies. The cost will be 10 cents per base for sequencing, plus 2 cents per base for developing raw materials for sequencing. Overall, the Genome will be completely sequenced within the next five years for $300 to $400 million. Among the scientists, Robbert Waterston, director of the Genome Center at

Washington University in St. Louis, said that his own lab is now able to process 6×10^5 to 7.5×10^5 bp per week. By scaling up his present lab by 5 folds and running three labs in tandem to achieve a gross sequencing rate of 40 million bp per week, Waterson projected that it would take 5 years to cover 99% of the entire Genome with an overall accuracy of 99% to 99.9%.

Although the projections have raised some controversy over the figures of the cost and the time estimated from the experiences of sequencing the nematode genome which was done by the joint team of Robert Waterson and John Sulston, "the whole mentality of the field has undergone a substantial change in the past year...towards a feeling that it's time to start doing it," says Francis Collins, director of the NCHGR.

It is true that we don't have to wait. One will be more optimistic if he/she looks at the most recent advances in highly multiplexing DNA sequencing technologies. For example, with the combination of fast separation by capillary electrophoresis (30 bp per min) and simultaneous monitoring 100 capillaries, Yeung's group (17, 27, 28, 40, 42, 68, 69) has developed a multiplexed capillary electrophoresis technology capable of producing a raw sequencing rate of 3000 bp/min in a single instrument. The technology is readily scaleable to 1000 capillaries to achieve a gross sequencing rate of 40 million bp per day. That is 280 million bp per week. The 3-billion bp Genome can be sequenced within TWO MONTHS instead of five years with at least 99% accuracy.

Before proceeding to the chapters in the dissertation, it is helpful to provide some biological and analytical background of DNA sequencing. The remaining of this general introduction thus gives a glimpse of the principal developments of DNA sequencing techniques.

DNA sequencing history The major developments of DNA sequencing are summarized in Fig. 4. The historically significant effort of sequencing nucleic acids is ascribed to Holley who first sequenced yeast tRNA in 1965 (70). The breakthrough of DNA sequencing rate occurred in 1977 when two sequencing methods were introduced independently by Sanger (71), and Maxam and Gilbert (72). These methods provide the theoretical and practical background for later developments of modern sequencing technologies. For the pure purpose of DNA sequencing, the Sanger method predominates due to its ease of automation. Another important event in sequencing technology development is the introduction of fluorescent dye-terminators and dye-primer which enables the rapid growth of automatic techniques based on Sanger chemistry (74-76). Although other sequencing concepts have also been introduced in recent five years which include single molecule approach (77), atomic probe microscopy (78), mass spectroscopy (79), and sequencing by hybridization (80), the Sanger method remains the mostly widely accepted method, especially for large-scale automatic sequencing.

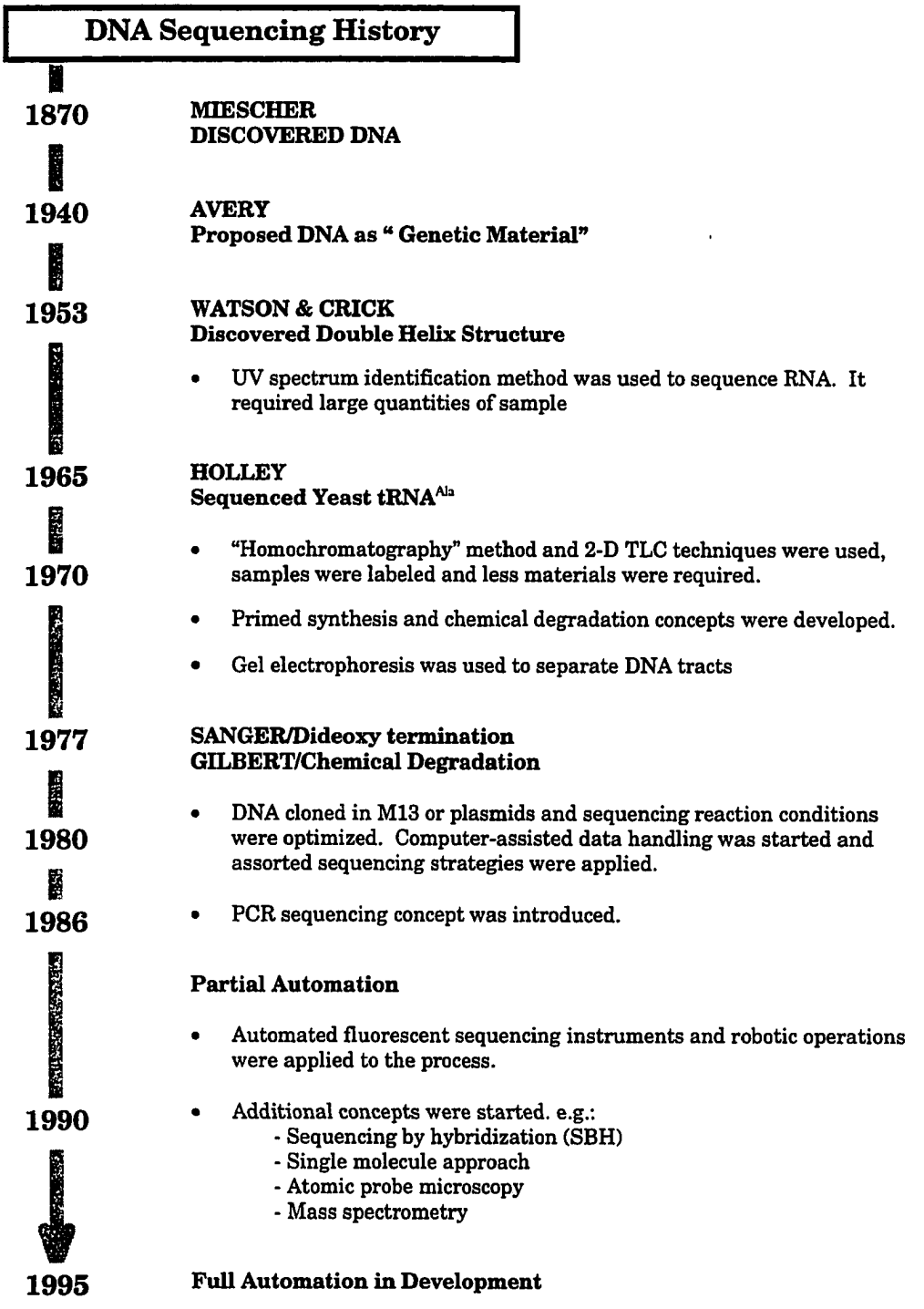
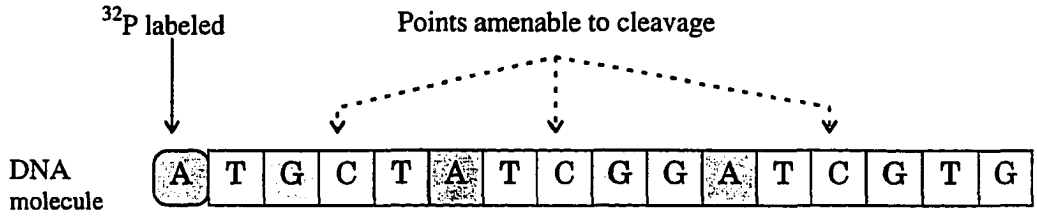


Figure 4. A historical view of DNA sequencing (adapted from ref. 73).

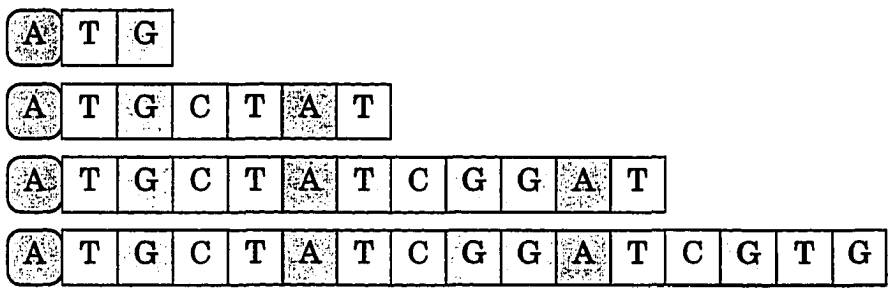
Maxam-Gilbert method The Maxam-Gilbert sequencing procedure (72) employs a chemical approach to cleave preexisting DNA for sequencing (Fig.5). First, the DNA to sequence is labeled with ^{32}P at the 5' end with polynucleotide kinase. Next, the DNA is treated with a chemical reagent that specifically reacts with one of the four bases. The reaction time is controlled so that on the average only a few bases on one DNA chain react with the reagent. The modified bases introduce the spots on the backbone where the DNA chain is amenable to cleavage by subsequent chemical treatment. Approximately all the bases of the given type in a DNA chain are equally susceptible to modification. The net result is the production of a family of 5'-end ^{32}P-labeled fragments that terminate at the points specific to the given type of bases. Electrophoresis of this product gives the location of the given type of bases on the original DNA chain.

To find out the complete sequence of the DNA chain, four reactions are used. The conditions can be adjusted so that the four reactions cleave DNA preferentially at guanines (G>A), adenines (A>G), cytosines alone (C), and cytosines and thymines equally (C+T). Electrophoresis of the four reaction products are carried out in four lanes on the same slab gel. After the autoradiogram of the sequencing gel is obtained, the whole sequence can be read by tracking the individual bands one step at a time. The bands on the C lane indicate the location of C bases. The T bases can be determined from the

a) DNA labeled at one end with ^{32}P and the points on the chain amenable to C-specific cleavage chemical reaction



b) DNA fragments produced after chemical cleavage. Only the fragments with the original ^{32}P label can be detected with autoradiography.



c) Pattern of the autoradiogram after the reaction product is separated on a gel.

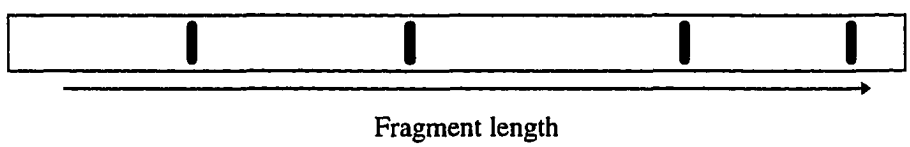


Figure 5. Diagrammatic representation of the Maxam-Gilbert method.

bands in the C+T lane by comparing the C lane and the C+T lane together. The C bands in the C+T lanes can serve as markers for more accurate determination of T bands since the pattern of the C bands have been known from the C lane. G and A bases can be determined in a similar way.

Sanger method The Sanger method (71) employs chain-terminating dideoxy-nucleotides to produce a continuous series of fragments in reactions catalyzed by polymerase. Dideoxynucleotides (ddNTPs) resemble deoxynucleotides (dNTPs) except that ddNTPs lack a 3'-OH group. The 5'-end phosphate group allows a ddNTP to add to a growing chain just like a dNTP during polymerization. But the ddNTP cannot be added to by a subsequent dNTP or a ddNTP because it does not have the 3'-end OH group. The growing chain thus stops at the point where a ddNTP is picked up during polymerization. The ddNTPs are thus called chain terminators, or simply terminators.

The Sanger reaction includes the following major steps (Fig. 6). First, a suitable primer is chosen so that the polymerization synthesis can start at the point of interest. The primer is annealed to the template DNA that is to be sequenced. The primer-template complex is in a reaction solution which also includes polymerase, all four types of dNTPs, and a small amount of one type of ddNTP. Starting from the primer and with the catalysis of the polymerase, a complimentary chain of the single-stranded template DNA keeps growing by picking up a dNTP that is complimentary to the base at each point. There

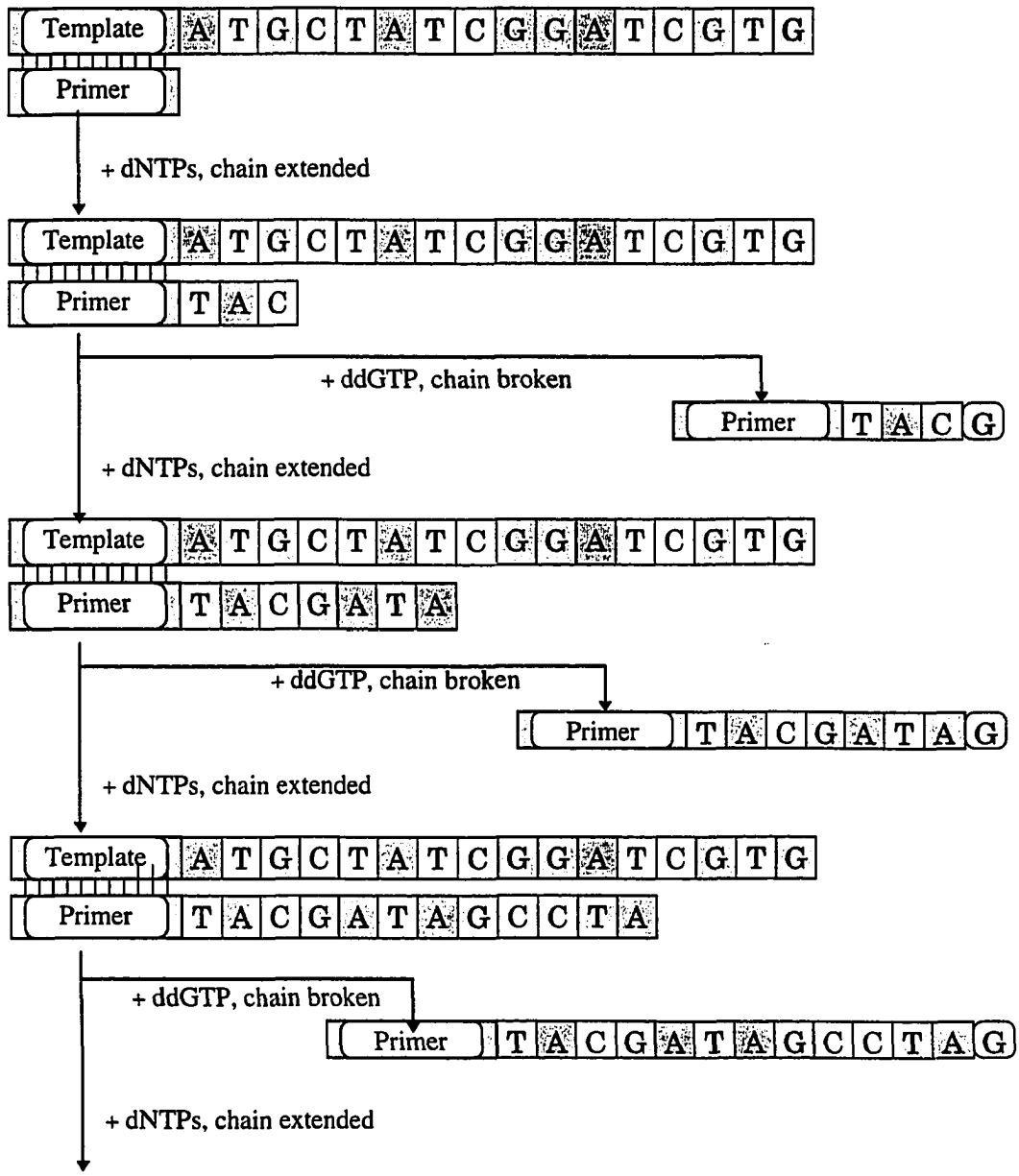
Figure 6. Diagrammatic representation of the Sanger method. The reaction solution contains template DNA, primer, four dNTPs, one ddNTP, and polymerase. In this figure, it is assumed that the primer is ^{32}P-labeled, and ddGTP is used. The reaction produces the set of fragments for sequencing C bases. The major steps include:

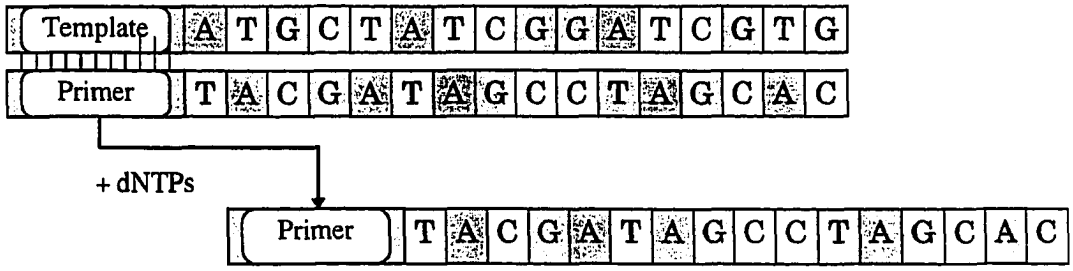
a) A suitable primer is attached to the point of interest on the template DNA, usually at the 5' end of the template. The polymerase picks up dNTPs complimentary to the bases on the template chain and synthesizes the complimentary chain by starting from the primer. If no ddGTP is picked up, the chain extension continues to the end. If a ddGTP is picked up at a point at which the base on the template is C, the chain extension stops.

b) The fragments produced after the Sanger reaction are separated on a gel.

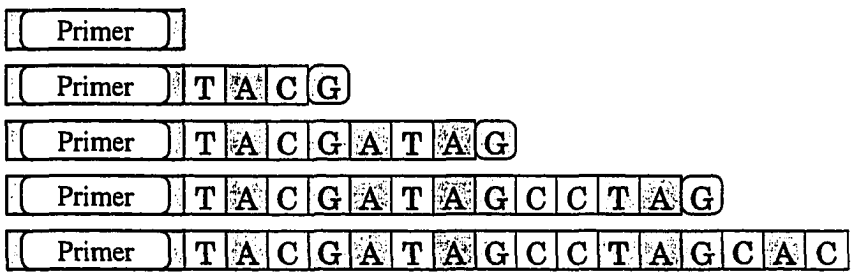
c) The autoradiogram shows the bands corresponding to the C bases on the template DNA.

a) Primer aneals to the template. Chain extention begins.





b) Fragments produced after the Sanger reaction



c) Autoradiogram of electrophoresis of the Sanger reaction product

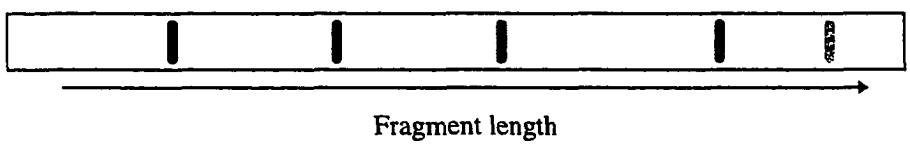


Figure 6. (continued)

is also a chance to pick up a ddNTP of the given type at the point where the base on the template DNA is complementary to the ddNTP. If a ddNTP is picked up, the chain stops. The length of the chain indicates the location of the base on the template DNA the type of which is registered by the ddNTP. Electrophoresis ladder of the reaction product shows the locations of the given type of bases on the template DNA. If four reactions are used and the reaction products are run respectively in four separated lanes on one slab gel, all of the bases on the template DNA can be sequenced.

Traditionally, the Sanger reaction products separated on slab gel are analyzed by autoradiography. Either the primer or the ddNTPs can be labeled radioactively. Alternatively, the dNTPs can be labeled. The bases are thus indicated by the bands on the autoradiogram.

Recent advances of Sanger chemistry introduce fluorescent dye-labeled primer and terminators. The advantage of these approaches is that the four reactions corresponding to an unknown DNA sequence can be carried out in a single tube and loaded onto a single slab-gel lane or a single CE column for electrophoresis. The four sets of fragments are then sorted spectroscopically. This has greatly facilitated the automation of DNA sequencing. It also provides the fundamental DNA sequencing protocols upon which multiplexing DNA sequencing techniques are developed. The most popular set of four dyes for

labeling primers are FAM, JOE, ROX and TAMRA (81). Dyes for labeling terminators vary but are basically fluorescein and rhodamine derivatives.

Base-calling schemes Based on the Sanger chemistry using fluorescent dye-primer or dye-terminators, various optical designs and algorithms have been developed to call bases for DNA sequencing. It is unlikely that the four sets of fragments produced by the Sanger reaction can be run in four different capillaries to determine the sequence like in a slab-gel because of the relatively large variation of migration time in CE among a group of capillaries. Basically all of the base-calling schemes using fluorescent dye labels employ color sorting to distinguish the four sets of fragments that are run on one column or one lane.

Single-color four-intensity coding In this scheme (75,82-85), the primer is labeled with one fluorescent dye. Four reactions are carried out separately and then mixed together. By adjusting the dNTP/ddNTP ratio in each reaction, the sample is prepared to yield a nominal ratio of peak height for the four types of bases, e.g., 8:4:2:1. A simple optical setup is used and high light-collection efficiency can be achieved. The system (75) has been commercialized by Pharmacia. One problem associated with the method is that the variation of incorporation efficiency of a ddNTP causes ambiguity in this peak height coding, even though the manganese-T7 polymerase reaction is used. Because of the large peak height difference between the four types of fragments and the limited

dynamic range of the detector, the S/N for the set of fragments encoded by the smallest peak height is usually poor. Ghost peaks associated with false pauses in the Sanger reaction causes particular problems for the bases encoded by this smallest peak height. Therefore, this scheme is presently the least desirable.

Two-bit binary coding The 2-bit binary coding method (86) employs two-color detection and requires two dyes that can be used to label primers. Binary combinations of different dye-labeled primers are used to distinguish the four sets of DNA fragments. As an example, FAM-labeled primer and JOE-labeled primer can be used to produce a coding system of (1,1), (1,0), (0,1), (0,0), where 1 indicates that a band (peak) appears, and 0 as null. The two dyes are excited with one laser and two-spectral-channel detection is employed with the red spectral channel for monitoring JOE and the green channel for FAM. The A-fragments are prepared with both FAM-labeled and JOE-labeled primers. These A-fragments are detected in both channels and thus denoted as (1,1). The T-fragments are prepared with only FAM-labeled primer. These fragments are detected only in the red channel and denoted as (1,0). Similarly, the-G-fragments are prepared with only JOE-labeled primer and can be denoted with (0,1). C-fragments are not synthesized and thus denoted with nulls (0,0) which correspond to the gaps in both channels. The disadvantage of the method is that the lack of detection of the null-coded fragments (0,0) would lead to ambiguity if there is residual secondary structure or anomalous migration of DNA fragments

due to insufficient denaturing power of the gel. One solution to reduce this ambiguity is to do two runs by exchanging the (1,0) and (0,1) coding for the T- and G-fragments while keeping the coding for A- and C-fragments . Since this approach requires an additional run, it is less desirable if high-throughput is a major concern.

Two-label two-color four-ratio coding Generalization of the binary coding method leads to a four-ratio coding system in which the four sets of fragments are labeled with four different mole fraction combinations of the two dyes (86). The null code (0,0) is replaced by a unique ratio. The ratios are determined both by the mole ratio of the two dye-labeled primers and the spectral filters used. The advantage of the method is that sequencing is done with only one run while the ambiguity caused by null coding is avoided. Since all peaks need to be measured quantitatively, the requirement of S/N is higher than that in the binary coding. More attention must be paid to fine tuning of the mole fraction ratio of the two dye-labeled primers in Sanger reactions.

Two-label two-color two-intensity coding Tabor and Richardson (87) found that use of Sequenase in Mn^{2+} containing buffer remarkably improves the consistence of termination probability for each set of fragments throughout the sequence. This makes it possible to use the amount of a set of fragments as an additional coding tag. If one uses two dyes to label primers and uses each dye-labeled primer to prepare two sets of fragments, the four sets of fragments can

be sorted into two groups by two-color detection. Within each group, if the two sets of fragments are prepared in distinctive amounts by using different dNTP/ddNTP ratio, the difference of the intensities of the two sets of fragments can distinguish one set from the other (24, 82, 84, 88). FAM and JOE are usually used for this two-color two-intensity coding system because the two dyes have minimal mobility shifts and both can be excited with a 488-nm laser efficiently. Due to some overlap of the fluorescence spectra of these two dyes, in order to achieve longer read, software correction for the spectral interference can be applied (24). The advantage of the two-color two-intensity coding method is that chemistry is relatively simple compared to the two-label two-color four-ratio coding method while accuracy is improved. However, using Sequenase in Mn^{2+} containing buffer produces shorter sequencing length compared with the standard Sequenase method (87).

Four-label two-color four-ratio coding This method was introduced by using a set of four succinylfluorescein dyes to label the four terminators and employing a two-color detection system (76). Generation of all four sets of fragments is carried out simultaneously in one single tube. A 488-nm laser beam is scanned across the slab-gel to excite the fluorescence from one or multiple lanes. Two PMT detectors are set up side by side above the separation lanes to collect fluorescence in each half of the collection solid angle. Two different band pass filters are used on each PMT. The ratio of the intensities of

each fragment at both spectral channels is calculated. Each set of fragments have a ratio different from another three sets of fragments. The four ratios, which are determined by the filter selection and the light collection efficiency of each detector, are used to encode the four sets of bases in the sequence. The intensity ratios are basically independent of the incorporation efficiency. The chemistry procedure is thus simplified. The system was commercialized by Du Pont. Twelve lanes can be run simultaneously with the overall sequencing rate of 600 bp per hr. Now, the system is no longer produced. This may be partly because the four dyes have quite close fluorescence spectra. Very narrow band pass filters need to be used to improve the ratio difference, but they also decrease the S/N. Similar spectra also cause interference between the two spectral channels. These two factors cause ambiguities.

Four-label four-color coding This method uses four different dyes to label the four sets of fragments followed by separation on one column and sorting with four-color detection (74, 89). The primer is labeled with the standard four dyes: FAM ($\lambda_{ex,\,max}$ = 493 nm, $\lambda_{em,\,max}$ = 519 nm), JOE ($\lambda_{ex,\,max}$ = 526 nm, $\lambda_{em,\,max}$ = 548 nm), TAMRA ($\lambda_{ex,\,max}$ = 559 nm, $\lambda_{em,\,max}$ = 578 nm), and ROX ($\lambda_{ex,\,max}$ = 580 nm, $\lambda_{em,\,max}$ = 605 nm). The four sets of fragments are generated with the four dye-labeled primers respectively. The four reaction products are then mixed and separated in one lane. The detection system is relatively complicated. Due to the difference of optimal excitation wavelength between FAM/JOE pair and

ROX/TAMRA pair, multiple excitation lines are usually preferred. For example, both 488-nm and 514-nm lines from an argon ion laser are used to excite all of the four dyes in the ABI 373A sequencer, which is marketed based on Smith's system (74, 89).

One straightforward approach (89) to achieve four-color detection is to split the fluorescence image of the detection window into four using three 50/50 beam splitters. Four PMT detectors are used to monitored the split images. A narrow band pass filter centered at the emission maximum of one dye is placed in front of each PMT so that each set of fragments are detected in one spectral channel but another three.

A modification of the four-PMT optical setup is to use a rotating filter wheel to transmit sequentially fluorescence in four spectral bands to a single PMT (34). This detector offers the advantage of low component count and simple alignment compared with the beam-splitter arrangement.

Alternatively, the 488-nm line from an Ar^+ laser and 532-nm line from a YAG laser can be combined to excite FAM, JOE, ROX and TAMRA together. Wedged prisms can be used to split the fluorescence image to obtain four distinctive images on an array detector (41). Four band pass filters are attached on the wedged prisms to form the four-color detection scheme.

To call bases, the four chromatograms for the four spectral channels are overlaid. Ideally, each set of fragments only show up in one spectral channel.

Base-calling is very straightforward because each base is indicated by its color and its sequence is its position in the overlaid graphs. But due to the overlap of the fluorescence spectra of the four dyes, some extent of spectral crosstalk between the four channels exists. Another problem is the mobility shift caused by the four dyes attached to the DNA fragments. These two problems need to be corrected before the accurate sequence can be determined. The former problem is solved by running the same data points in the four channels through a multi-variate analysis algorithm to calculate the actual pure single-color intensity in each channel. The latter is solved by normalizing the time scales of the four channels. The system thus requires calibration before performing DNA sequencing. After the two software corrections, the system provides high base-calling accuracy. But the complicated base-calling calculation takes computer time which becomes one of the major concerns when the system is scaled up for large-scale sequencing.

ABI has also implemented dye-terminator chemistry for the four-color coding method on its 373A sequencer by using the PRISM fluorescent dyes. With either dye-primer or dye-terminator chemistry, twenty-four lanes can be run simultaneously on the 373A Sequencer. The sequencing speed is typically 10-12 hrs per run. With 24 lanes running together, the raw sequencing rate can be 1200 bp/hr.

Monochromator-based coding Instead of sorting the four dyes by using four optical filters, the spectrum can be dispersed with a monochromator and the four dyes can be sorted out by integrating the correct ranges in the spectrum that is recorded with an array detector. In one approach (90), a 488-nm line from an Ar+ laser and a 543-nm line from a helium-neon laser were used. The two laser lines do not have to be made collinear to excite one detection window. Instead, the two lines can be used to illuminate two detection windows respectively. The image of each detection window is focused with a microscope objective and goes through a spectrometer. The dispersed fluorescence spectrum is recorded with a cooled photodiode containing 1024 diodes each of which has the size of 25 μm x 2 mm. The two dispersed spectra from the two windows are aligned so that they can be monitored with different parts of the same photodiode with a resolution of 1 nm per pixel. The four electropherograms each of which corresponds to one set of fragments are built from the recorded on-the-fly spectra. The time scale is corrected for the two detection windows by software. Four-color coding is used to identify the bases.

Another approach (91) uses only one 488-nm line for excitation. The fluorescence image is dispersed with a monochromator and monitored by a cryogenically cooled CCD detector. The CCD is operated in binning mode with the resolution of 1.6 nm per detector element. The linear array of 96 detector elements covers the spectrum range between 500 nm and 650 nm where the

emission spectra of the four dyes fall. Before DNA sequencing, a 96 x 4 standard spectral response matrix is calibrated by running one dye at a time through the gel-filled capillary. Each recorded on-the-fly spectrum is used to solve for the 4 x 1 vector each element of which represents the intensity for a dye. The four electropherograms each of which corresponds to one set of fragments are built from the series of calculated 4 x1 vectors. The remaining base-calling step is a straightforward four-color coding procedure.

High-throughput techniques From the aspects of optical setup, base-calling method, and sequencing speed, the currently available commercial sequencers can not be used for the HGP without fundamental improvements. Nowadays, the sequencing rate achieved on one single CE column is comparable to that on the 24 lanes in a 373A sequencer. By running 100 to 1000 capillaries in parallel, several order improvement in sequencing throughput can be expected. However, such a highly multiplexed CE system can only be made possible by developing new multiplexing detection system, base-calling algorithm, and CGE separation system that are amenable to multiplexing and full automation.

CHAPTER 1. CONTAMINATION CONTROL IN CAPILLARY ELECTROPHORESIS AND QUANTITATIVE MEASUREMENT OF POTASSIUM AND SODIUM IN SINGLE HUMAN ERYTHROCYTES

A paper published in the Journal of Capillary Electrophoresis

Qingbo Li and Edward S. Yeung*

ABSTRACT

A highly sensitive detection scheme such as laser-induced indirect fluorescence detection is prone to interference from contamination. Contamination due to particulate matter in the determination of inorganic cations in single cells was investigated in detail. Cell manipulation was also optimized to minimize the loss of intracellular analytes. The mass detection limit of the system was 110 amol for potassium and 70 amol for sodium. The signal-to-noise ratio for potassium was high enough for quantitative measurement in individual human erythrocytes, although that for sodium was marginal. The measured quantities were used to obtain the population distribution of the cells.

INTRODUCTION

The world we live in is a contaminated one. Contamination from particulate matter is very common in the environment since it can be present everywhere. In most materials and processes, particulate contamination is negligible because of the small amounts involved. It is well known that particulate contamination is a serious problem in the semiconductor and electronics industry. The problem is effectively controlled with cleanroom technology. On the other hand, particulate contamination has not drawn much attention in certain microchemical methods, for example, the determination of ultratrace inorganic ions by capillary electrophoresis (CE). This is partly because the detection systems have not been sensitive enough to sense the contamination of inorganic ions from particulate matter. Although some detection systems can detect very low concentrations, the absolute detection limit is high because the total (vs. injected) sample volume required is still quite large. Soluble inorganic ions from a particle, if any, will not normally cause interference because these would be diluted in the sample of a relatively large volume. However, in a highly sensitive detection system such as laser-induced indirect fluorescence detection, both low concentration detection limit and low mass detection limit can be achieved (1). If a particle is injected, the amount of soluble inorganic salts would be in the range of the mass detection limit. At this point, particulate contamination is no longer negligible.

Laser-induced indirect fluorescence detection has been employed to detect intracellular cations and small organic ions in single human red blood cells (2,3). Here we used a smaller cationic fluorophor, 2-amino-pyridine (2AP), as the background electrolyte for detection and quantification of intracellular potassium and sodium in single human erythrocytes. Better detection limits were obtained than when a larger fluorophor was used. The detection system has thus also become sensitive enough to detect soluble metals in particulates if a particle is inadvertently injected. In the early stages of this study, the reproducibility for determining K and Na was very poor. This was later found to be due to the contamination of K and Na from the environment. After narrowing down the contamination sources, we realized that particulate matter was the major source of interfering K and Na. Soluble inorganic salts indeed exist in significant amounts in airborne particles (4a-4d).

In addition to the contamination problem, the procedure for cell sample preparation was the other major concern of this work. It was observed that the freshness of cell samples affected the measured K and Na levels in individual cells. Intracellular K and Na levels would decrease over long times or during low-temperature storage of a blood sample (5). Untimely suspension of cells in a low ionic strength solution also caused the loss of intracellular K and Na. In this paper, optimization of the sample handling, separation and detection conditions are described. The results of single-cell studies are treated statistically and com-

pared with those obtained by the method of laser vaporization combined with atomic emission. The advantages and limitations of this method are examined.

EXPERIMENTAL SECTION

Instruments The CE separation system used was similar to previously described instruments (1,2). Excitation was provided by 334-nm light in 125 mW of combined 334-350-360-nm radiation from an Ar ion laser (Coherent, Innova 90, Palo Alto, CA). A 1-cm focal length quartz lens (Melles Griot, Irvine, CA) was used to focus the beam within the 15-μm capillary. A 20X microscope objective (Edmund Scientific, Barrington, NJ) was used to collect the fluorescence image. A 395-nm long-pass filter (Melles Griot) was used together with a 405-nm interference filter (Edmund Scientific). A photomultiplier tube (Model R928, Hamamatsu, Bridgewater, NJ) was used to detect the fluorescence signal. An electrometer was used to convert the PMT current to an analog voltage output. A 24-bit A/D interface (ChromPerfect, Justice Innovations, Palo Alto, CA) was used to acquire data. Data was collected and stored in a 286 PC computer (Electra Data System). For CE separation, a 50-ml polyethylene narrow-mouth bottle (Nalgene, Nalge, Rochester, NY) was used as the buffer reservoir at the ground end. A 250-ml polyethylene wide-mouth bottle (Nalge) was used as the buffer reservoir at the positive high-voltage end. The 50-ml bottle was made airtight with a septum. The separation capillary was inserted

through the septum. A 20-ml gas-tight sterile syringe (Becton Dickinson, Rutherford, NJ) was coupled to the 50-ml bottle via the needle of the syringe. Vacuum or pressure at the 50-ml bottle was provided by manually controlling the syringe. To avoid contamination from hands, plastic gloves were always worn during the experiments. By use of a micropipet with a 200-µl pipet tip (USA/Scientific Plastic), a droplet of 5-10 µl of cell suspension solution was transferred onto a clean plastic slide (Thomas Scientific, Swedesboro, NJ) which was located on a microscope stage. For cell injection, the injection end of the capillary was mounted onto a micromanipulator (The Micromanipulator Co., Model 512). The orifice of the capillary was guided close to the selected cell and a vacuum pulse was applied at the ground end of the capillary to inject the cell. The capillary was quickly dismounted from the micromanipulator and put back into the 250-ml bottle. A 30 kV high voltage was applied immediately with a high voltage power supply (Spellman, Plainview, NY) to begin the separation. The cell injection process usually took 1-2 min.

Cell sample preparation A 2-µl fresh whole blood sample was collected from the cleaned fingertip of an apparently healthy adult male by use of a lancet (Boehringer Mannheim, IN) and a micropipet. The collected blood sample was immediately diluted in 14 ml wash buffer (250 µM 2-amino-pyridine and 8% glucose at pH 7.2) which was in a 15-ml polyethylene centrifuge tube (Corning, Corning, NY). The tube was inverted immediately to prevent clogging.

After centrifugation at 1200 rpm for 4 min, the supernatant was removed with a polyethylene transfer pipet (Fisher Scientific, Fair Lawn, NJ). The washed cells were suspended in 3 ml wash buffer for immediate single-cell injections. The cell sample preparation process took 5-6 min.

Separation A 15 μm i.d., 365 μm o.d. bare fused-silica capillary (Polymicro Technologies, Phoenix, AZ) of 60 cm total length and 40 cm effective length was used without surface pretreatment. The running buffer was made up of 250 μM 2-amino-pyridine and 1.5% glucose, and adjusted to pH 4.8 with acetic acid. Before use, the capillary was flushed with a 1:1 methanol-water mixture for 30 min. The capillary was then filled with the running buffer and equilibrated at 20 kV for 3 hours before use. Standards were injected with gravity at a height of 16 cm for 20-60 s.

Reagents 2-Amino-pyridine was obtained from Aldrich (Milwaukee, WI). Acetic acid, potassium chloride, sodium chloride, lithium chloride, and glucose were all obtained from Fisher.

RESULTS AND DISCUSSION

Separation and detection In indirect fluorescence detection, a fluorescing ion is the main component of the running buffer. The fluorophor is selected to have several characteristics. First, the fluorophor should have a high fluorescence quantum yield, so that excitation and stray light rejection do not

present problems even at low concentrations. Second, the electrophoretic mobility of the fluorophor should be close to those of the analytes so that sharp and symmetrical peak shapes can be obtained (6). Close matching of mobility between the fluorophor and an analyte also favors the transfer ratio between them. In addition, indirect detection should be carried out under moderate pH conditions, i.e., pH 3 to pH 9. Matching the pH of the buffer with that of the sample will avoid artifacts due to a pH jump at injection. With well-behaved physical and chemical properties (7), 2-amino-pyridine (2AP) is very suitable for the determination of alkali and alkali earth metals. 2AP has a pK_b value of 7.18 (8). This is ideal for indirect detection because 2AP becomes singly charged at pH 4.8. Higher sensitivity and a lower detection limit are thus achievable with 2AP compared to previous results (1,2).

The fluorescence intensity of 2AP reaches a maximum at pH 4.8. Further increase of acidity did not result in a higher fluorescence intensity. At lower pH, the transfer ratio between 2AP and metal ions was not favorable. The detection limit at pH 3.0 degraded by a factor of 3 compared to that at pH 5.7. The transfer ratio between alkali metal ions and 2AP was found to be close to unity, with a value ranging from 0.8 to 0.2. The metals with mobilities closer to that of 2AP had a higher transfer ratio, and thus exhibit larger peak areas and sharper peak shapes.

Hydrochloric acid and acetic acid (HAc) were used to adjust the pH of the running buffer. No difference was observed in the detection limit and sensitivity by using HCl or HAc. Since the pK_a of acetic acid is 4.7, which is very close to the pH we selected for the running buffer, acetic acid was selected so that the buffer capacity of the solution can be higher.

The CE system was primarily optimized for sensitivity instead of separation. According to earlier studies (9), a lower concentration of the background electrolyte favors detection performance. However, when the buffer concentration was lowered, electroosmotic flow velocity increased and the separation efficiency decreased. Between pH = 4 to pH = 5, when the buffer concentration is below 200 μM, potassium and sodium could not be separated. The buffer concentration was thus adjusted to 250 μM for this work.

Cell suspension matrix and cell contents affected the stability of the performance of a bare fused-silica capillary. In previous work (2), a capillary was coated with a hydrophobic polymer and CTAB was included in the running buffer to alleviate the matrix effect. However, due to the reversal of electroosmotic flow, migration times of alkaline metal ions were long. Peak dispersion was serious for late eluting ions such as potassium. In addition, coating small inner diameter capillaries required a difficult and tedious procedure. In this work, we found that adding a low concentration (20 to 100 μM) of CTAB into the running buffer significantly alleviated the matrix effect

without reversing the direction of electroosmotic flow. Analyte peaks were eluted before system peaks. Peaks became sharper and interferences were minimized. However, after CTAB was added into the running buffer, the baseline became noisier and detection sensitivity decreased by a factor of 2. When cells were injected, the migration time of analytes increased by 35% after forty runs. The continuous change in migration times indicates that CTAB did not totally prevent the matrix material from adsorbing onto the capillary walls. Although the pattern of the electropherogram was reproducible and the peaks of potassium and sodium were not difficult to assign, significant changes in migration times can affect the accuracy of peak area measurements. It was found that glucose could be used as an additive in the running buffer to alleviate the matrix effect in cell analysis. Adding glucose into the running buffer also decreases electroosmotic flow and helps to improve separation efficiency, although the improvement was not as large as that when CTAB was used. Adding glucose into the running buffer did not cause extra baseline noise, but a slow drift in the baseline was sometimes observed, especially when smaller i.d. capillaries were used. This is probably due to the higher viscosity of the running buffer. With glucose in the running buffer, sensitivity was not affected. Therefore, 1.5% glucose was included in the running buffer. This was necessary because the concentration of 2AP was only 250 μM and the separation of potassium and sodium was marginal without electroosmotic flow modifier. No equili-

bration was needed between consecutive runs in single-cell analysis. After seventy cell injections, the migration times only increased by 12%. An experienced experimenter could analyze up to 12 cells/hr on such a system.

We also tried to include both CTAB and glucose in the running buffer so that lower 2AP concentrations could be used to achieve a lower detection limit. However, with glucose in the running buffer, adding CTAB had little effect on the migration time and the separation efficiency. This may be due to the acidic pH of the buffer and the competitive adsorption of glucose onto the capillary wall.

Figure 1 and Figure 2 show the typical separation and detection for standards and for a single cell. The pattern of the electropherograms is quite characteristic. There are two system peaks. The latest eluted system peak (SP2) corresponds to the neutral components (electroosmotic flow). It indicates the end of the separation. The other system peak (SP1) corresponds to the migration of 2AP. The two system peaks may be positive, negative, derivative shaped, or absent entirely. This depends upon the pH and composition of the injected sample. When water was injected, both system peaks were negative. When the buffer itself was injected, SP1 did not appear and SP2 was positive. When a sample prepared in the running buffer was injected, SP1 did not appear if the sample concentration was low. SP1 appeared as a positive peak when a high concentration of sample was injected. If the injected sample was at a pH

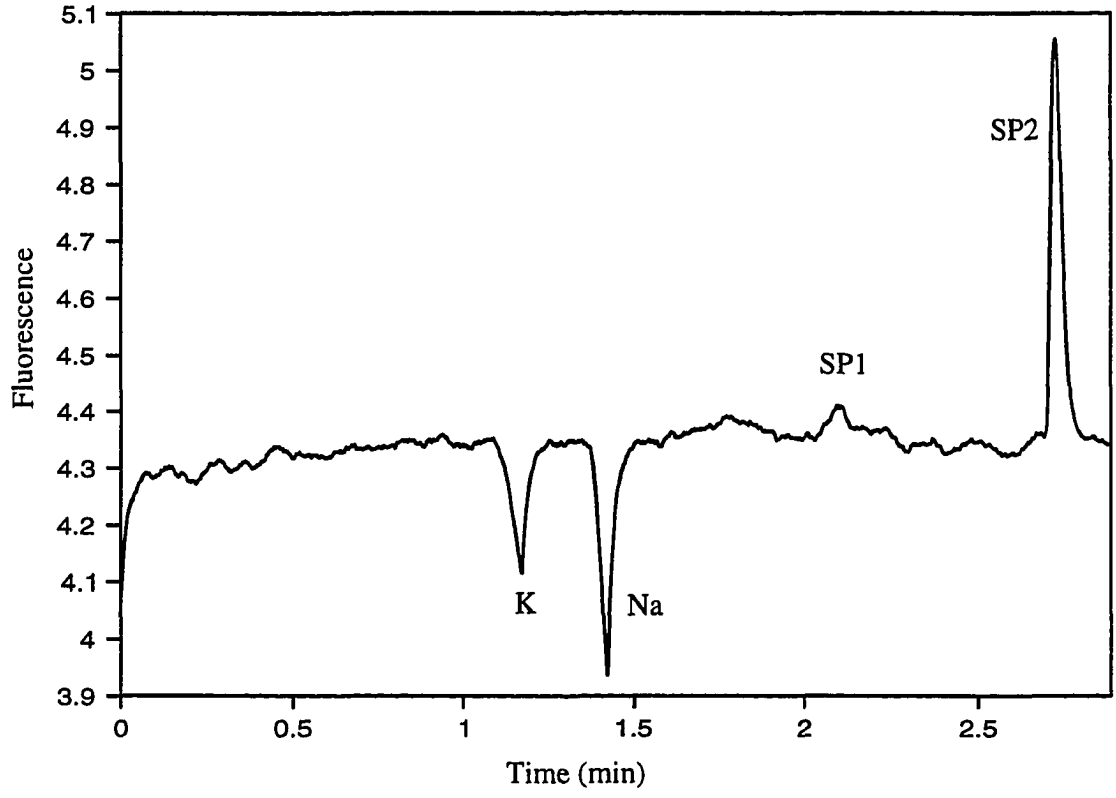


Figure 1. Electropherogram of the separation and detection of 1.06×10^{-4} M KCl and 1.02×10^{-4} M NaCl.

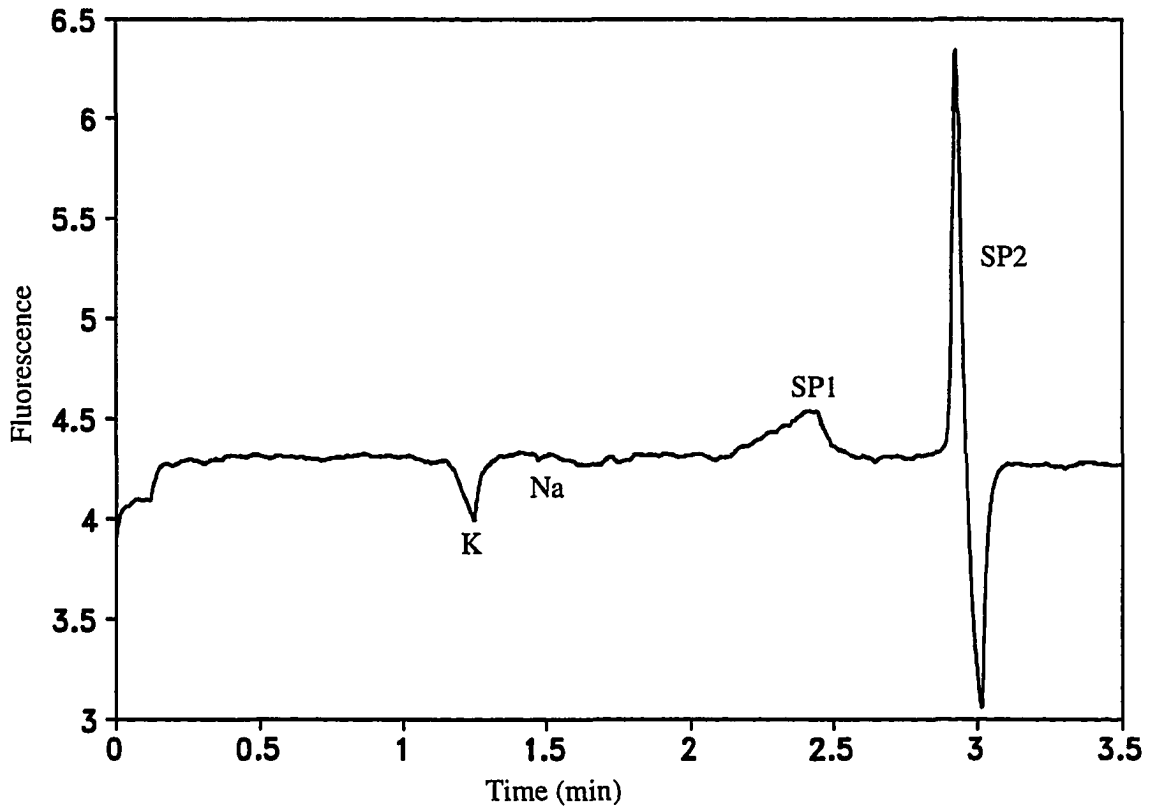


Figure 2. Electropherogram of the separation and detection of K and Na in a single erythrocyte.

different from that of the running buffer, SP2 might show a derivative shape. These phenomena suggest that SP1 is related to the amount of 2AP displaced by the analytes and SP2 is mainly due to disturbance of the equilibrium of adsorbed 2AP.

While SP2 always indicates the end of the separation, SP1 does not always lag behind the metal ion peaks. The relative positions of SP1 and a metal ion depend upon the pH of the running buffer. At pH 2.5, without any electroosmotic flow modifier in the running buffer, SP1 elutes earlier than most of the alkali and alkaline earth ions. It overlaps the peak of the fastest ion, cesium. At this pH, silanol groups on the capillary surface are almost neutral. Interaction of 2AP with the capillary surface is decreased and the mobility of 2AP increased. At lower pH, 2AP may also become doubly charged and migrates even faster. It is therefore important to select an appropriate pH range so that the system peaks do not overlap with an analyte peak.

Contamination of the running buffer In the early stages of this work, occasionally a system peak would appear at the position of the sodium ion. This system peak could be positive even when a standard solution of sodium was injected. Sometimes it appeared as a broad band or as a stepwise change in the background signal. This system peak was not as predictable as SP1 and SP2. It seriously affected the accuracy and reproducibility of sodium determination. This was later found to be related to sodium contamination in the running

buffer. In those experiments, 20-ml glass sample vials were used as buffer reservoirs. Sodium contamination came from the glass. Even when a glass container was washed very carefully as suggested in previous work (1), it was hard to completely eliminate the release of sodium from glassware during use. Polyethylene substitutes were therefore used later on and the system peak at the position of the sodium peak no longer appeared.

The cause and behavior of the system peak in these experiments were further characterized. 50 μM of KCl and 100 μM NaCl were deliberately added to the running buffer of 250 μM 2AP and 1.5% glucose at pH 4.8 that was originally free of contamination. After the running buffer was contaminated with K and Na, the electropherograms became more complicated than expected. The exact mechanism was not clear. When clean buffer was injected, two positive peaks appeared at the positions of potassium and sodium respectively (Figure 3). Since the contaminated running buffer contained more sodium than potassium, the peak at the potassium position was smaller than that at the sodium position. SP1 and SP2 appeared as negative peaks. This is because 2AP in the injected zone replaced the potassium and sodium ions that were in the buffer. The zones corresponding to potassium and sodium thus contained more 2AP molecules than in the buffer stream. SP1 and SP2 became negative because the replaced

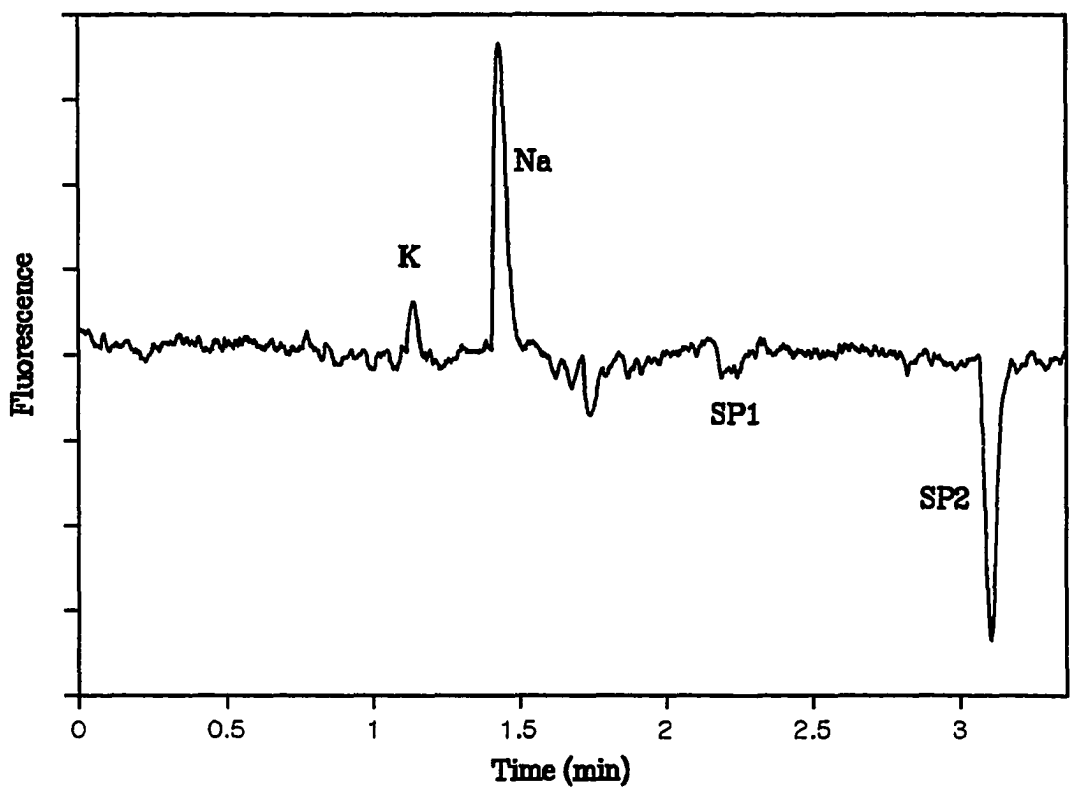


Figure 3. Electropherogram of clean buffer injected into the running buffer which was contaminated with 50 µM KCl and 100 µM NaCl.

potassium and sodium ions subsequently replaced other 2AP molecules. When only CsCl prepared in clean buffer was injected, the Cs peak had a shoulder which corresponded to a K peak (Figure 4). The Na peak appeared as negative or derivative shaped depending upon the amount of CsCl injected. This can be explained by the disturbance of the adsorbed Na on the capillary walls.

Contamination due to the injection process In the beginning, it was very difficult to obtain reproducible measurements of potassium and sodium even in standard solutions. Not only did the peak areas of K and Na deviate substantially, but the ratio of peak areas of K and Na was also not consistent. The ratio should be about 0.63 if K and Na existed in equal concentrations. It was assumed that the outside surface of the first several cm of the capillary at the injection end might have memory effects after coming in contact with high concentration salt solutions. After the capillary was treated with alkaline solution before and after each injection, about 5-10 cm of the injection end of the capillary was carefully washed with deionized water before it was put back into the running buffer. Fresh buffer and samples were changed from time to time. However, the random fluctuation of K and Na signals was still present. Moreover, peaks sometimes would appear at the positions of K and Na even when the capillary was just pulled out from and put back into the buffer vial. Originally, a micropipet tip was inserted through the cap of a buffer vial to guide

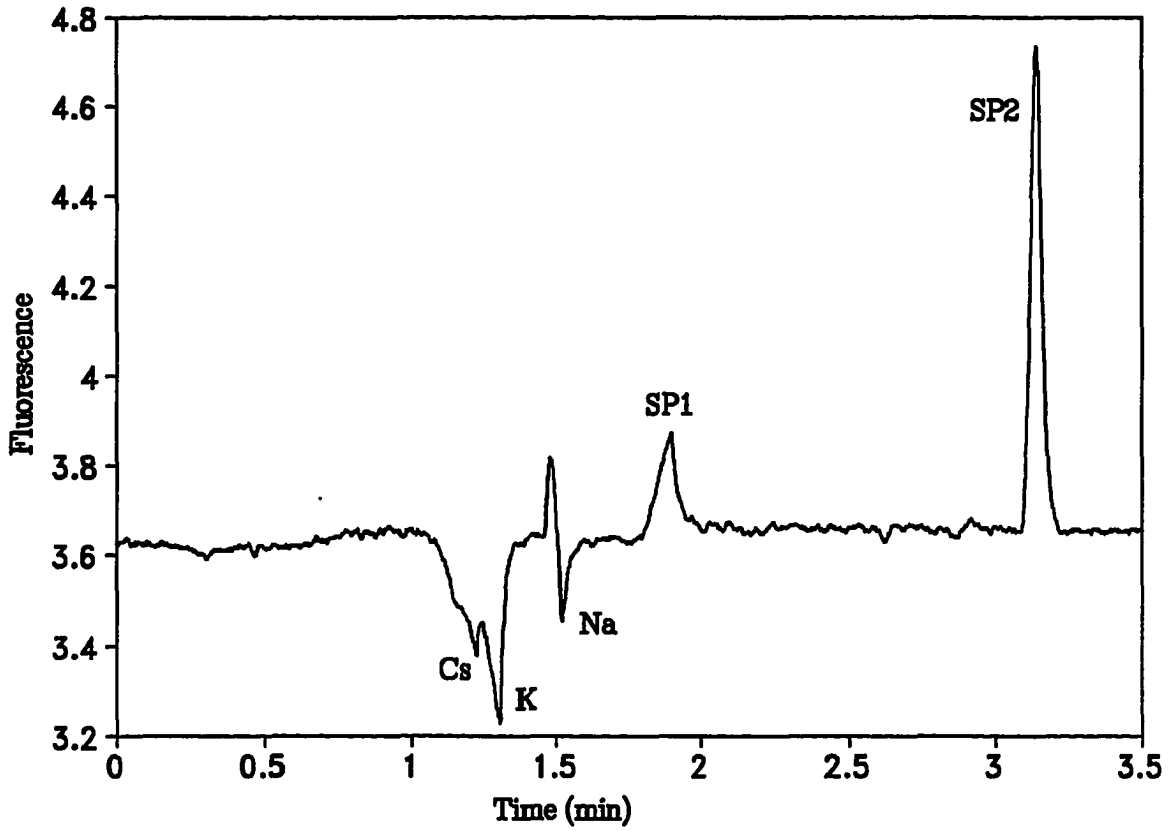


Figure 4. Electropherogram of 5 x 10^{-4} M CsCl injected into the running buffer which was contaminated with 50 µM KCl and 100 µM NaCl.

the capillary into the buffer vial. It was possible that the micropipet tip was not clean. But even when new micropipet tips were used, the problem was still not solved. One important observation is that the impurity peaks did not appear in every experiment. If the detection system was not optimized and the detection limit was poor, no impurity peaks were detected. In other words, impurities were only present in very small amounts.

In contrast to sodium and potassium, the reproducibility of determination of lithium was much better. Calibration curves were obtained with a linear range of three orders and R^2 of 0.999. Obviously, there was no lithium contamination in the environment. The interesting thing is that when only LiCl standard was injected, the impurity peaks at the positions of potassium and sodium still appeared. Analysis of these impurity peaks with lithium as the internal standard showed that these corresponded to concentrations of no more than 1×10^{-5} M (Figure 5). 90% of these impurity peaks corresponded to an amount of no more than 60 fmol of cations when a 75-μm capillary was used. In about 50% of injections of lithium standards, there were impurity peaks at the position of either potassium or sodium or both. The results suggest that the impurity peaks must be caused by contamination in the surrounding environment.

To identify the exact cause of contamination, the capillary injection end was deliberately brought into contact with different surfaces and then put back

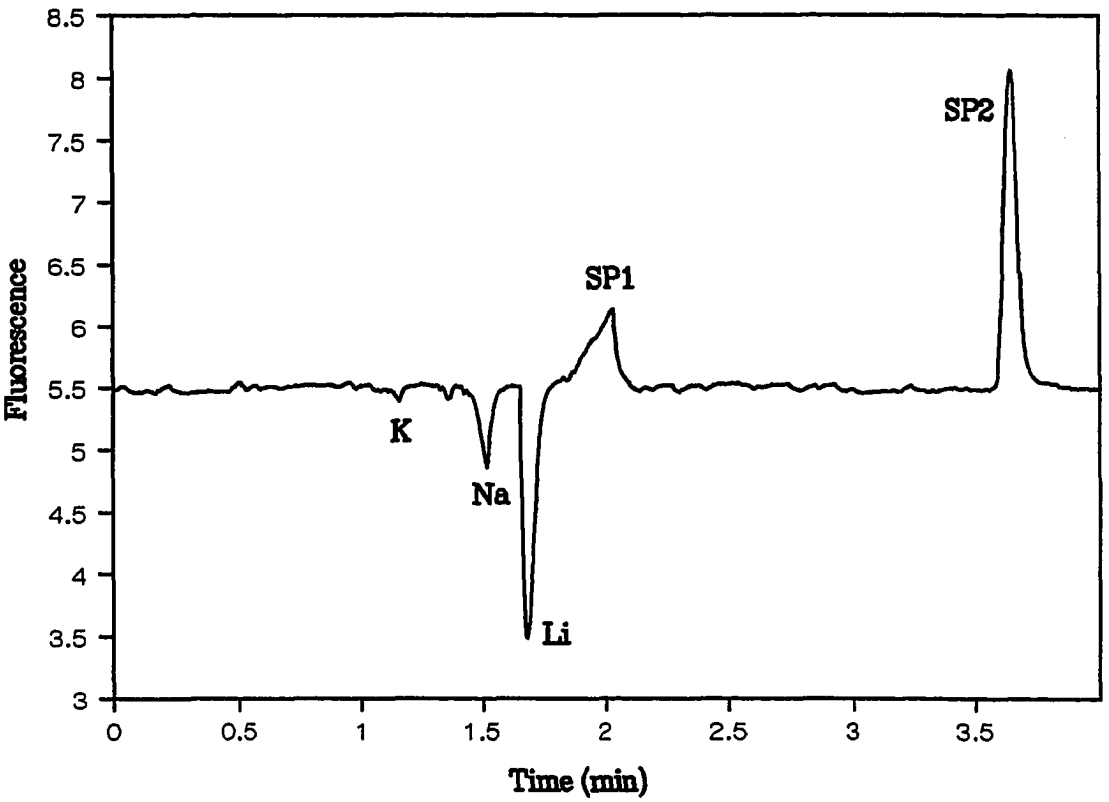


Figure 5. Contamination from particles when LiCl was injected. The K and Na contamination did not affect the calibration of LiCl.

into the buffer vial to start electrophoresis without sample introduction. As shown in Figure 6, the touching process indeed introduced significant amounts of impurities from the contaminated surfaces.

The results indicate that Na and K are the two major soluble cationic contaminants. This is consistent with the typical elemental compositions of particulate matter, which contain about 1-2% of K, 2-4% of Na, and 8.3% of Ca, as estimated from the data in the literature (4a-4d). The major component of an airborne particle is silica. Soluble inorganic salts usually form a layer of coating around the particle. If a 2 μm diameter particle with average elemental composition was injected into a capillary, it would release about 1-2 fmol of K and Na. The amount is high enough to interfere with single-cell analysis. It is also possible that multiple particles were injected together.

There are millions of particles with diameters greater than 0.01 μm in every ml of aged air. In the size range of 0.5 μm to 10 μm diameter, there are thousands of particles per ml of aged air (10). Since such a large number of particles exist in ambient air, they can deposit quickly on any open surface. The particle deposition process can be more pronounced in an area with high static electricity, such as the CE system. Therefore, there is substantial opportunity for the capillary to pick up some particles when contacting any surface which had been exposed in the air for some time.

To avoid contamination from particles, the capillary injection end should

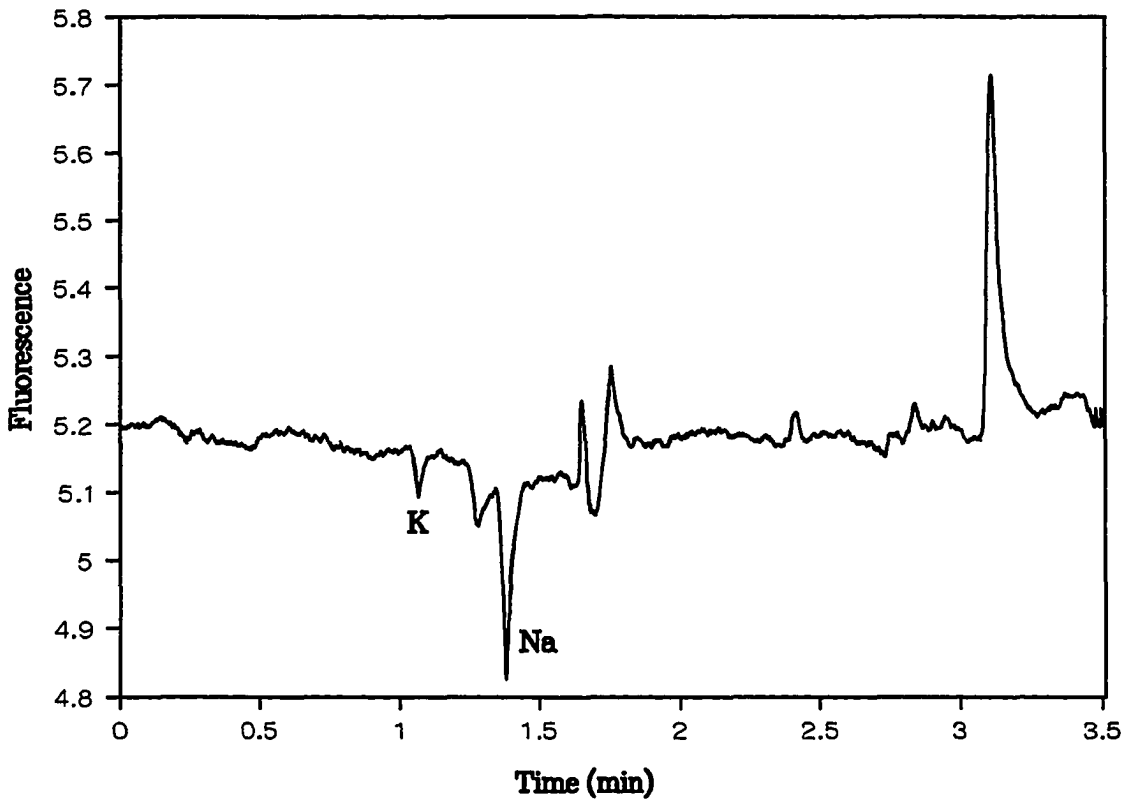


Figure 6. Contamination from the inner surface of a new, unwashed 20-ml glass sample vial.

avoid contact with any surface. The micropipet tip which was used to guide the capillary into the buffer vial was the major source of contamination since it could easily collect airborne particles. Therefore, the buffer vial was changed to a wide-mouth polyethylene bottle so that a capillary could be taken out and put in freely without touching the mouth of the buffer vial.

We further investigated if it is possible to carry out single-cell analysis for intracellular K and Na in the open air without additional shielding. It was found that glass slides were not suitable for the cell injection operation because they released sodium salts even after being washed with hydrochloric acid and deionized water several times. To verify this, a droplet of deionized water was put onto a clean glass slide. If the water sample was injected from the glass slide, contamination peaks appeared. Plastic slides were thus used. Before use, the plastic slide was rinsed in deionized water to remove any particles on its surface. The water beads were then shaken off. To test the benefits of using plastic slides, a droplet of deionized water was put onto a washed plastic slide. The capillary injection end was then immersed in the water droplet for about 30 s to inject the water sample by gravity. Such injected blank runs were found to be contamination free. However, if the water droplet was placed on the slide for over 10 min, impurity peaks would appear when the water sample was subsequently injected. This indicated that the water droplet on the slide collected airborne particles during long exposures to the open air. Therefore,

each slide should be used for only one single-cell injection. Great care was also taken when transferring a capillary between the microscope stage and the buffer vial so that the injection end of the capillary did not touch the micromanipulator, microscope stage and the outer surface of the buffer vial. Therefore, it is feasible to do single-cell analysis of intracellular K and Na in an open air environment. However, specific contamination control devices such as a clean room or a proper ventilation system are desirable if the procedure is to be implemented for routine operation.

Single-cell analysis In single-cell studies, an adequate amount of whole blood was washed and centrifuged several times to get rid of the extracellular matrix. Since cell leakage, which was partly addressed in previous work (2), is more significant for small inorganic ions than for organic ions and macromolecules, only fresh blood samples should be used and manipulation should be rapid in the analysis of intracellular K and Na. In this study, it was found experimentally that multiple washing cycles and prolonged suspension of cells in solutions with low ionic strength decreased the signals of intracellular K and Na.

We then analyzed about 70 cells over a time period of about nine hours. In this experiment, 20 µl fresh blood was collected and washed with 14 ml wash buffer for four cycles. Washed cells were then stored in the suspension solution at a dilution factor of 500 at room temperature. A small amount of this cell

suspension solution was transferred for single-cell injection every two to three hours. The first cell injection began 35 min after the blood sample was collected. We can discern that there was a gradual decrease of the signals over the long time period. This tendency is qualitatively consistent with the results of a leakage study performed by using UV absorption detection in which lysed cells were analyzed instead of single cells (data not shown). The decreasing tendency was not due to the deterioration of the performance of the capillary or the detector because the signal response for standards injected throughout the time period did not show the same trend. The average intracellular K amount was calculated to be 4.4 fmol. K was not detected in 18% of the cells. The results indicate that cells lost their intracellular K and Na because of improper handling and storage.

To minimize leakage, a different cell preparation procedure was used in another experiment. In this experiment, 2 μl of fresh blood was collected and washed only once in 14 ml wash buffer. The washed cells were then diluted to 3 ml for single-cell injection. The first cell injection was begun about 4 min after the blood sample was collected. After one hour, the old cell suspension solution was centrifuged. The supernatant solution was injected to check if there was contamination. The old cell sample was then discarded. Another fresh 2-μl blood sample was collected for the next one-hour experiment. Five batches of cells were collected and 49 cells were analyzed. About 10 cells were analyzed in

every batch (every hour). Before each batch of cells were analyzed, 10^{-4} M and 10^{-5} M standards were injected three times respectively. The RSD of peak areas for the injected standards was 10% to 15%. Average peak areas were calculated. A two-point calibration was then used to calculate the amount of K and Na present in each cell.

As shown in Figure 7, the separation and detection system was very reproducible. Excellent signal-to-noise ratios were obtained for K measurement. K was detected in all of the cells and Na was detected in 88% of the cells. The results are summarized in Figure 8. The results show that contamination was well under control. The leakage problem was minimized by using fresh cells and reducing washing cycles. The average intracellular K and Na were 10.8 fmol and 0.33 fmol respectively. The observed standard deviation of the K and Na amounts were 6.3 fmol and 0.31 fmol respectively. The average Na:K ratio was 0.035. The RSD of Na:K ratios was 0.032. The ratios of Na to K in the cells were lower than the macroscopic value, as were the absolute amounts of Na. This is probably because the S/N ratio for the Na peak was low, affecting the determination of peak areas.

Recently (11,12), we have determined the distribution of intracellular K and Na in individual human erythrocytes by laser vaporization/atomic emission in a flow stream. There we found similarly large variations of K and Na in the population of cells studied. The fact that these two independent techniques

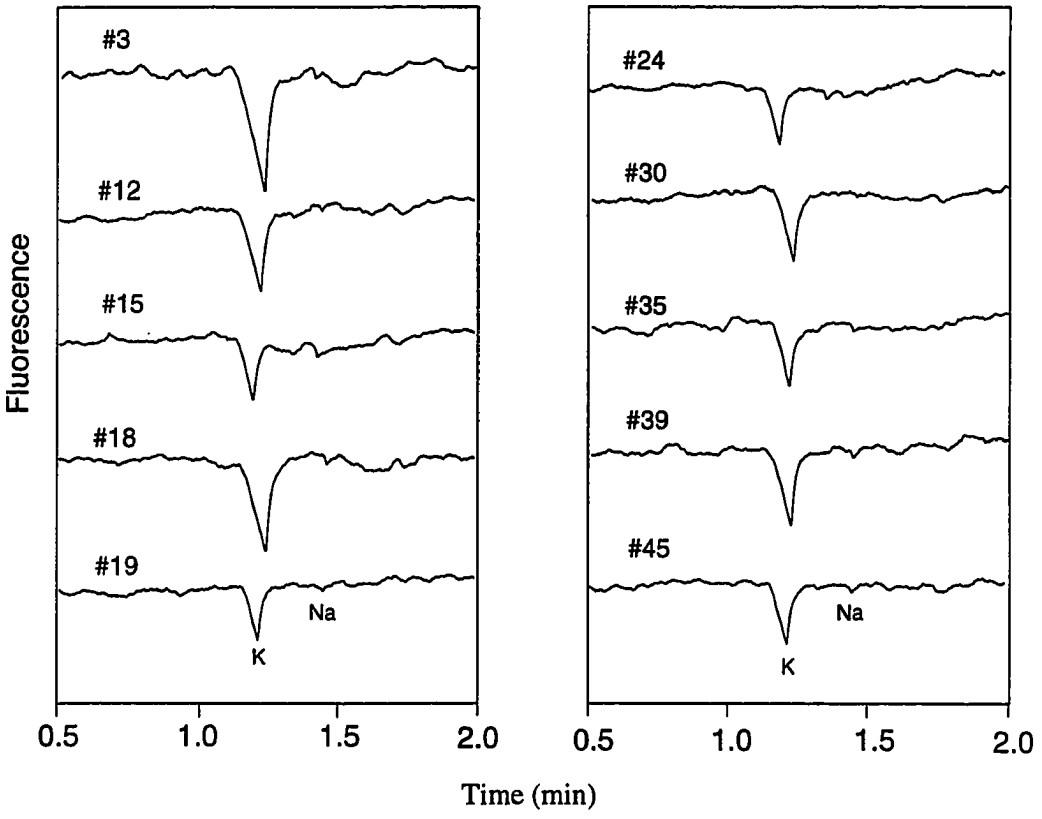


Figure 7. Selected electropherograms of single-cell analysis from an experiment involving 49 cells.

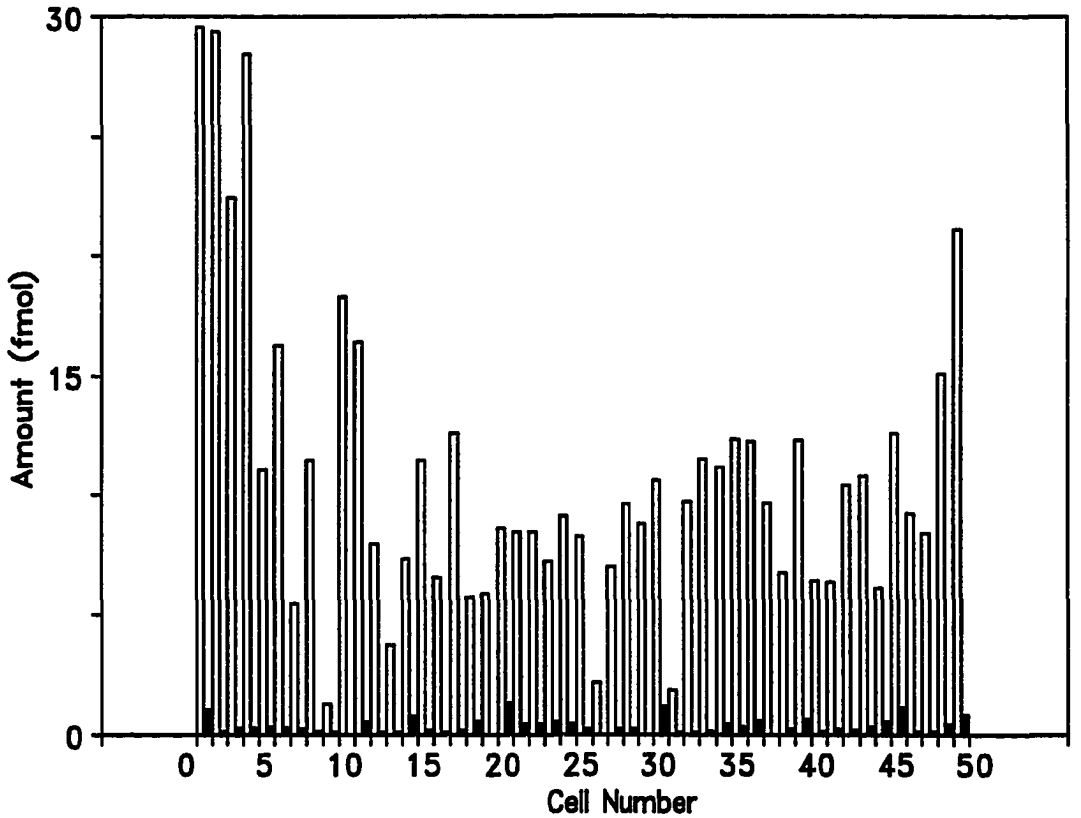


Figure 8. Distribution of the amounts of potassium and sodium in 49 single cells from five batches of blood samples. Solid bars, Na; and open bars, K. See text for details.

produced essentially the same results indicate that the large variations are not due to experimental artifacts in either technique. The conclusion is that cell age is responsible for these variations. The major regulation mechanism of intracellular K and Na is activity of the enzyme Na^+/K^+-ATPase (13), with minor influences due to other ions such as Cl^- (14), Ca^{2+} (15) and Mg^{2+} (16). As the enzyme activities degrade as a function of cell age (17), one expects the intracellular K levels to decrease. So, Figure 7 depicts cells ranging from 0-120 days old in the blood sample.

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 Basic Energy Sciences, Division of Chemical Sciences.

REFERENCES

1. L. Gross and E. S. Yeung, "Indirect Fluorescence Detection of Cations in Capillary Zone Electrophoresis", Anal. Chem., 62 (1990) 427.
2. B. L. Hogan and E. S. Yeung, "Determination of Intracellular Species at the Level of a Single Erythrocyte via Capillary Electrophoresis with Direct and Indirect Fluorescence Detection", Anal. Chem., 64 (1992) 2841.

3. Q. Xue and E. S. Yeung, "Indirect Fluorescence Determination of Lactate and Pyruvate in Single Erythrocytes by Capillary Electrophoresis", J. Chrom., 661 (1994) 287.

4. (a) M. Iwatsukl, S. P. Tilleberatine, T. Fukasawa and T. Fukasawas, "X-Ray Analysis of Airborne Particulates Collected by an Andersen Sampler. Compound and Elemental Distribution vs. Particle Size of Laboratory Particles", Environ. Sci. Technol., 18 (1984) 818; (b) S. E. Lindberg, M. Bredemeier, D. A. Schaefer and L. Qi, "Atmospheric Concentration and Deposition of Nitrogen and Major Ions in Conifer Forests in the United States and Federal Republic of Germany", Atmospheric Environment, 24A (1990) 2207; (c) W. Baeyens and H. Debeurwaerder, "Particle Trace Metals Above the Southern Bight of the North Sea—I. Analytical Procedures and Average Aerosol Concentrations", Atmospheric Environment, 55A (1991) 293; (d) G. A. Norton, E. L. Dekalb and K. L. Malady, "Elemental Composition of Suspended Particulate Matter from the Combustion of Coal and Coal/Refuse Mixtures", Environ. Sci. Technol., 20 (1986) 604.

5. D. Querido, "In Vitro Loss of Potassium from Erythrocytes During the 0-180h Postmortem Period in Rats: Relationship Between Potassium Loss and Postmortem Interval", Forensic Science International, 51 (1991) 111.

6. S. Hjertén, "Zone Broadening in Electrophoresis with Special Reference to High-Performance Electrophoresis in Capillaries: An Interplay Between Theory and Practice", Electrophoresis, 11 (1990) 665.

7. S. R. Meech and D. Phillips, "Photophysics of Some Common Fluorescence Standards", J. Photochemistry, 23 (1983) 193.

8. D. R. Lide, in Handbook of Chemistry and Physics; CRC Press: Boston, 1990.

9. E. S. Yeung and W. G. Kuhr, "Indirect Detection Methods for Capillary Separations", Anal. Chem., 63 (1991) 275A.

10. A. Lieberman, in Contamination Control and Cleanrooms; Van Nostrand Reinhold: New York, 1992, p. 68.

11. N.-H. Cheung and E. S. Yeung, "Single-Shot Elemental Analysis of Liquids Based on Laser Vaporization at Fluences Below Breakdown", Appl. Spectrsc., 47 (1993) 882.

12. N.-H. Cheung and E. S. Yeung, "Variation in Sodium and Potassium Contents Within Individual Human Erythrocytes", Anal. Chem., 66 (1994) 929.

13. L. E. Hokin, in Membrane Bioenergetics (C. P. Lee, G. Shatz and L. Ernster, eds.), Addison Wesley: Reading, Massachusetts, 1979, p. 281.

14. P. B. Dunham, G. W. Stewart and J. C. Ellory, "Chloride-Activated Passive Potassium Transport in Human Erythrocytes", Proc. Natl. Acad. Sci. (USA), 77 (1980) 1711.

15. G. Gardos, "The Function of Calcium in the Potassium Permeability of Human Erythrocytes", Biochem. Biophys. Acta, 30 (1958) 653.

16. P. W. Flatman and V. L. Lew, "Motor Unit Organization of Human Medial Gastrocnemius", J. Physiol. (London), 287 (1979) 33.

17. P. A. Marks, A. B. Johnson and E. Hirscherg, "Effect of Age on the Enzyme Activity in Erythrocytes", Proc. Natl. Acad. Sci. (USA), 44 (1958) 529.

CHAPTER 2. ADAPTING THE CHARGE INJECTION DEVICE FOR APPLICATION IN CAPILLARY ELECTROPHORESIS

INTRODUCTION

In the last 20 years, charge transfer devices (CTDs), which include charge coupled devices (CCDs) and charge injection devices (CIDs), have evolved to the most common two-dimentional array detector available today for spectroscopic applications. Although the first CCD device (1) was developed two years earlier than the first CID (2), both the first commercially available CCD and the first CID were introduced in 1973. The first scientific uses of the CCD and the CID occurred in 1976 in the area of astronomy (3,4). The low light level detection capabilities of the CTDs demonstrated in astronomical observations were soon utilized in other areas such as microscopy (5-10), molecular luminescence (11-15), molecular absorbance (16-18), Raman (19), atomic spectroscopy (20-23), and high-energy photons and particles (24-25). These array detectors also have been utilized in chromatographic applications (26-39).

Both the CCD and the CID are two-dimensional solid-state array detectors that receive incoming photons, collect and store photon-generated charge, and then determine the quantity of accumulated charge which is quantitatively related to the signal. They have many similarities, but they also have very different performance characteristics and modes of operation.

In the CCD, the charges accumulated in high-capacitance sensing elements (pixels) are transferred sequentially through adjacent pixels to a specialized low-capacitance output node that is located at the corner of the detector array. Once the readout process is initiated, all the charges in the array must be read out. Individual pixels or separated small groups of pixels cannot be selectively read by skipping other void pixel information. The charges in the pixels are eliminated immediately after the readout process.

Contrastingly, pixels in the CID are individually addressed. A pixel can be randomly accessed. Generally, it is the pixel accessing and readout process that distinguishes the CID from the CCD. The major performance features of the CID are summarized in Table 1.

The unique features of the CID determine that it is not just an alternative of the CCD. In some applications such as atomic spectroscopy, astronomy, and military surveillance, the CID provides imaging advantages that the CCD cannot. The major limitation of the CID, however, is its read noise. It cannot be determined generally whether the CID or the CCD is better. The choice must be made according to the specific application and the major imaging problems to be solved in the application. The comparison of the two devices are summarized in Table 2.

The random pixel accessing capability and the large dynamic range of the CID suggest that this device has high potential of application as an array

Table 1. The major performance features of the CID

	Features	Advantages/Disadvantages
1.	Random pixel accessing	Reduce data volume Increase frame rate
2.	Charge confined within a pixel	No charge transfer loss effects
3.	Low dark current	Long integration time Operation at room temperature
4.	Nondestructive readout	Reduce read noise Dynamic image acquisition control
5.	Defects confined to each pixel	Defect tolerance
6.	Optical overloads confined to illuminated pixels	Anti-blooming
7.	Contiguous pixel structure (no opaque regions)	Virtually all of the sensor area produces useful photocurrent
8.	Simple 3-mask process for array fabrication	Inherently high quantum yield
9.	Large pixel well capacitance	Large dynamic range Contribute to read noise

Table 2. Comparison of the CID and the CCD

	CID	CCD
Structure simplicity	Simple	Complicated
Defect tolerance	High	Low
Pixel size	Comparable	Comparable
Quantum yield	High	Low unless back-thinned
Spectral response range of one device	Wide	Narrow
Read Noise	High	Low
Dark current	Low	High
Anti-blooming capability	High	Low
Spatial resolution	Comparable	Comparable
Charge transfer loss effects	None	Yes
Dynamic linear range	Large	Narrow
Sensor metric stability	Comparable	Comparable
Operation flexibility	High	Low

detector in on-column laser-induced fluorescence detection in multiplexed capillary electrophoresis (40). This is the first time that the CID is used for capillary electrophoresis application, although that of the CCD was five years ago (29).

This chapter documents the process of adapting the CID for laser-induced fluorescence detection in capillary electrophoresis, mainly from the hardware and software development aspects. More details are discussed elsewhere (40).

EQUIPMENT

The CID used in this work is the SCM5000E scientific grade β-system provided by CID Technologies Inc. (Liverpool, NY). The system carries a CID38 imager. The CID38 imager contains up to 524H x 524V active pixels. According to the manufacturer, 512H x 512V pixels can be used for image acquisition. Along with the hardware system, the DOS version function library and the macro development kit are provided but without complete customized application software. The PC computer used for this CID system is a 486 DX/33MHz computer with 16MB RAM, 250MB hard drive and a super VGA monitor (Electra, Ames, IA). Application software is developed by using Microsoft QC2.5 as the compiler.

HARDWARE

Imager architecture In the CID imager, pixels can be addressed individually in the X-Y addressable pixel array (Fig. 1). Each pixel consists of two metal-oxide-silicon capacitors. One capacitor is designated as the column charge storage pad and the other as the row charge storage pad. The column pads of all pixels in one column are connected to a column bus, and the row pads of all pixels in one row are connected to a row bus. The column buses are connected to the column shift register (H-scanner) and the row buses are connected to the row shift register (V-scanner). Pixel selection is accomplished by selecting the proper column and row through the H-scanner and V-scanner. Charge collected in a pixel is manipulated independently by driving the column bus and the row bus to execute various image functions such as readout, injection, or knock-down. This random pixel access is the first unique feature that distinguishes the CID from the CCD.

Fig. 2 shows the complete process of integration, readout and injection. The column and row reference voltages labeled in the figure are the actual voltages used in the SCM5000E CID system. At integration mode (Fig. 2a), the column pad is set at a lower voltage (Col Ref_L) than the row pad (Row Ref_L). Since the imager surface is N-doped semiconductor, photon-generated or thermally-generated charge is collected in the lower potential well under the column pad. The charge can be measured without removing it from the pixel.

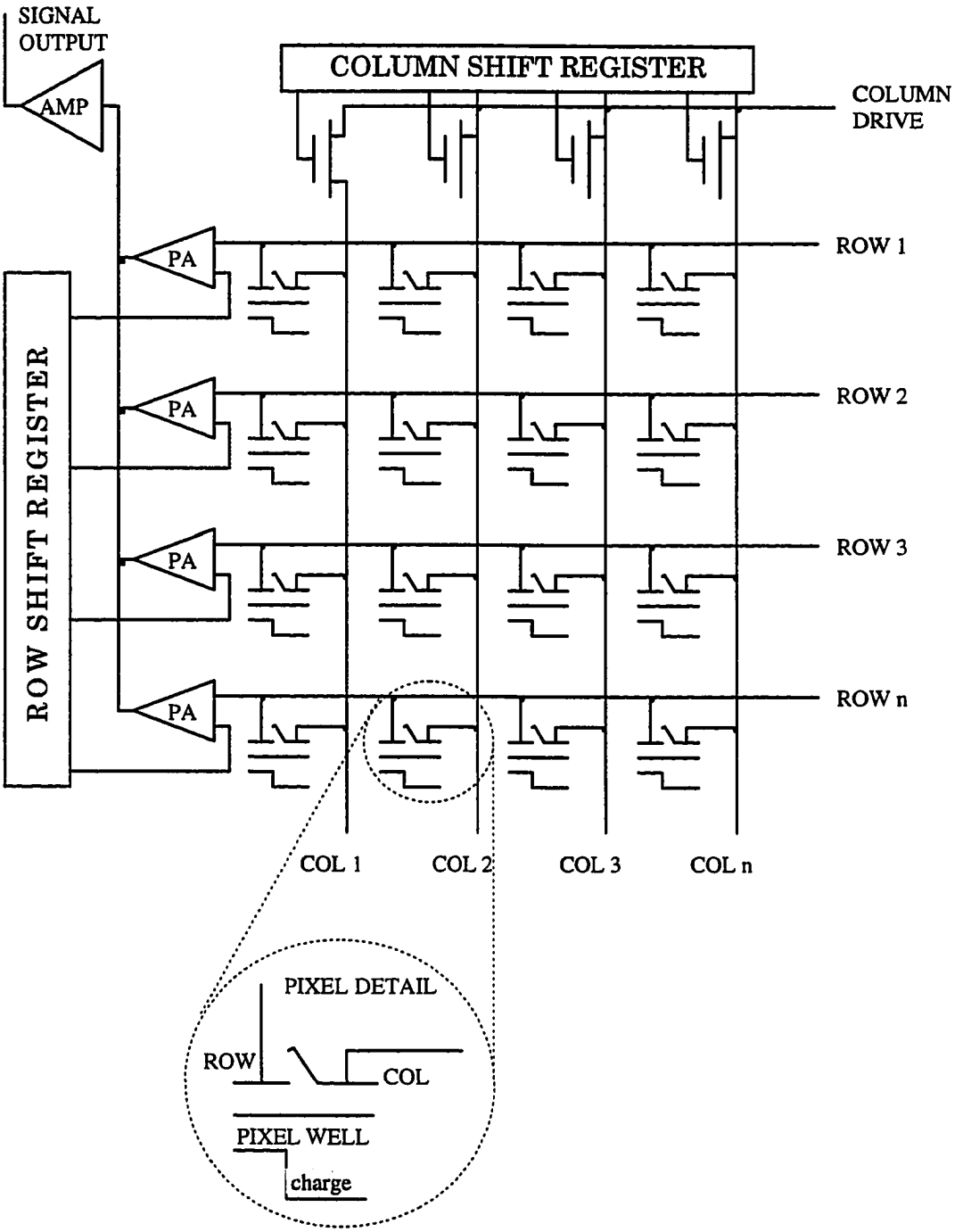


Figure 1 Random pixel addressing in the CID imager architecture

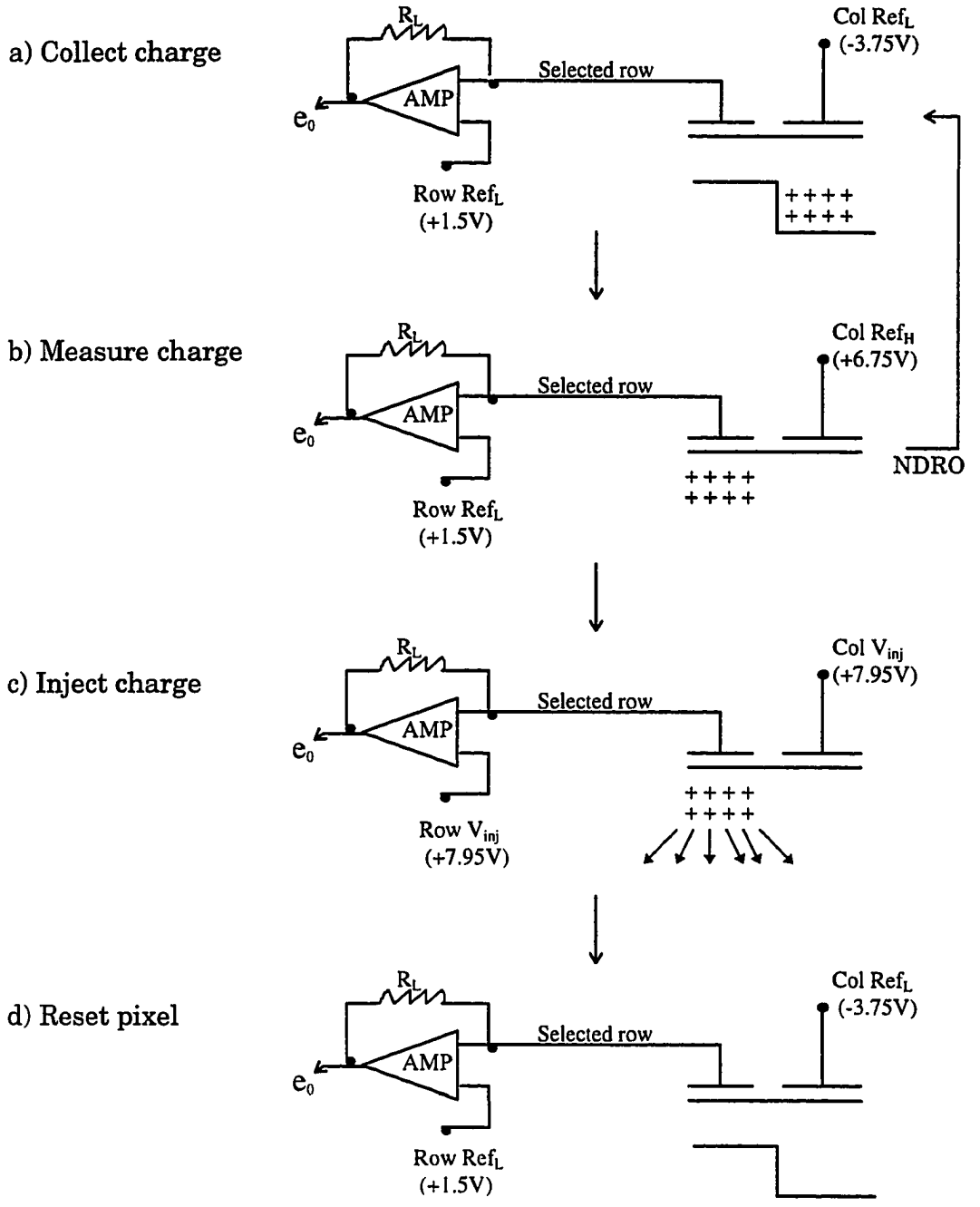


Figure 2 Schematic diagram of readout sequence of a pixel in the CID

At readout mode (Fig. 2b), the voltage of the column pad is raised to Col Ref_H, where Col Ref_H > Row Ref_L. The charge is thus driven to the potential well under the row pad. This charge transfer within the pixel causes a current flow in the external circuit and is sensed by the pre-amplifier for this row. After the readout, the charge remains in the pixel. The charge can be returned to the column capacitor by resetting the voltage of the column pad to Col Ref_H. The charge can thus be measured nondestructively many times by repeating the above two steps. This nondestructive readout (NDRO) operation is the second unique feature that distinguishes the CID from the CCD.

After the charge in the pixel is read one or multiple times, the charge can be injected into the imager substrate by raising both potentials of the column and row pads to charge injection voltage V_{inj} (Fig. 2c). The pixel is then reset to integration mode (Fig. 2d). Unlike in the CCD where charges in pixels are eliminated automatically following readout, charges in CID pixels need to be cleared in a separate step after readout.

The complete and operable CID system consists of three major parts: the CID camera head, the camera control unit, and the host computer as shown in Fig. 3.

Camera head The camera head unit contains the shutter assembly, the imager, the preamplifier board, and the liquid nitrogen dewar chamber. The imager is hosted within the liquid nitrogen dewar chamber. The preamplifier

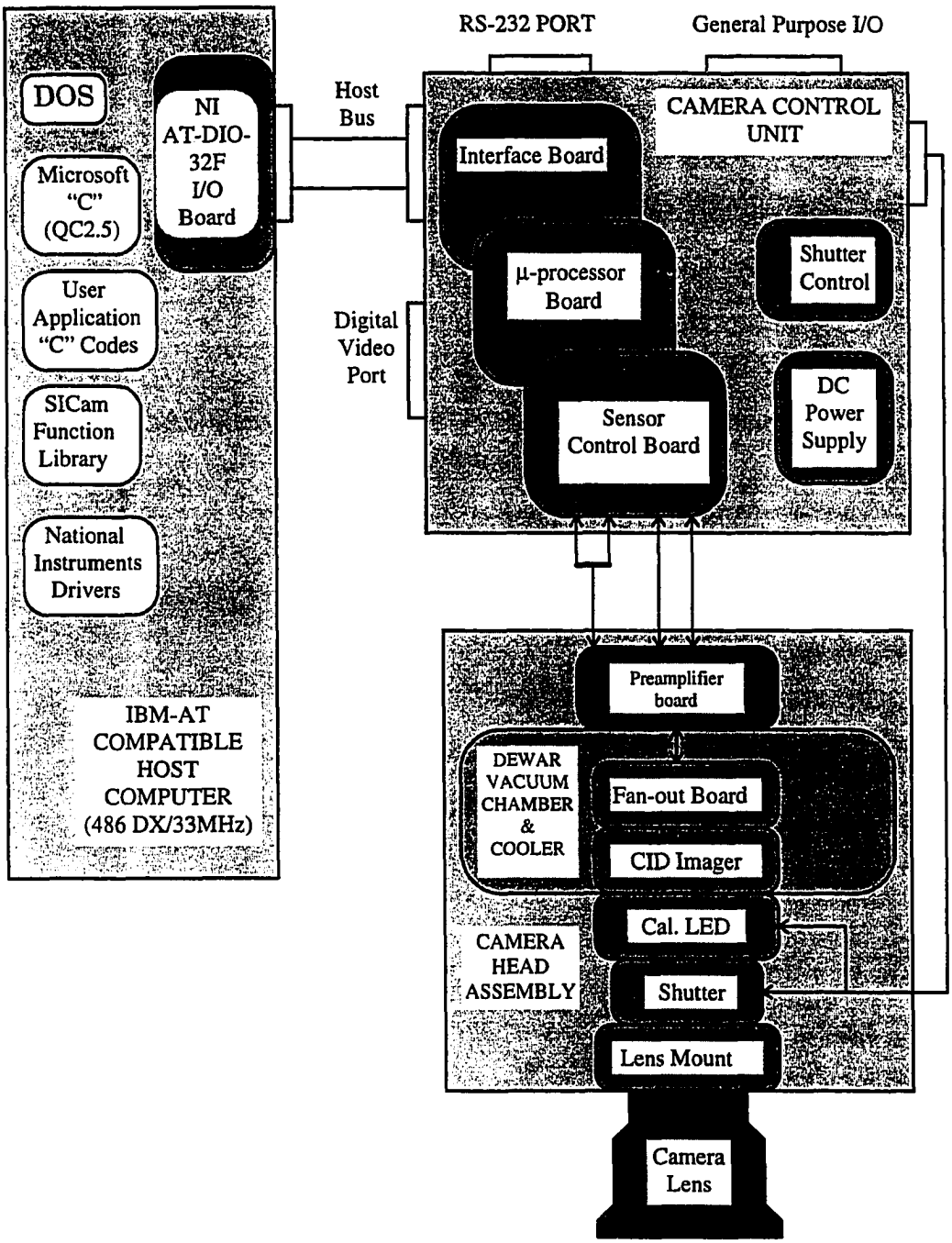


Figure 3 SCM5000E system block diagram

board is located at the base of the dewar. If the CID needs to operate at liquid nitrogen temperature, the dewar chamber is evacuated down to 10^{-5} to 10^{-6} Torr with a diffusion pump, or preferably, an oil-free ion pump. The dewar is then filled with liquid nitrogen, which usually takes 30 min. A full dewar of liquid nitrogen can last for 12 to 16 hrs, depending upon the vacuum inside the dewar chamber and the ambient temperature. The time the vacuum can be maintained depends upon how the sealing of the dewar and how the cleaness inside of the dewar. Typically, the vacuum should be preserved at least one week without causing dramatic evaporation of liquid nitrogen or even frosting on the CID camera head when the dewar is filled. The CID can also be operated at room temperature without causing any harm to the device.

Camera control unit The camera control unit houses all the control electronics for the camera system. The major parts include the processor board, the sensor control board, the interface board, the shutter control board, and the DC power supply.

The heart of the processor board is an 80L286 10-MHz microprocessor. On the board, a RAM socket and an EPROM socket can accept from 64K (8K x 8) to 1MB (128K x 8) RAM and from 512K (64K x 8) to 1MB (128K x 8) EPROM. These two memory devices are supported for program and data storage. The 286 microprocessor can continuously accept a stream of commands from the host computer, translate each one into primitive codes, and send the primitive codes

to the logical cell array on the sensor board to execute. Alternatively, a macro consisting of a series of operations can be downloaded from the host computer into the memory of the 286 microprocessor at one time and run locally at the 286 microprocessor. This process eliminates the need for the host computer to send the commands one by one. For time critical applications, this process is efficient. In effect, the functions in the standard library are written in this way. Because each function only carries out a very specific task, a user-defined operation usually calls multiple functions to accomplish it. But if a user-defined operation can be completed in a single macro, the code of this operation can be pre-downloaded to and reside in the 286 microprocessor. After a triggering word is sent from the host computer to the 286 microcomputer, the operation is executed without any interruption by the communication between the host computer and 286 microcomputer.

The heart of the sensor control board is the programmable logical cell array (LCA). The LCA receives the primitive codes from the processor board and executes them. The major function of the LCA is to generate the clocking waveforms to operate the CID sensor in various modes necessary to scan the imager. The LCA drives the vertical and horizontal scanners to slew to an address, conducts pixel reading, and performs other successive or auxiliary operations necessary to complete the image scanning process. It also provides a time base for its operation.

The interface board provides monitoring of system interface signals and power supplies, interface signal termination, and signal conversation for specific drive signals.

Signals scanned from the imager are carried in certain waveforms. These waveforms flow through the preamplifier circuit, go to the analog processor and 14-bit A/D converter in the sensor control board, and are finally digitized. There are three ways to access the digital signals. They can be monitored from a digital video I/O port in the camera control unit, be transferred to the 286 microprocessor, or go directly from the 14-bit A/D converter through the pixel data FIFO to the host computer memory. The last mode is used to collect data from the camera to the host computer in this work.

Host computer The host computer should be a PC/AT compatible computer of a model higher than 386/20MHz with a VGA monitor. An AT-DIO-32F interface board is plugged into a 16-bit slot on in the host computer for communication between the host computer and the camera. DOS-compatible "C" software is used to control the CID operations.

OPERATION

Initialization This process initializes the communication between the host computer and the camera system, and sets the CID operation parameters including the active imager area, X and Y direction skew pixel, column and row

reference voltages, even and odd row offset voltages, and the amplifier offset voltages. The default values of these parameters are preset in the RAM based 286 code by the manufacturer. Although these values can be modified by the user in application software, the user usually just needs to adjust the active imager area as necessary. The user sometimes also needs to adjust the charge injection voltage for more complete charge clearance. Others can be used as preset by the manufacturer. If the user wishes, the user can re-optimize these operation parameters according to a specific application.

Calibration The analog focal plane electronic circuits of CID camera require calibration with the camera head assembly at a stable operation temperature. After the camera is turned on or reset, it should be calibrated before it is used to acquire images. If the camera head is to be cooled down to liquid nitrogen temperature, after the dewar chamber is filled with liquid nitrogen, it usually needs to wait for one to two hours before calibration and operation. The calibration process is to set the coarse amplifier offset voltage, the pre-amplifier offset voltages and the row offset voltages. The pass status of the calibration is also recorded. The calibration data are then downloaded to the 286 microprocessor. The data can also be stored on hard disk so that the next program can read and download the data to the 286 microprocessor without the need of calibrating the camera again.

Image acquisition After an exposure, the charges are stored in the array of individual pixels. There are flexible ways to read and manipulate these pixel charges. The major flexibility includes:

- Random pixel access: One can selectively read only those interesting areas (subarrays). A subarray can be as small as one pixel.
- Nondestructive readout (NDRO): The charge injection function can be suspended so that a pixel can be read repeatedly without changing the charge accumulated in the pixel. NDRO is primarily for reducing read noise because read noise is in reciprocal proportion to the square root of the number of NDRO. Secondly, NDRO is useful in dynamic control of image acquisition, especially in atomic spectroscopy applications.
- Subarray injection: Charges in a subarray can be cleared selectively without affecting other subarrays.
- Dynamic change of gain: according to signal strength, background level, and digitization sensitivity requirement, different gain can be selected for each subarray according to the amount of charge accumulated in the subarray. This is useful in extending the dynamic range of CID.

Data storage Data flowing out of the CID controller can be accepted by the host computer into either base memory or extended memory (XMS). Because pixel intensity is expressed as integer and a maximum contiguous data segment

that can be defined in base memory is 64KB, base memory can only accept an subarray image with up to 32K pixel data. Any subarray image larger than 32K pixel data must be read into extended memory.

The maximum accessible XMS in Microsoft DOS environment is 16MB. Larger XMS can be reached through other environments such as Microsoft Windows. At the time when the SCM5000E CID system was purchased, only MS-DOS based software interface was available. Although Windows based software can manage larger XMS for data storage, DOS based software usually runs more efficiently and faster.

In time critical applications, it is desirable to store multiple images in XMS without accessing the hard disk during data acquisition because data transfer from memory to hard disk is much slower than that in between memory blocks. Since a full CID frame contains 262K pixels (512x512), as many as 30 full frames can be stored in XMS if each frame is read only once. If smaller subarrays are read, more frames can be stored until 16KB XMS is full.

For applications in CE, it is common that thousands of frames need to be stored in XMS. In such a case, it is not practical to allocate one XMS handle for each frame because of limited number of XMS handles available in the DOS environment. To overcome this limitation, the solution is to allocate one XMS handle for each subarray. A series of frames for this subarray are stored into one single XMS block allocated by the XMS handle designated to this subarray.

Any calculation for image data can only be performed after the data to calculate are transferred into base memory. When multiple NDROs are used in subarray readout, if the image size does not exceed 64KB, the image can be read into base memory for averaging the NDRO. The averaged image is then transferred to XMS for storage. This process reduces the data volume.

Data can be processed after they are permanently stored on the hard disk, or while they are still in XMS with data processing functions integrated in the CID control software. The latter procedure takes advantage of fast data access in memory. When a large amount of data need to be manipulated, the later procedure reduces data processing time.

SOFTWARE

Development tools The SCM5000E CID camera system is still under development. Its software system was not totally commercialized at the time of our purchase. Although the hardware and software manuals are supplied by the manufacturer, the instruction of operation is very concise. For a specific application purpose, efforts are still needed to develop proper software to run and evaluate the device for the applications.

The operation platform for the CID system is PC microcomputer. Speed and data storage capacity discussed in the following text is specific for a desktop computer. Because a desktop microcomputer has the advantage of low cost and

operation simplicity, it is the choice if the speed and data handling capability meet the requirement. Nowadays, PC computers can easily be connected to a network system. Data acquired on a PC computer can be loaded to the network for further complicated analysis and information exchange.

When the SCM5000E CID camera system was purchased in July 1993, the manufacturer only supplied the MS-DOS version standard function library and a simple example program (as *example.c* in the manual). Image display software was not provided. The manufacturer gives users the flexibility of developing user-defined software according to specific needs. But on the other hand, the programming procedure takes significant effort. One year later, the manufacturer introduced Labview (National Instruments) version application codes. In the Labview environment, programming becomes more intuitive and less demanding on language-oriented programming skills. In this environment, library functions are assigned to different icons according to their roles in a program. These icons resemble electronic elements. Writing a program is just like connecting an electronic circuit with these virtual electronic elements. A user can develop software without having to know much about programming languages. However, if a function is not within the standard library, the user still need to write the function with a standard programming language and add it into the standard library or preferably a user-defined function library.

Although assisting programming tools such as Labview makes programming easier, MS-DOS based software remains the most efficient as long as running speed is a major concern. In this work, available memory for data storage, sampling speed and cycle duty of the CID system are very critical for high-speed high-throughput DNA sequencing. Software overhead time and the memory occupied by operation systems must be minimized. Therefore, the software for CID operation in this work is developed with MS-DOS version of C compiler (Microsoft QC2.5). Appendix A gives an overview of the organization of the software.

Function library *mce.lib* Building a function library to organize many specific tasks in a large program is an efficient way most programmers prefer to write and maintain complicated software. All of the functions written for the CID operation and data analysis in this work are organized into the library *mce.lib*. A main program can call the functions in the library by linking the object of the main program with the library. This procedure reduces repeated tasks, makes the software easier to maintain, and reduces errors. The description and "C" codes of these functions are listed in Appendix B.

CID calibration program *cal.c* When a CID control program is run, calibration data need to be downloaded to the 286 microprocessor before acquiring images. Since the calibration data of the CID camera will not change as long as the temperature of the CID head remains stable and the camera is not

reset, the CID only needs to be calibrated at the first time. The calibration data is saved to hard disk. Subsequent programs only need to reload the calibration data from hard disk and download them to the CID controller without having to calibrate the camera again. This saves time because the calibration process takes several minutes. The exact time for calibration depends upon the active imager size defined in the initialization step. The "C" code of the program is listed in Appendix C.

Image display program *img.c* In order to focus and choose the subarray locations and sizes, an image display software is needed. Although image analysis programs are commercially available, it is not practical to use the software to display the images taken from the CID system for the purpose of focusing. This is because in order to look at the image with the software, the image must been saved, the CID control program must quit and a different image analysis software needs to be initialized before the image can be displayed. To avoid this interruption and delay, program *img.c* is written to integrate both image acquisition function and image display function. The image display function incorporated is a basic image display function without advanced image analysis capabilities. For focusing and subarray allocation, this program is sufficient.

To acquire the image of an experimental setup, the size of the image is usually larger than 32K pixels. A memory block that can be allocated in base

memory is not large enough to handle the image. In program *img.c*, an image is always read into XMS. The program can read one or multiple subarrays the sizes and locations of which can be defined in the *imgparam.cfg* data file. The subarray parameters can also be changed during image acquisition. For the purpose of focusing in this work, only one subarray needs to display during focusing. Although the program *img.c* is currently for acquiring and displaying the image of one subarray, the program structure for acquiring and displaying multiple subarrays has been embedded in the program. With some modifications, the program can easily be upgraded for handling multiple subarrays if necessary.

Pixels in the CID imager have uneven bias levels if not corrected. The difference in bias level introduces spatial noise in an image display plane. In order to compensate for the uneven bias level, a background image needs to be acquired and subtracted from the subsequent images. The background subtraction also eliminates the shadowing effect in an image due to dark current accumulation in the lower part of a subarray. This effect is only important when the CID is operated at room temperature and the subarray is large. When cooled with liquid nitrogen, dark current is suppressed.

Before taking an image, the shutter is closed and a background image is acquired and saved in XMS for background subtraction. An object image is acquired and stored in XMS for displaying with correct scale and magnification.

Before display, each pixel data in the object image is subtracted by the corresponding pixel data in the background image. The image can be displayed as many times as necessary as long as it is still in the XMS. If the size, location, read gain and exposure time of a subarray are not changed, it is not necessary to acquire a background image before taking each image, except for the first time.

To display the subarrays, the CID system uses the host computer VGA monitor as image display output. The VGA monitor is set to 320 x 200 pixels, 256 color video mode. To set up a linear black and white (B&W) gray scale, equal intensity of red, blue and green is mixed in a 24-bit word packet. Although each color is represented with an 8-bit integer, because two high bits in the 8-bit word must be 0, there are only 64 intensities instead of 256. The B&W linear gray scale thus has 64 intensities. In the order from lowest to highest, each intensity is "painted" into 4 "holes" of the 256-color palette. A pixel intensity can thus be displayed with these 256 "colors" in the palette.

The full-scale pixel intensity of the CID is 14-bit. Since the display color scale is 8-bit, it is necessary to bring the pixel intensity scale down to the 8-bit display scale. This can be done by a bitwise right shift operation on each pixel data. Although the 6-bit right shift guarantees all pixel intensities are within the display scale, shifting 6 bits downwards erases all intensities below 64. This is not desirable when the signal is very weak or the low intensity features in an image is interesting. In the program *img.c*, the bit number to shift can be

selected during display so that the signals can be trimmed to fit the best scale and display the clearest picture.

The program allows changing most of the image acquisition parameters during display. The flow chart of the program is shown in Fig. 4. The "C" code of the program is listed in Appendix D.

Multiplexed CE application program *mce.c*. This program is written for acquiring a series of fluorescence images from the laser illuminated windows on one or multiple capillaries during capillary electrophoresis, building electropherograms from the images, and processing the electropherograms if necessary.

The subarrays corresponding to the detection windows on capillaries are allocated during focusing with program *img.c*. The subarray parameters are saved in the input data file *subparam.cfg*. For multiplexed CE application in this work, a subarray size usually does not exceed 32K pixels. Each subarray can be read out independently with different read gain, different read number or even with different exposure time.

In program *mce.c*, two exposure modes can be used. In one mode, the exposure time for all frames are the same. In the other mode, exposure time gradient is applied (40). For the frames following the first frame from which exposure time gradient is applied, the exposure time for each frame increases by the increment calculated in the function *hms()*. If the increment is the same for

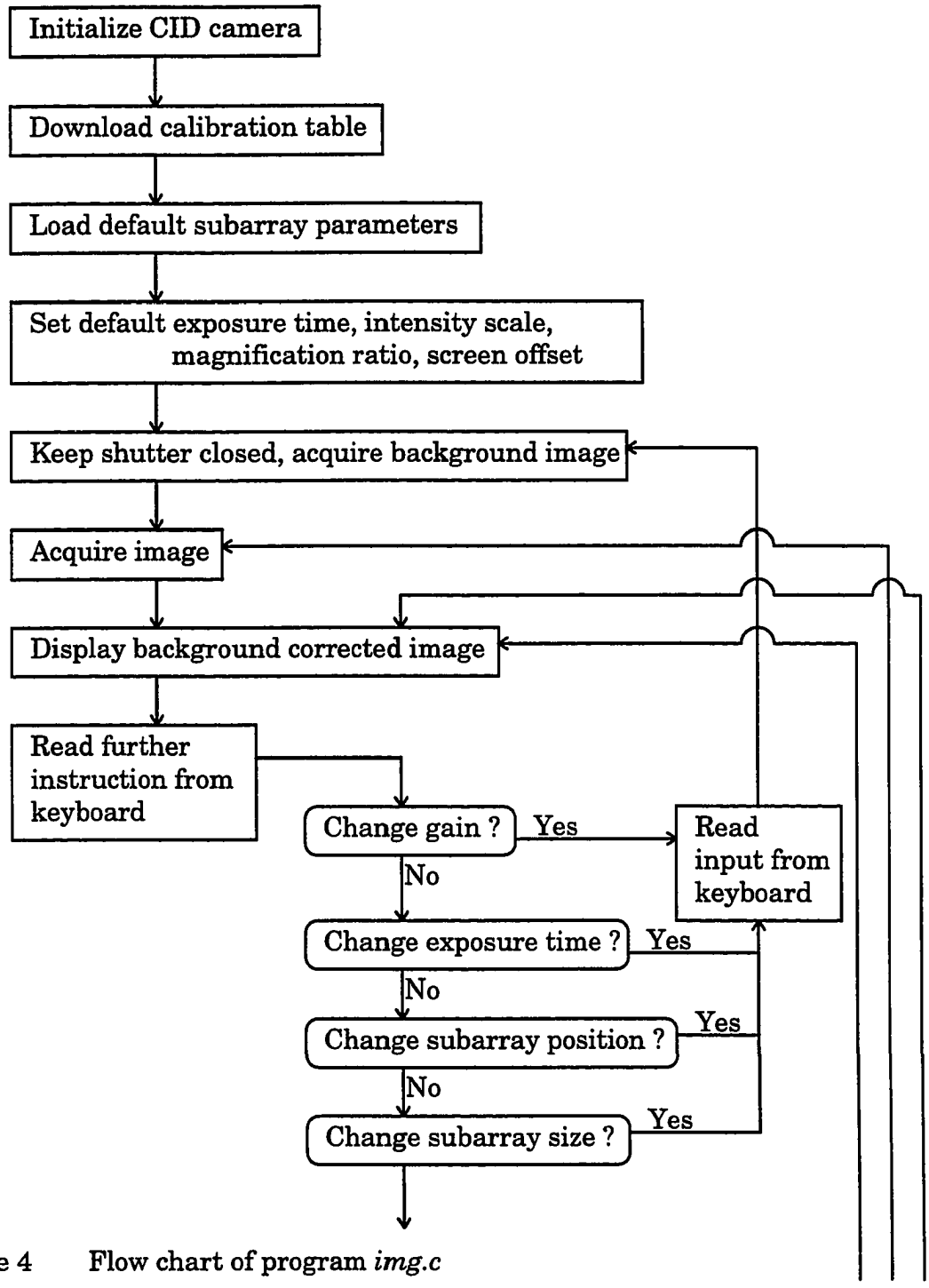
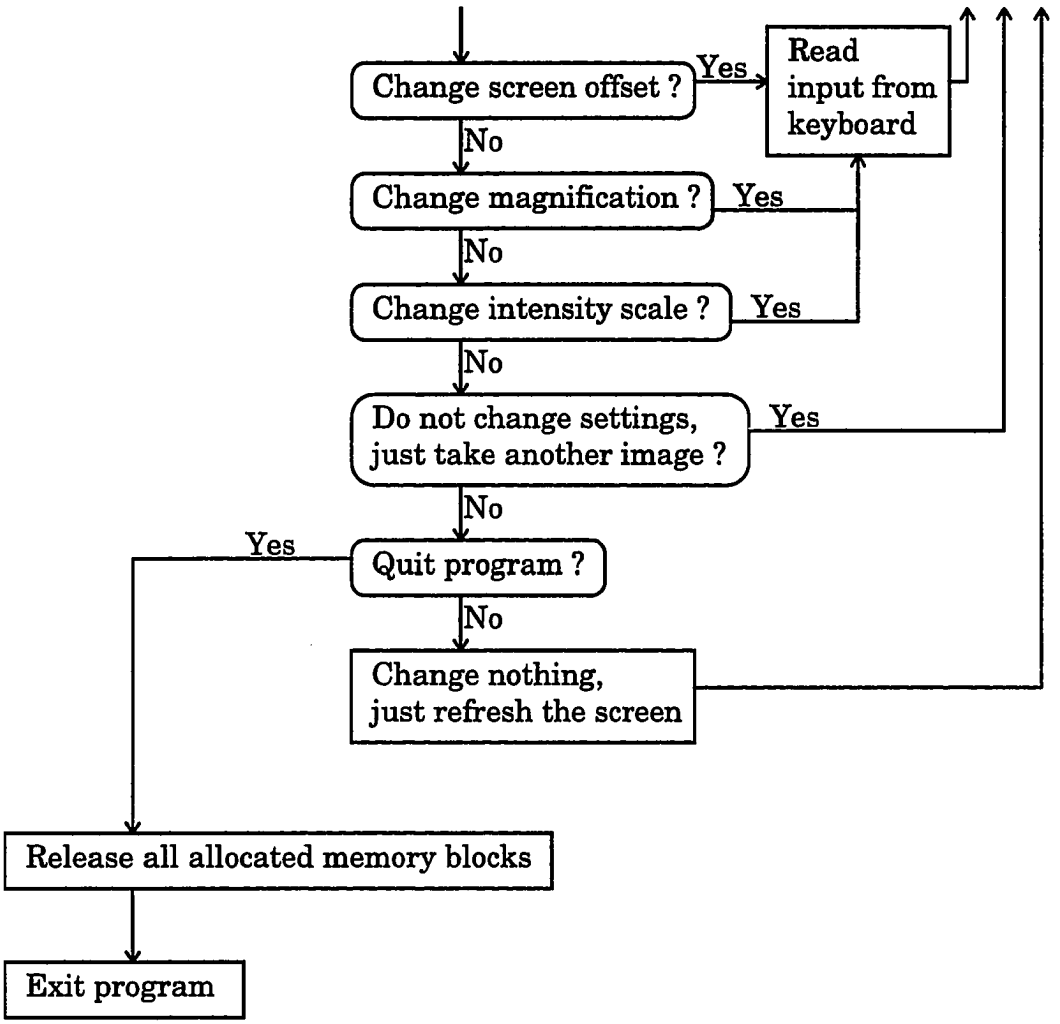


Figure 4 Flow chart of program *img.c*



Figue 4 (continued)

all of the frames, the exposure time gradient is considered linear. If the increment changes, the exposure time is considered nonlinear. Due to the nature of DNA sequencing, linear exposure time is sufficient. Nonlinear exposure time function is not supported in program *mce.c*. It can be added readily if necessary.

During data acquisition, the pixel data of a subarray is first read into base memory. Averaging is carried out for NDRO data, then the data is transferred into XMS for fast storage. Only one dynamic base memory buffer needs to be allocated for reading subarrays, even though there may be hundreds of subarrays, since subarrays are read sequentially according to their order in the input file *subparam.cfg*. After one subarray is read, the subarray data is stored in XMS, and the base memory buffer is resized according to the next subarray size. The base memory buffer works as a shuttle moving between the CID pixel data FIFO and the XMS of the host computer.

After a run is finished, all the data stored in XMS are sequential images. The electropherogram of a capillary is built by reading the intensity of the corresponding pixel in all of the frames into a base memory buffer. The data in the buffer is then saved to hard disk in a preferred format. The format can be binary, ASCII or a chromatogram format such as Chrom Perfect format so that the data can be analyzed with the chromatographic software. Alternatively, data processing functions can be called to treat the data before they are saved.

The flow chart of the program is shown in Fig. 5. The "C" code of the program is listed in Appendix E.

Appendix F lists the format of the input data files *miscpara.cfg*, *subparam.cfg*, and *imgparam.cfg*.

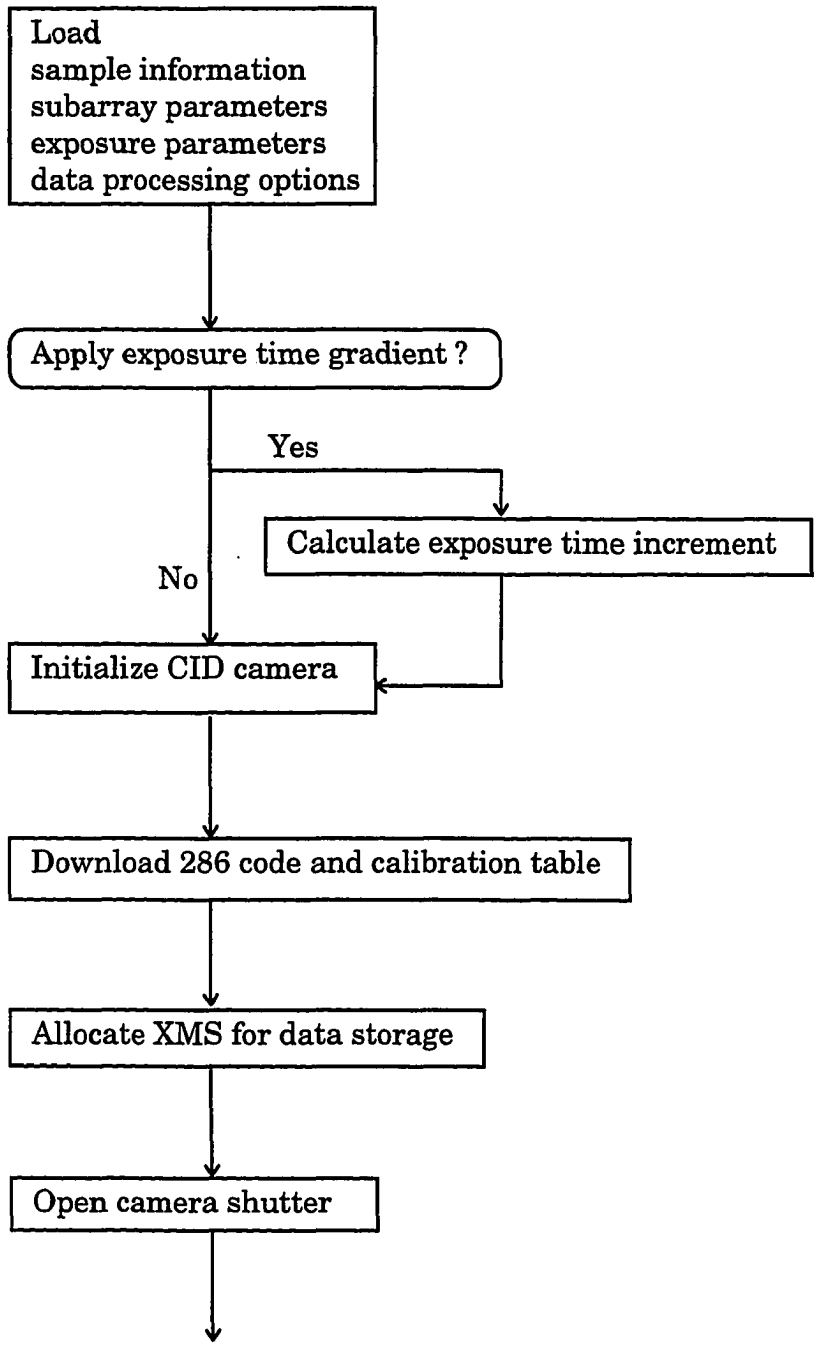


Figure 5 Flow chart of program *mce.c*

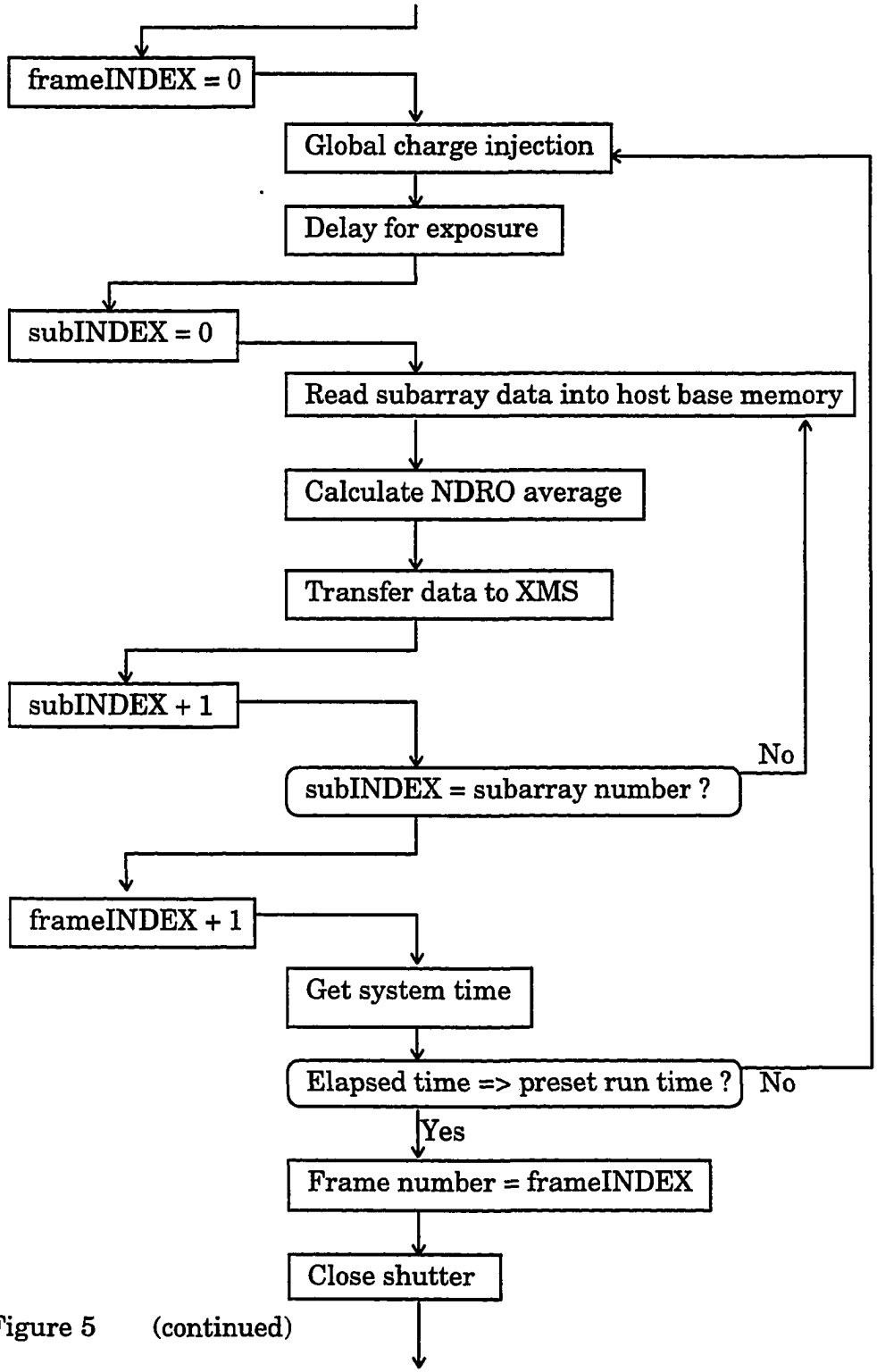


Figure 5 (continued)

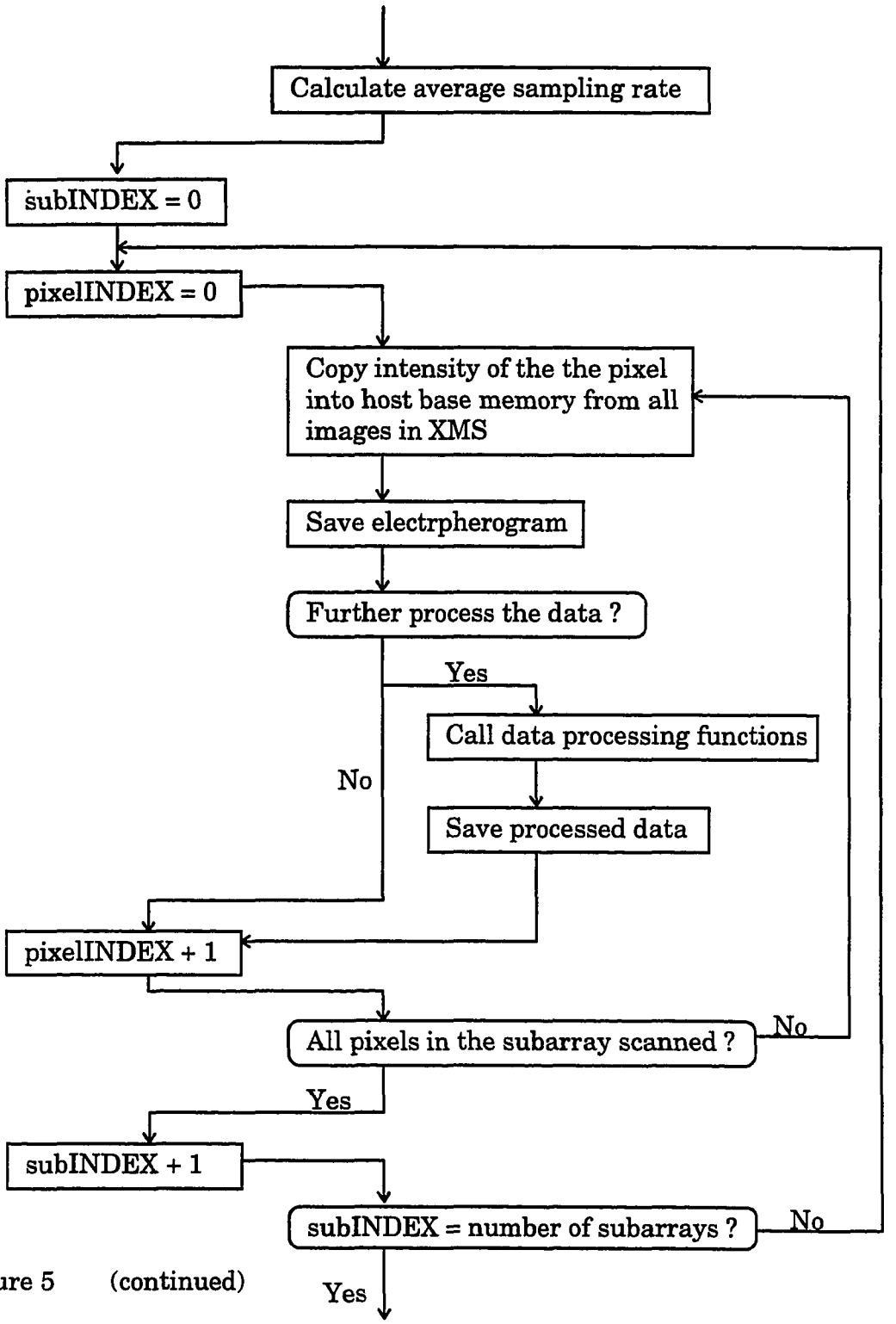


Figure 5 (continued)

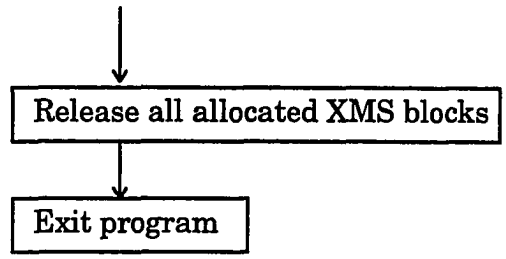


Figure 5 (continued)

REFERENCES

1. G. F. Amilio, M. F. Thompsett, and G. E. Smith, Bell Syst. Tech. J., 49, 593 (1970).
2. G. J. Michon and H. K. Burke, Dig. IEEE, Int. Solid State Circuits Conf., 16, 138 (1973).
3. J. R. Janesick and T.S. Elliot, History and Advancements of Large Area Scientific CCDs; Astronomical Society of the Pacific: Tucson, AZ, 1991.
4. R. S. Aikens, C. R. Lynds, and R. E. Nelson, Proc. SPIE, 78, 65 (1976).
5. D. J. Arndt-Jovin, M. Rober-Nicoud, S. J. Kaufman and T. M. Jovin, Science, 230, 247 (1985).
6. Y. Hiraoka, J. W. Sedat, and D. A. Agard, Science, 238, 36 (1987).
7. T. M. Jovin and D. J. Arndt-Jovin, Annu. Rev. Biophys. Chem., 18, 271 (1989).
8. J. J Linderman, L. K. Harris, L. L. Slakey, and D. J. Gross, Cell Calcium, 11, 131 (1990).
9. B. R. Masters, Proc. SPIE, 1448, 98 (1991).
10. T. M. Jovin and D. J. Arndt-Jovin, Proc. SPIE, 1439, 109 (1992).
11. P. M. Epperson, R. D. Jalkian, and M. B. Denton, Anal. Chem., 61, 282 (1989).
12. P. M. Epperson and M. B. Denton, Anal. Chem., 62, 1513 (1989).

13. K. L. Ratzlaff, Anal. Chem., 52, 1415 (1980).

14. R. D. Jalkian, R. S. Pomeroy, J. D. Kolczynski, M. B. Denton, J. M. Lerner, and R. Grayzel, Am. Lab., February, 47 (1989).

15. R. D. Piccard and T. Vo-Dinh, Rev. Sci. Instrum., 62(3), 584 (1991).

16. P. M. Epperson, J. V. Sweedler, R. S. Bilhorn, G. R. Sims, and M. B. Denton, Anal. Chem., 63, 327A (1988).

17. J. A. Cosgrove, and R. B. Bilhorn, J. Planar Chromatogr., 2, 362 (1989).

18. M. Baker, and M. B. Denton, Planar chromatography and electrophoresis, in Charge Transfer Devices for Chemistry, J. V. Sweedler, K. L. Ratzlaff, M. B. Denton, Eds., VCH: New York, 1993.

19. R. L. McCreery, CCD array detectors for multichannel Raman spectroscopy, in Charge Transfer Devices for Chemistry, J. V. Sweedler, K. L. Ratzlaff, M. B. Denton, Eds., VCH: New York, 1993.

20. G. R. Sims and M. B. Denton, Opt. Eng., 26, 999 (1987).

21. R. B. Bilhorn and M. B. Denton, Appl. Spectrosc., 44, 1538 (1990).

22. R. S. Pomeroy, R. D. Jalkian, and M. B. Denton, Appl. Spectrosc., 45, 1120 (1991).

23. T. W. Barnard, M. I. Crockett, J. C. Ivaldi, P. L. Lundberg, D. A. Yates, P. A. Levine, and D. J. Sauer, Anal. Chem., 65, 1231 (1993).

24. J. R. Janesick, T. Elliott, R. Bredthauer, and C. Chndler, X-ray instrumentation an astronomy, Golub, L. Ed., SPIE, Bellingham, WA, 1 (1988).

25. M. Ellia and K. Pollari, Nucl. Instrum. Methods Phys. Res. A, 288, 267 (1990).

26. J. C. Sutherland, L. Bohai, D. C. Monteleane, J. Mugavero, B. M. Sutherland, and J. Trunk, Anal. Biochem., 63, 446 (1987).

27. P. Jackson, V. E. Urwin, and C. D. Mackay, Electrophoresis, 9, 330 (1988).

28. L. B. Koutny and E. S. Yeung, Anal. Chem., 61, 1931 (1989).

29. Y.-F. Cheng, R. D. Piccard, and T. Vo-Dinh, Appl. Spectrosc., 44, 755 (1990).

30. J. V. Sweedler, J. B. Shear, H. A. Fishman, R. N. Zare and R. H. Scheller, Anal. Chem., 63, 496 (1991).

31. P. Jackson, Anal. Biochem, 196, 238 (1991).

32. K.C. Chan, L.B., Koutny, and E.S. Yeung, Anal. Chem, 63, 746 (1991).

33. R.D. Jalkian, and M.B. Denton, Proc. SPIE, 1054, 91 (1991).

34. A. E. Kargar, J. M. Harris, and R. F. Gestland, Nucl. Acid Res., 19, 4955 (1991).

35. D. A. McGregor and E. S. Yeung, Anal. Chem., 64, 1 (1992).

36. L. B. Koutny, and E. S. Yeung, Anal. Chem., 65, 183 (1993).

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

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

39. X. Lu and E. S. Yeung, Appl. Spectrosc., 49, 605, (1995).

40. Q. Li and E. S. Yeung, Appl. Spectrosc., 49, 825, (1995).

CHAPTER 3. EVALUATION OF THE POTENTIAL OF A CHARGE INJECTION DEVICE FOR DNA SEQUENCING BY MULTIPLEXED CAPILLARY ELECTROPHORESIS

A paper published in Applied Spectroscopy

Qingbo Li and Edward S. Yeung*

ABSTRACT

Despite the rapid growth in the use of imaging detectors in spectroscopy, the charge-injection device (CID) has unique features that have not been fully exploited. The advantages of CID as a two-dimensional array detector for laser-induced fluorescence detection in highly multiplexed capillary electrophoresis are evaluated. In such a system, the CID maintains both high sensitivity and high sampling rate which are usually difficult to achieve simultaneously with other array detectors. Applying the electronic windowing function significantly improves the scan rate and greatly reduces the volume of data generated. With 1-s exposure time and 488-nm excitation, the detection limit of the system is 10^{-12} M fluorescein with the device cryogenically cooled and 10^{-11} M fluorescein at ambient temperature. The low dark current of the CID imager allows operation at room temperature without significantly affecting sensitivity when combined with moderate laser powers. We demonstrate that the CID is well suited for

high-speed, high-throughput DNA sequencing based on multiplexed capillary electrophoresis with on-column laser-induced fluorescence detection.

INTRODUCTION

Highly multiplexed capillary electrophoresis (CE) is an attractive approach to overcome the throughput limitation of current DNA sequencing instrumentation in sequencing the human genome (1). The advantages of capillary electrophoresis for DNA sequencing in terms of high speed and low sample requirement have been demonstrated by many research groups (2-9). Highly multiplexed CE imposes great demands on the detection system. Laser-induced fluorescence detection (LIF) has been the major method employed in the automation of DNA sequencing (10,11). To be compatible with the high speed provided by CE and the high throughput of a large capillary array, a fast, sensitive two-dimensional array detector is required.

Charge-coupled devices (CCDs) have long been used for on-column LIF detection in liquid chromatography, slab gel electrophoresis and capillary electrophoresis (12-15). Recently, by using CCDs as array detectors, several research groups have developed capillary array electrophoresis with different configurations to pursue the same goal—high-speed, high-throughput DNA sequencing as part of the Human Genome Project. In this work, we investigate the possibility of using another type of charge transfer device—charge-injection

device (CID)—as an array detector for DNA sequencing. CID is a solid-sate imaging device similar to a CCD but has unique characteristics which CCDs do not have (16). The most important features of the CID include random pixel addressing, flexibility of user programmable architecture, large dynamic range, low dark current, anti-blooming imaging, high tolerance to irradiation, high quantum yield over a wide wavelength range, and non-destructive readout.

Most of the applications of CID reported are in astronomy and atomic spectroscopy (17). Unlike CCDs, no applications of CID in chromatography have been reported. One reason may be because low-noise scientific-grade CID cameras have only become available recently. While there are many CCD manufacturers offering various models with well-engineered software, few CID manufacturers exist. In CE, when only a single capillary or a small number of capillaries are involved, there is no obvious reason for using a CID while various CCD cameras are available. However, when a large number of capillaries need to be monitored simultaneously in an array format, the unique features of a CID camera can make a significant difference.

EXPERIMENTAL SECTION

The array detector used is the SCM5000E scientific grade CID camera system (CID Technologies Inc., NJ). The system includes a controller and a camera head. The camera head contains a 512H x 512V imager which is

installed in a dewar chamber to provide for cooling the camera with liquid nitrogen. The system can also be operated at ambient temperature. The system is connected to a 486 DX/33MHz host computer (Electra, Ames, IA). All commands and configuration parameters to operate the camera system are programmable from the host computer. Operation software is developed by using the CIDTEC standard function libraries. Software is written in C language by using Microsoft Quick C 2.5 as the compiler.

A Nikon F/1.4 lens (Nikor 28-mm f.l. wide-angle) is attached to the CID camera head. For different magnification factors, a Nikon extension tube is connected between the CID lens mount and the lens.

To operate the camera in the vertical direction, the head assembly is placed downward on a 9" x 9" x 1" aluminum plate. The lens mount and the lens protrude from underneath the plate through a cut hole. The aluminum plate is supported with four 1" diameter scaled bronze poles. Each pole can slide through a 1" diameter hole at the corner of the plate and can be tightened by a screw, so that the camera can be adjusted to the desired height. To make the camera more stable, the upper part of the dewar chamber is clamped between two pieces of half-circle aluminum plates. One of the half-circle plates is attached to two bronze poles which are fixed onto the main aluminum plate.

A U-shaped capillary holder is fixed on a translational stage which is attached to a magnetic stage. A capillary is placed horizontally across the

U-shaped holder. One or more capillaries can be placed in parallel onto the holder. For separating the DNA samples, a capillary column is coated with polyacrylamide and filled with a matrix made from a mixture of polyethelyneoxide polymers, TBE buffer solution and 5 M urea. Detailed procedures for the preparation of polyacrylamide-coated capillaries and the separation matrix are reported in the work of Chang and Yeung (18). A 45 cm long, 360 μm o.d., and 75 μm i.d. capillary is used here. A 1-cm section of the polyimide coating is burned off 30 cm from the injection end to form a detection window. The capillary is filled with water and then the polymer matrix under 400 psi pressure provided by a compressed nitrogen gas tank. Two 10-ml vials containing TBE buffer are put at both ends of the capillary. Samples are injected at –9 kV for 6 s. To run electrophoresis, –9 kV is applied at the injection end of the capillary with the other end grounded. After the capillary is filled with the polymer matrix but before a sample is injected, the capillary is pre-run for 10 min to stabilize the baseline. After each separation, the matrix is pushed out of the capillary by compressed N_2. The capillary is rinsed with water and filled with fresh matrix again for the next separation. The replacement procedure usually takes about 20 min. To preserve the capillary while not in use overnight or over several days, the capillary is rinsed with water and blown dry. A capillary can be used for more than one week with proper care.

Excitation is provided by the 488-nm line from an air-cooled Ar^+ laser (Uniphase, San Jose, CA). A 488-nm line filter is placed in front of the laser head to remove plasma emission. The laser line is focused onto the capillary with a 10-cm lens (Melles Griot) from the horizontal direction. The position of the camera is adjusted to a proper height so that the image of the inner bore of the capillary can be totally focused onto one pixel on the CID imager. A 515-nm cut-off filter is used to discriminate against scattered laser light.

RESULTS AND DISCUSSION

The primary purpose of this work is to evaluate and optimize the operation of the CID camera for highly multiplexed capillary electrophoresis for DNA sequencing. Only one capillary is used in the setup in this work. However, the hardware and software were developed to accommodate a 100 capillary array without further modification. Actual sequencing with 100 simultaneous channels will be published elsewhere.

Hardware In multiplexed capillary electrophoresis, in order to achieve high sampling rate and to skip over irrelevant pixels, only one row of pixels which contain the images of the illuminated windows of the capillary array are to be scanned. The rigidity of the setup is important. In this work, the capillary is clamped between two halves of a U-shaped holder. Since there are no moving parts in the setup, after focusing, the image of each capillary is maintained on

the same pixel throughout many cycles of manipulation such as filling with the matrix, matrix replacement, and sample injection. The capillaries used in a 100 capillary array will be 75 μm i.d., 150 μm o.d. By focusing each capillary onto two pixels, the CID camera can accommodate up to 250 capillaries. Alternate pixels therefore represent the capillary bore and the capillary walls, respectively. In this work, 75 μm i.d., 360 μm o.d. capillaries were used for ease of handling. This does not affect the evaluation of the CID camera.

When a laser beam irradiates a capillary array, one can focus the illuminated path along an imager row or an imager column. The two orientations may result in different subarray readout rates. In CCDs, the readout rate is different if one directs an n¥1 subarray along the parallel registers compared to the same subarray along the serial registers. To appreciate the difference in a CID camera, one needs to understand the subarray scanning sequence of the CID imager. When the camera begins to read a subarray, it first resets the horizontal (H) and vertical (V) scanners. The V-scanner slews to the first row in the subarray. The H-scanner then slews to the first column in the subarray. The whole row (line) is then read out. The V-scanner then does a line increment to slew to the second row in the subarray. The H-scanner is reset and again slews to the first column of the subarray. This loop repeats until all of the pixels in the subarray are read. If the subarray is oriented along a column, the imager needs to go through n times of horizontal slewing, n times of H-scanner reset and n–1 times of line

increment in order to readout all the pixels. Although the slew rate of the CID camera is as high as 5 MHz, the time spent in repeated slewing is not negligible when n is large and the n¥1 subarray is not at the edge of the imager. For example, when the camera is operated at 9 kHz pixel rate, if a 400¥1 subarray is located at the 250th column between the 51st row to the 450th row, the experimentally measured subarray readout time is 47 ms for row orientation and 86 ms for column orientation.

In contrast, the column orientation would be preferred if spatial cross-talk in a CID imager is significant. It was reported that row-column cross-talk increased linearly with signal level while column-column cross-talk remains insignificant until the signal level approached saturation. The row-column cross-talk in an unilluminated pixel was only 0.11% of the signal in the illuminated region even for the early model CID. In the current model CID system, spatial cross-talk, if any, is not found to interfere with the measurement of laser-induced fluorescence in the CE system.

It is because the SCM5000E CID system employs a pseudo-random scheme to achieve random access that led to such a subarray readout rate difference. In future models, truly random access mechanism will be implemented. The difference between the two orientations will therefore be eliminated.

Noise For CID cameras, the limit of detection (LOD) is determined by background fluctuations, integration time, duty cycle, read-noise, dark current and quantum yield at the specific wavelength range. Under different conditions, the predominant noise source is different. The CID is a noisy detector compared to the CCD. With single read, the read-noise is 250 e while the read noise of a typical CCD is 5-10 e. Since read-noise decreases with the square-root of the number of reads, with 100 non-destructive readouts (NDRO), the noise is reduced to 25 e, which is still higher than that of the CCD. However, this does not mean that one can obtain better LOD by using the CCD camera for a specific situation. In fact, other factors play more important roles in determining the LOD of LIF detection in CE.

It is natural to think about applying multiple NDRO to decrease CID read-noise in any case. But, multiple NDRO increases readout time, decreases the duty cycle and in turn degrades LOD and sensitivity. For a 400 x 1 subarray, if 100 NDRO is employed in order to reduce the read-noise by 10-fold, the subarray readout time will be about 4.4 s. This will obviously be too slow for any CE applications.

Fig. 1 shows the relation between the measured baseline noise and the number of NDRO. Curve 1 represents the case when the camera is operated at liquid-N_2 temperature without exposure but with 1000 ms integration time. Since liquid-N_2 cooling basically eliminates dark current, the measured baseline

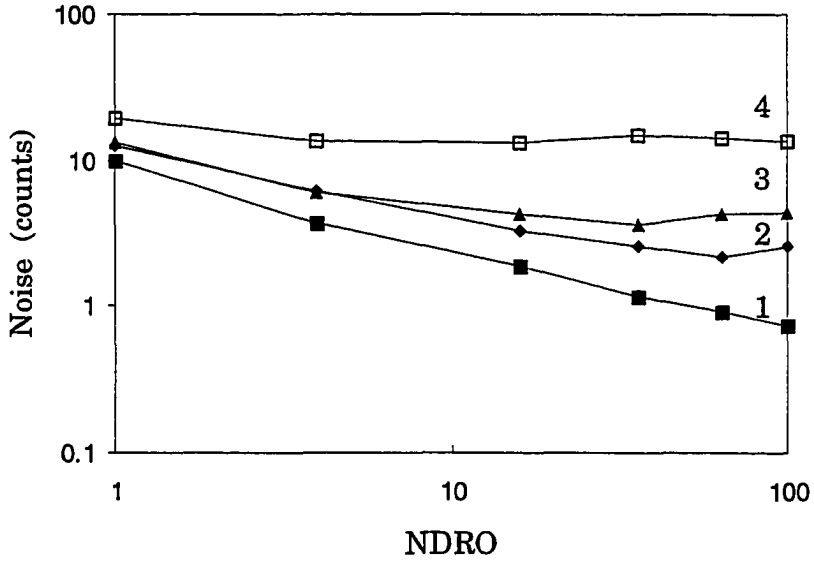


Figure 1. Noise as a function of the number of nondestructive readouts (NDRO). Labels are explained in the text.

noise is only read-noise, which decreases linearly with the square-root of the number of NDRO. Curve 2 represents the same measurements at ambient temperature. The entire curve is raised by 2 units. When the NDRO number increases, the contribution of dark noise increases due to the prolonged pixel dwell time. Therefore, the curve reaches a minimum after 36 NDRO and rises after 64 NDRO. This is why at ambient temperature, multiple NDRO is beneficial only within a limited range. Curve 3 represents the situation of on-column LIF detection for CE with 20 mM sodium phosphate buffer. The exposure time was 1000 ms and the camera was operated in the snapshot mode. When NDRO is fewer than 25, the baseline noise is dominated by read-noise. After that, the baseline noise is determined by background fluorescence and dark noise. Curve 4 represents the case of on-column LIF detection in a polymer-matrix-filled capillary. The camera was operated at ambient temperature with 450 ms exposure time. With more than 4 NDRO, there is no further reduction in baseline noise. This means that background noise has become comparable to the read-noise. So, for on-column LIF detection in DNA sequencing with polymer matrices, if one compares CID with CCD, the larger read-noise of the CID is no longer of concern. The CID is thus able to show its other advantages.

Sensitivity CID cameras have high quantum yields over a wide spectral range. The typical quantum yield is higher than that of the best

phototubes and higher than or comparable to that of a typical CCD. Therefore, by cooling the camera to liquid nitrogen temperature to suppress dark current and applying multiple NDRO to reduce read-noise, CIDs are expected to be at least as sensitive as PMTs and conventional CCDs.

The LOD of the camera at liquid nitrogen temperature for CE measurements is 10^{-12} M of fluorescein with 20 mW excitation. The camera runs in the snapshot mode with 1000 ms integration time and 100 NDRO. The pixel containing the image of the irradiated point of the capillary is read before and after an exposure. Then the charge is cleared. The net signal is the difference of the two readouts. The LOD was also measured at different laser excitation powers up to 20 mW, and was found to be inversely proportional to the laser power in this range.

We also tested the sensitivity of the CID camera when it is operated at ambient temperature. The camera is operated in the snapshot mode with 1000 ms integration and single readout. The LOD is 10^{-11} M estimated from the peak heights and the standard deviation of the baseline. The baseline noise under this condition is 9.9 counts (gain = 250), which is 10X higher than that for liquid-N_2 cooling and with 100 NDRO. Accordingly, the LOD degrades by a factor of 10.

Operation modes For high speed DNA separation, the sampling rate needs to be as high as 2 Hz. Unlike in CCD where pixels are scanned by

destructive readout, in CID, a separated charge-injection step is needed to clear the pixel. Therefore, the sampling rate of the CID is not only determined by its pixel-read rate, but also affected by the charge injection speed.

To test the system, we ran the camera at liquid-N_2 temperature in the snapshot mode with 1000 ms integration time. Charge injection was carried out by applying 100 cycles of global injection. Unfortunately, due to insufficient charge-injection efficiency, the CE peaks and the baseline were distorted and quantitative information was lost. Even following 100 cycles of global charge injection with 20 cycles of subarray injection, the integrated charge is still incompletely cleared (Fig. 2). It was suggested that 800-ms continuous global injection be applied in order to achieve complete charge injection. Such a delay is impractical for operating the camera at a reasonable sampling rate for CE.

Alternatively, one can keep integrating the signal without clearing the charge. The continuously integrated signal is read at the desired sampling rate. The actual integrated signal during each frame period is obtained by difference. When the overall signal approaches either digitization or pixel-well saturation, the charge is cleared and the cycle is repeated. In this mode, charge-injection need not be complete because the signal for each frame is based on a difference. However, there are disadvantages to operating the CID camera in this manner. First, because the absolute standard deviation of repeated measurement of the same signal is proportional to the signal level, the noise level in different parts

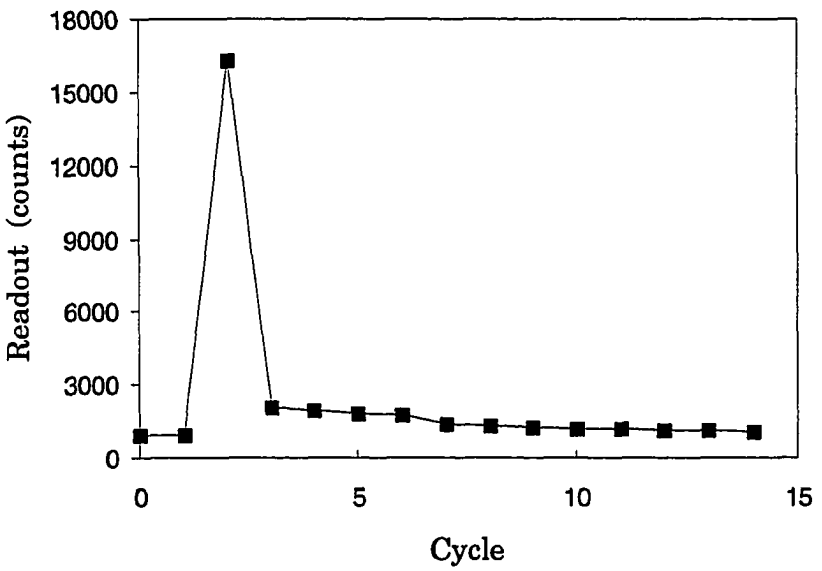


Figure 2. Charge injection efficiency of CID at liquid-N_2 cooling temperature. Each charge injection includes 100 global injection and 20 subarray injection. Only one exposure was taken at the #2 frame period.

of the reconstructed electropherogram will be different. Second, because the signal for each frame is the calculated difference of two readouts, the read noise is increased by a factor of the square-root of 2. Third, in a case such as DNA sequencing, there are hundreds of peaks in each run and the temporal spacing between adjacent peaks is very small. There is also a large background signal. So, after only a few peaks the accumulated charge needs to be cleared. Since each charge-injection cycle needs an additional readout, such an operation decreases the duty cycle.

One may use asynchronous scanning mode to overcome the above limitation. For example, to read a 400 x 1 subarray at 9 kHz pixel-read rate, if each pixel needs to be read 9 times per frame, the total readout time will be 400 ms. With a frame rate of 2 Hz in the snapshot mode, the exposure duty cycle is only 20%. If one keeps the shutter open all the time and keeps scanning the subarray continuously without waiting between two frame readouts, the actual exposure time will be 399 ms or an exposure duty cycle of 99.75%. The 400 pixels are exposed and read at different times, i.e. asynchronously. Although the same principle can also be applied to the CCD, when the subarray is n x m (n>1, m>1), blurring may occur in the CCD because the charge in a pixel is shifted and not confined within a fixed location like a CID.

To clear the charge for each frame, one needs to apply subarray charge injection. The charge in each pixel is cleared individually during each frame

without disturbing the other pixels. We found that it took more than 3.5 ms to clear the charge in a pixel completely by subarray charge injection. Although the exposure duty cycle can still be above 99% for a 400 x 1 subarray, the frame rate cannot be higher than 0.65 Hz, which is not adequate for high speed DNA sequencing.

Fortunately, we found that charge injection is more efficient at ambient temperature than at liquid-N_2 temperature. At ambient temperature, 100 cycles of global injection, which takes only 2.8 ms, can completely clear all levels of integrated charge. Fig. 3 shows the separation of Hind III digested ϕX174RF DNA fragments at ambient temperature. It clearly shows that after each exposure, the pixels can be restored to the previous baseline level.

Dark noise of the CID is generally lower than that of the CCD because of the difference in pixel structure of the two kinds of imagers. The dark noise of the CID is not serious even at ambient temperature. With 1000 ms integration per frame, the dark noise is 2.0 to 2.6 counts (gain = 250). With 10 s integration per frame, the dark noise is 8.1 to 8.3 counts (gain = 250), which is still lower than the single-read noise. It takes about 40 s for dark current to saturate the digitization scale (gain = 250). Therefore, if the CID is operated at ambient temperature, dark current is not the major source of noise, especially in matrix-filled capillaries.

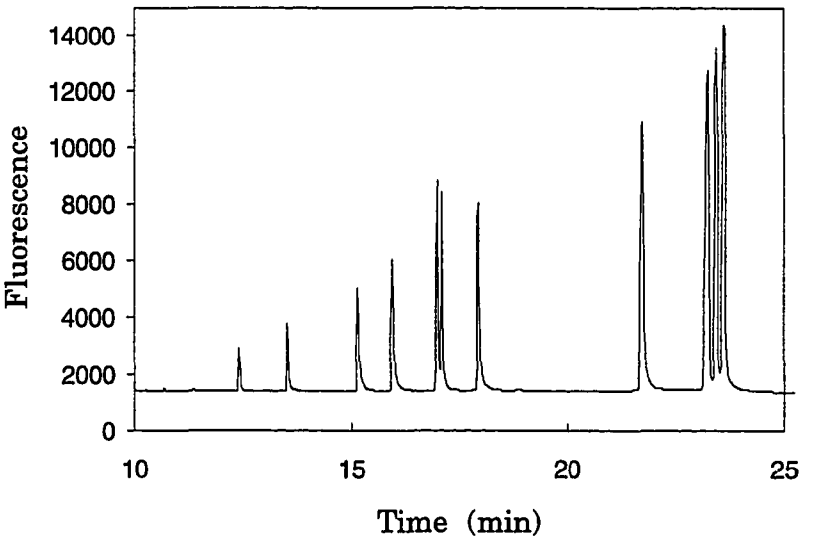


Figure 3. Separation and detection of Hae III digested ϕX174RF DNA fragments with the CID operated at ambient temperature. A cut-off filter was used in fluorescence detection.

After optimization, we establish the operation sequence as shown in Table I for a gain setting of 250. The time for each step is also measured. The subarray configuration is set by assuming a two-channel system as the one in previous work (2). The scheme of two 200 x 1 subarrays is for a 100-capillary array. The scheme of two 500 x 1 subarrays is for a 250-capillary array. Because the CID has 512 x 512 pixels, as many as 256 capillaries can be set up by focusing each capillary onto two pixels.

In the axial direction of the capillaries, if the fluorescence images cover more than one pixel each, there is no need to read all of these pixels and combine the intensities because the measurement is background-noise limited. For the same reason, if one uses a CCD for detection, there is no need to bin several pixels. Binning unilluminated or less intense pixels with the most intense pixel would in fact decrease S/N because dark current is increased.

In CCDs, if two separated lines (n x 1 subarrays) need to be read, the pixels between these two lines also need to be read out. This tremendously increases the number of pixels to be read and thus the volume of data to handle. The further apart the two lines are, the more extra pixels need to be read. In contrast, in CID, only the subarrays or pixels containing useful information are selectively read out. If the two laser lines need to be located far from each other for optimal optical coupling (2), the sampling rate is not affected because the number of pixels to be read does not change.

Table I. Sequence of CID operation for LIF detection in multiplexed capillary array

	Timing for each step	
Operation step	100-capillary array	250-capillary array
Global charge injection	3 ms	3 ms
Exposure	448 ms	384 ms
Read subarray A	23 ms	58 ms
Transfer data to XMS	3 ms	5 ms
Read subarray B	23 ms	58 ms
Transfer data to XMS	3 ms	5 ms
Repeat above steps	500 ms/frame	500 ms/frame
Transfer data from XMS to hard disk after run	5 min for 7200 frames	10 min for 7200 frames
Duty cycle for the first pixel in subarray A	90%	77%
Duty cycle for the last pixel in subarray B	99%	99%

Timing of CID operation The SCM5000E CID system can run with a pixel-read rate of 8.8 kHz to 100 kHz, which corresponds to a gain of 255 to 0. Gain defines the digitization sensitivity of the camera. At different read rates, the pixels require the same amount of light to saturate. In other words, the same amount of light produces the same number of charge carriers in a pixel independent of digitization gain. But the digitization sensitivity of this amount of charge carriers is different at different read rates. The faster the camera reads, the less sensitive the digitization process is. In addition, read-noise increases rapidly when the camera reads faster than 33 kHz. Below 33 kHz, the read-noise is relatively constant. As shown in Table 1, our readout time is small compared to exposure time even if the camera reads at the slowest read rate. Therefore, we choose to operate the camera at 9 kHz which is selected by setting the gain to 250, so that the digitization sensitivity is maintained at the highest level available.

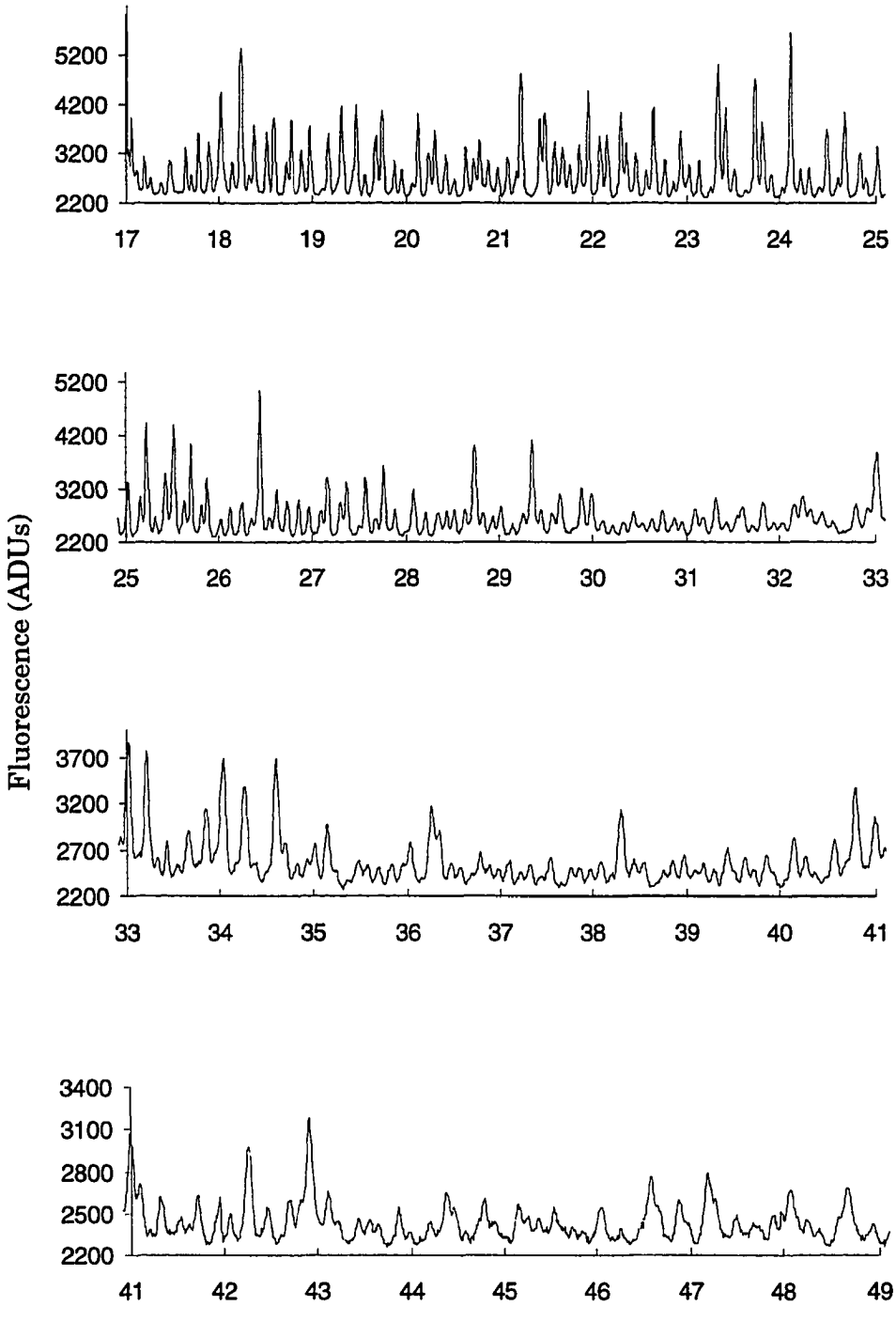
In the operation sequence shown in Table 1, the exposure duty cycle for each pixel is different because their actual exposure times are slightly different. For the first pixel in the first subarray, the duty cycle is 90% because the exposure time for this pixel is only 448 ms. For the nth pixel to be read, it gains an extra (n–1)*110 μs exposure time while the camera reads the first (n–1) pixels. For the last pixel, the actual exposure time is 494 ms and the duty cycle is 99%. The result indicates that it is advantageous to keep the shutter open. If

the camera operates at 10-Hz frame rate, the duty cycles for the first pixel and the last pixel to be read are 55% and 95%, respectively, which are excellent taking into account the very fast sampling rate.

The flexible reading mode of CID has additional advantages. If the light intensity is higher at one side of the capillary array than at the other side due to absorption or light scattering, the corresponding subarray can be scanned from the side where the light intensity is higher so that S/N is more even across the capillary array. If the capillary array is illuminated by a Gaussian-shaped light profile (2), the fluorescence from the center of the capillary array may saturate the middle pixels before the pixels on the sides have accumulated sufficient charge. What one can do is to read the middle pixels with shorter integration times and read the pixels on the sides after longer integration times. This can be accomplished easily by software programming.

Fig. 4 shows the data acquired from 1 channel of separation of DNA fragments from a Sanger reaction following the optimized sequence of CID operation, as shown in Table I, column I. The test sample was the set of DNA fragment ladder prepared by the Iowa State University Nucleic Acid Facility using the standard dye-labeled terminators (ABI) and TAQ polymerase. The sample preparation procedure was not altered in any way from that used to produce samples for the commercial DNA sequencing instrument 373A

Figure 4. Separation and detection of a PGEM/U DNA sample after Sanger reaction with standard dye-labeled terminators. A cut-off filter was used in fluorescence detection. The fgure is plotted in four sections to show the details.


5200
4200
3200
2200
17 18 19 20 21 22 23 24 25
5200
4200
3200
2200
25 26 27 28 29 30 31 32 33
3700
3200
2700
2200
33 34 35 36 37 38 39 40 41
3400
3100
2800
2500
2200
41 42 43 44 45 46 47 48 49
Fluorescence (ADUs)
Time (min)

Sequencer (ABI). The injected sample is identical in concentration and composition to those suggested for loading into the commercial instrument. The test sample was independently analyzed (sequenced on the commercial instrument) and was found to be well-behaved. Judging from this side-by-side comparison, the sensitivity and temporal resolution is clearly sufficient for DNA sequencing applications.

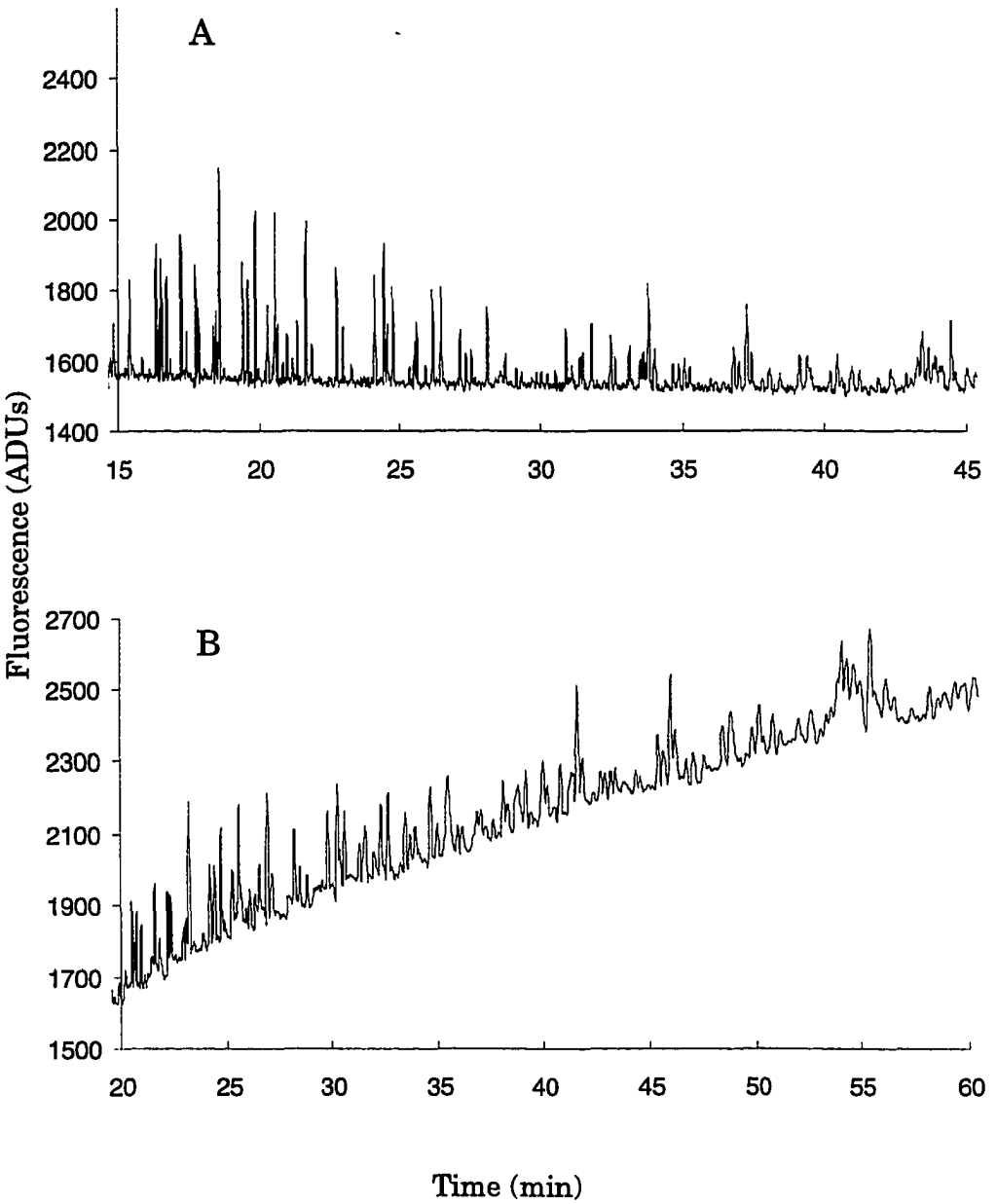
Exposure-time gradient In DNA separation by CE, different sizes of fragments migrate at different velocities. Their residence times at the detection window are different. The larger fragments elute later and stay at the detection window longer. In addition, peaks of the larger fragments become broader because of a loss in separation efficiency. If the camera runs at the same frame rate throughout, the peaks of the larger fragments contain more data points than those of the shorter fragments. However, if the camera runs at a slower frame rate to monitor a larger fragment, the total fluorescence from that fragment can be concentrated into fewer sampling points. The S/N will thus be improved. This is very useful because the larger fragments in a DNA sequencing sample are typically at lower concentrations than the shorter fragments due to the nature of the polymerase reaction. In addition, the peak intensity of a fragment is inversely proportional to its migration time due to electrokinetic injection bias (19). With the flexibility of the user programmable features of the CID, we are able to change the exposure time dynamically

during electrophoresis to account for the migration velocity differences of the DNA fragments.

Programming to achieve an exposure-time gradient is simple. The exposure time can be programmed to as fine as several microseconds. In these applications, a linear exposure-time gradient is adequate. To achieve a finer gradient, we can fit the relation between exposure time and migration time to a numerical function derived from experimental data, and adjust the exposure time according to the fitted function in real time. To demonstrate this, we monitored the same separation of DNA sequencing fragments through a band-pass filter to reduce the number of peaks and to reduce the signal levels. From the electropherograms in Fig. 5, we can see that the S/N is improved for the large fragments after implementing exposure-time gradient. There is a sloping baseline in Fig. 5b, since the background is also integrated for progressively longer periods. This is however easily corrected by software after the run.

Data manipulation There are several ways to save and process the pixel data from a camera. The data during the run can be stored in RAM or onto the hard disk of the host computer. To store in RAM, the data size of each frame must be small. Otherwise the installable extended memory of a typical desktop computer will not be enough to store the data of a complete run. The speed

Figure 5. Detection of DNA fragments after Sanger reaction (A) without and (B) with exposure-time gradient. A line filter was used in fluorescence monitoring.


A
B
Fluorescence (ADUs)
Time (min)
2400
2200
2000
1800
1600
1400
15
20
25
30
35
40
45
2700
2500
2300
2100
1900
1700
1500
20
25
30
35
40
45
50
55
60

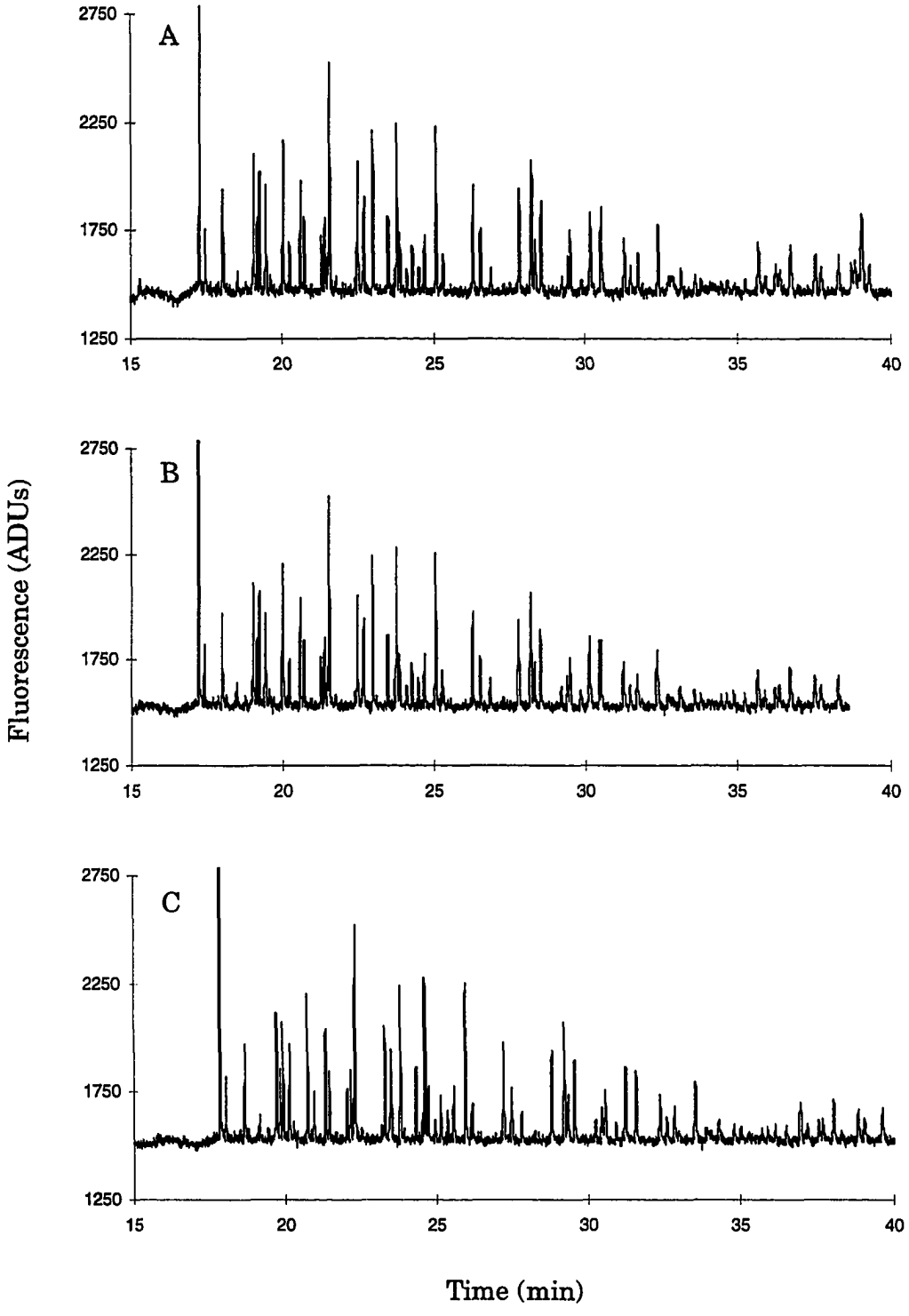
advantage of storing data during the run in RAM is obvious. Also, the computer does not have to re-read the large number of image files of thousands of frames from a hard disk in order to convert the image files into corresponding electropherograms during post-run data processing, which would take hours of computer time.

When the CID camera reads a subarray, pixel data is transferred to the base memory or extended memory of the host computer. For a 2-color 100-capillary array, only 400 pixels are to be read. For a desktop computer with 16 MB extended memory available, 20000 frames can be stored as floating-point numbers. If the camera is operated at 2-Hz frame rate, this is equivalent to 2.7 hrs. For a 250 capillary array, the maximum available run time will be 1.1 hrs. This is more than enough for the separation of a DNA sequencing sample, which takes less than 50 min. By using exposure-time gradient, the number of data points are also reduced. The image of one capillary occupies two pixels, with the centers of two adjacent capillaries being focused onto alternating pixels. The pixel between these two pixels corresponds to the capillary walls and contains useless information. The data size can therefore be further reduced by 50% by skipping over the void pixels. By proper software development, one can allocate up to 64 MB in a desktop computer for data storage. Therefore, the present system can be easily adapted to several hundred or a thousand capillaries in an array.

In our system, pixel data for a frame is first transferred to the base memory of the host computer. If multiple NDRO is applied, the average intensity for each pixel is calculated and only the averaged pixel data is saved. After electrophoresis is complete, the electropherogram for each capillary is built by directly reading the intensity of the corresponding pixel in all of the frames from the extended memory. The reorganized data for all of the capillaries are then saved to hard disk in binary format, which can be directly processed with standard chromatography software such as ChromPerfect (Justice Innovations, Palo Alto, CA). Alternatively, one can use built-in peak-finding algorithms to integrate the peaks while the data is still in memory in order to save data processing time. We have not put an effort to develop the latter algorithm.

By using two spatially separated excitation laser beams, the labeled DNA fragments can be excited twice to achieve two-channel, two-color base calling without the need to split the image (2). However, the appearance time of the same DNA fragment at the two channels is different. It is necessary to convert the time scale of one channel to match that of the other channel in order to do base calling. For the case in which the frame rate does not change, time-scale correction can be accomplished simply by multiplying the time scale of the first channel by the ratio of the effective length at the second channel to that at the first channel. After time-scale correction, the migration time of a peak in the two channels can be matched accurately, as shown in Fig. 6. With exposure-

Figure 6. Time scale correction for two channels. A and C are signals from the first and the second channels, respectively. B is the migration-time corrected plot of C. A line filter was used in fluorescence monitoring.


A
B
C
2750
2250
1750
1250
15
20
25
30
35
40
Fluorescence (ADUs)
Time (min)

time gradient, time-scale correlation for the two channels becomes more complicated but entirely tractable with the proper conversion algorithm. The important condition for accurate correlation of the time scales of the two channels is to keep the whole capillary at the same temperature and to make sure that the coating and the polymer matrix in the capillary is uniform, because time-scale correction is based on the assumption that a DNA fragment migrates along a capillary with constant velocity.

CONCLUSIONS

Compared with the multiple-wavelength fluorescence detection scheme where a spectrometer is employed (13) or with a multi-color split-image system (9), our system has the advantage of good S/N because the image of a capillary is focused onto only one pixel instead of dispersing it over a hundred pixels or splitting it into several images. Optical throughput is maximized and dark noise is minimized. In our system, the data size generated in each run is small. For a 1-hr run at 2-Hz frame rate in a 100 capillary array with two detection channels, only 1.44 MB data is generated. If one uses the multiple-wavelength fluorescence detection system (dispersing the fluorescence over 100 pixels) for a 100 capillary array, the number of pixels to be stored per frame will be 20000, and the volume of data generated will be 144 MB.

The CID camera can be operated at ambient temperature. This makes it simpler and more compact to incorporate into an automated DNA sequencing instrument. This is because without the liquid-N_2 dewar, the CID focal plane array occupies only a very small space. We showed that exposure-time gradient helps to improve the S/N for the larger DNA fragments and to reduce the volume of data generated. The random-access or electronic-windowing function of the camera is unique to CID cameras. With this capability, although the camera is operated at a pixel-read rate slower than a conventional CCD, it can achieve very high sampling rate with high exposure duty cycle and thus high sensitivity. Compared with CCDs, the advantages become more obvious when the size of the capillary array becomes larger and several spatially separated subarrays need to be read out. However, the CID camera suffers from relatively large read-noise. In order to overcome the limitation of read-noise when using CID in CE, one needs to apply higher laser powers for excitation.

At the present time, a prototype CID system costs around $25,000, which is comparable to good quality scientific CCD systems with similar data rates and sensitivity. This naturally did not take into account the effort that went into developing the software needed to run the CID system from the macro modules supplied by the manufacturer. According to the manufacturer, five times faster scientific grade CID system with lower read-noise and truly random pixel access is in development. When the new generation of CID systems becomes available, the detection performance in DNA sequencing will be even more impressive.

ACKNOWLEDGMENT

We thank S. VanGorden for helpful discussions about software development, and H. T. Chang and N. K. Fung for help in the capillary separations. 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. Joint U.S. Department of Energy and U.S. Department of Health and Human Services Report DOE/ER-0452P, "Understanding Our Genetic Inheritance—The U.S. Human Genome Project: The First Five Years"; Washington, DC, April 1990.
2. K. Ueno and E. S. Yeung, Anal. Chem., 66, 1424 (1994).
3. A. S. Cohen, D. R. Najarian, A. Paulus, J. A. Guttman and B. L. Karger, Proc. Nat. Acad. Sci. U.S.A. 85, 9660 (1988).
4. H. Swerdlow and R. F. Gesteland, Nucleic Acids Res. 18, 1415 (1989).
5. H. Drossman, J. A. Luckey, A. J. Kostichka, J. DíCunha and L. M. Smith, Anal. Chem. 62, 900 (1990).
6. N. J. Dovichi, H. Swerdlow, J. Z. Zhang, D. Y. Chen and H. R. Harke, Anal. Chem. 63, 2835 (1991).
7. X. C. Huang, M. A. Quesada and R. A. Mathies, Anal. Chem. 64, 967 (1992).

8. H. Kambara and S. Takahashi, Nature 361, 565 (1993).

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

10. L. M. Smith, J. Z. Sandes, R. J. Karser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent and L. E. Hood, Nature 312, 674 (1989).

11. J. M. Prober, G. L. Trainer, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jemsen and K. Baumeister, Science 238, 336 (1987).

12. Y.-F. Cheng, R. D. Piccard and T. Vo-Dinh, Appl. Spectrosc. 44, 755 (1990).

13. J. V. Sweedler, J. B. Shear, H. A. Fishman, R. N. Zare and R. H. Scheller, Anal. Chem. 63, 496 (1991).

14. A. E. Karger, J. M. Harris and R. F. Gesteland, Nucl. Acids Res. 19, 4955 (1991).

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

16. J. V. Sweedler, R. B. Bilhorn, P. M. Epperson, G. R. Sims and M. B. Denton, Anal. Chem. 60, 282A (1988).

17. P. M. Epperson, J. V. Sweedler, R. B. Bilhorn, G. R. Sims and M. B. Denton, Anal. Chem. 60, 327A (1988).

18. H.-T. Chang and E. S. Yeung, J. Chromatogr., in press.

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

CHAPTER 4. INVESTIGATION OF DETECTION CONFIGURATION FOR MULTIPLEXED CAPILLARY ELECTROPHORESIS WITH THE CHARGE INJECTION DEVICE

INTRODUCTION

Sequencing the 3-billion base pair human genome has become one of the major undertakings in the scientific community in the 1990s (1). After the introduction of fluorescent dye-labeled primers and dideoxynucleotides for the Sanger chemistry , laser-induced fluorescence (LIF) detection has been the major detection method in commercial automated DNA sequencers. With the existing automated DNA sequencers that use slab gel electrophoresis with LIF detection (2,3,4), the sequencing rate can be 10,000 to 100,000 bp per sequencer per year. Sequencing the whole human genome would take a thousand years even with 10 sequencers running simultaneously. Running more and more sequencers simultaneously can eventually reduce the time, but the expense will never be reduced. Although capillary gel electrophoresis has shown to be 25 times faster than slab gel electrophoresis (5), running one capillary at a time does not provide throughput advantages over slab gel electrophoresis because as many as 50 multiple lanes can be run simultaneously on one slab gel in a commercially available sequencer (6). Therefore, great interest has been focused on scaling up the capillary gel electrophoresis by running multiple capillaries simultaneously

in a capillary array (7-12) in order to both speed up the sequencing rate and reduce the cost. These multiplexed CE systems approach the same goal by using quite different detection schemes.

Post-column detection using a sheath-flow cuvette (13) has shown impressive LOD in CGE (14). The system has been adapted for a multiple sheath-flow gel-filled capillary electrophoresis system for DNA sequencing (11) in which 20 capillaries are monitored simultaneously. Excitation and detection occur in the gel-free sheath flow region with a cross-beam excitation and an orthogonal detection geometry. While the multiple sheath-flow system produces excellent detection limit, the complexity and the lack of flexibility of the system make it difficult to scale up for automatic high-throughput DNA sequencing.

For high-speed high-throughput DNA sequencing, the multiplexed CE system should have those merits including high sensitivity, high sampling rate, high reproducibility, compatibility with gel-filled capillary electrophoresis, and flexibility for automation. It seems that on-column detection is more promising to provide the best compromise among the above requirements.

One on-column approach utilizes an epi-illumination format in which a laser is focused onto a capillary by a microscope objective and the fluorescence is collected with the same objective by confocal detection (7-8). The 25 capillaries in the array are scanned sequentially by moving the capillary array across the optical detection region. Scattered light from the capillary wall is minimized by

focusing the probing beam down to only one tenth of the inner diameter of the capillaries. No crosstalk exists between capillaries. Laser power needed is the same as that required for single capillary detection since only one capillary in the array is monitor at a time. One disadvantage for using the system in high-speed high-throughput DNA sequencing is that after the capillary array is expanded to larger scale, the scanning mechanics may not be able to catch up with the high speed DNA sequencing separation.

Another on-column approach utilizes axial-beam excitation configuration in which an optical fiber is inserted into a capillary to introduce laser light to the detection window (10). Ten capillaries are run and monitored simultaneously. Fluorescence is monitored orthogonally above the detection window with a CCD camera. Introducing laser light into capillaries with optical fibers simplifies the optical alignment of the system. Once the optical fiber is inserted into a capillary, the optical alignment is settled. There are no moving parts in the system. Maintaining the capillary array is relatively easy. The system has high potential to scale up to accommodate hundreds of capillaries running simultaneously. However, for DNA sequencing where gel-filled capillaries are used, the system is disadvantageous because intrusion of optical fibers into gel-filled capillaries disturbs the separation system and possibly cause column blockage or contamination.

Ueno and Yeung (12) investigated and evaluated different on-column excitation schemes for distributing a laser beam to a large number of capillaries in an array. The investigations finally led to a two-window two-wavelength detection scheme potentially adaptable for two-color, two-intensity base-calling based on the standard four-dye Sanger chemistry, although actual DNA sequencing had not been carried out in the system. In the system, 100 capillaries were set in the 100 triangle-shaped grooves machined on an aluminum block. Each of the two laser beams was expanded and then focused to a fine line with a cylindrical lens so that the line cut across the array to illuminate all the capillaries. The capillaries were monitored with a CCD camera through a wide-angle camera lens. 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 with a larger chip. The work demonstrated the major features of a truly multiplexed CE system without moving parts and its potential for significantly increasing the speed of DNA sequencing. However, there are still many major steps need to be accomplished before the multiplexed CE system can be actually used for high-speed high-throughput DNA sequencing.

- First, a reproducible and easy-to-automate procedure for column pretreatment and regeneration is needed for efficient use of large number of capillaries while providing high quality separation (15).

- Second, a stable and easy-to-replace separation matrix is necessary to achieve single-base resolution while facilitating high speed separation (16).
- Third, modification in optical setup is necessary to couple laser power more efficiently into each capillary in the array so that less laser power is needed or more capillaries can be setup in an array (17).
- Fourth, a detection scheme for base-calling compatible with the multiplexing setup needs to be developed (18).
- Fifth, high sampling rate as well as high duty cycle of the array detector (CCD or CID) must be accomplished in order to catch up with the separation speed while a large number of capillaries are monitored with high sensitivity (19).
- Sixth, crosstalk between capillaries needs to be minimized.
- Seventh, a sample injection device is needed to inject 100 different samples into 100 capillaries simultaneously without causing contamination (15).
- Finally, an automatic and fast base-calling software is necessary to analyze the data for all the 100 capillaries.

Yeung's group has accomplished most of the above specifications (16-19). Compared with other published multiplexed CE systems for DNA sequencing (7-

11), the system here is very competitive and promising if evaluated comprehensively from all the above major aspects.

This chapter documents the investigation of excitation configuration, crosstalk, and alternative excitation laser source for high-throughput DNA sequencing. The chapter does not define the global optimum for the system. Instead, the experiences documented here may be used to guide further developments. All the work was done by utilizing the charge injection device (CID) as the array detector. From the other point of view, the work presented here may also be considered as continuing examination of the SCM5000E CID system for high-speed high-throughput DNA sequencing in addition to Chapter 3 in this dissertation.

EXPERIMENTAL SECTION

Separation Fused-silica capillaries of 150 μm or 350 μm o.d., 75 μm i.d. with polyimide coating or UV-transparent coating (Polymicro Technologies, Phoenix, AZ) are used for separation. A capillary is usually 45 to 60 cm total length, 35 to 50 cm effective length. Before filling a capillary with gel, the capillary is coated with polyacrylamide (20) or pretreated with 0.1 M HCl (15). The gel used is the mixture of 1.5% 8-million MW polyethylene-oxide (PEO) and 1.4% 0.6-million MW PEO dissolved in 1 x TBE running buffer with 3.5 M urea (15, 16). To fill a capillary with gel, a capillary is connected to a pressure device

such as a syringe or compressed nitrogen gas tank with a Teflon tubing. Gel filling usually takes about 10 min. After the gel is filled, the capillary is pre-run for 10 min before injection of a DNA sample. For electrophoresis of fluorescein, 2 mM Na_2HPO_4 solution of pH 9.3 is used as the running buffer. For a filling 100-capillary array in a preliminary experiment, one end of the array is bundled with a HPLC fitting. A Teflon tubing connects the fitting with a 20-ml syringe that is filled with the solution to be flushed through the capillaries. This method was not tested for filling gel into the capillary array. For filling gel into 100-capillary array, a pressure cell will be used (15).

Capillary array Different capillary mounts are used according to different excitation and detection geometry. It also depends upon whether the capillary array needs to be immersed in a refractive-index-matching liquid. Fig. 1 shows the mount used for two beam excitation in air. The aluminum block is 1.6 cm (W) x 6.0 cm (L) x 6.5 cm (H) . Two 0.3-cm wide, 3-cm deep grooves are cut on the mount to form two optical channels. The deep grooves helps to reduce scattered light background. The two channels are 0.9 cm apart to reduce interference of scattered light between the two detection channels when a two-wavelength, two-beam excitation scheme is employed. One or multiple capillaries are placed flat across the grooves and taped tight with Scotch tape (3M, St. Paul, MN). Since Raman and Raleigh scattering from air is negligible, there is no need to further isolate the two channels to reduce interference of

laser light scattered from one channel to the other channel. Fig. 2 shows the mount used for immersing the capillary array in a refractive-index-matching liquid. Two aluminum wings (plates) are attached perpendicularly to the mount shown in Fig. 1. These two plates are slightly higher than the mount. The two aluminum plates blot the scattered laser light propagating in liquid and prevent them from going from one channel to the other channel. For more complete isolation, a small piece of black plexiglass is shaped to fit into the opening between the two aluminum plates, as show in Fig. 2. The whole mount is then put into a liquid cell, as shown in Fig. 3. The bottom and two sides of the walls of the liquid cell are made of aluminum (in gray color). Two quartz plates are glued to another two opening sides of the U-shaped aluminum frame to provide the transparent path for incoming and outgoing laser beams. The top of the cell remains open for placing and removing the capillary mount. After the capillary mount in Fig. 2 is put into the cell and a liquid is filled, the top of the cell is covered with a piece of quartz plate. Optical filters are placed on the quartz plate.

The capillary mount shown in Fig. 4 is made for one-beam excitation configuration in air. On a 2"x2"x3" aluminum block, a 3/16" (W) x 2" (L) x 1.5" (H) groove is cut on the block to provide the optical path for incoming laser light. The groove is cut to such a depth that the scattered laser light can be dissipated into the environment without causing bright background in the

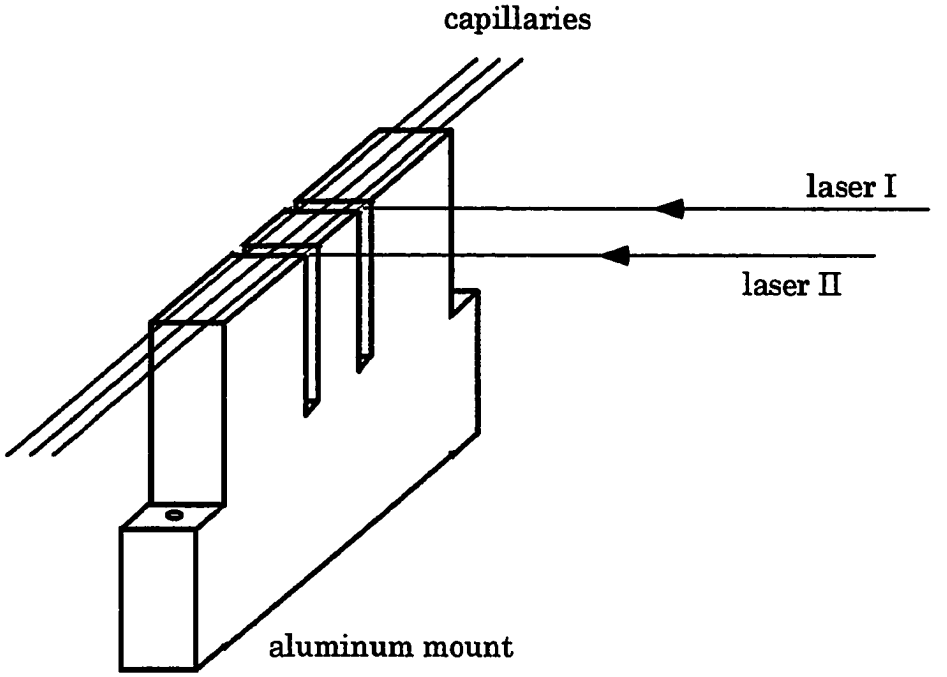


Figure 1. Capillary mount for two-window detection with two different laser beams in air.

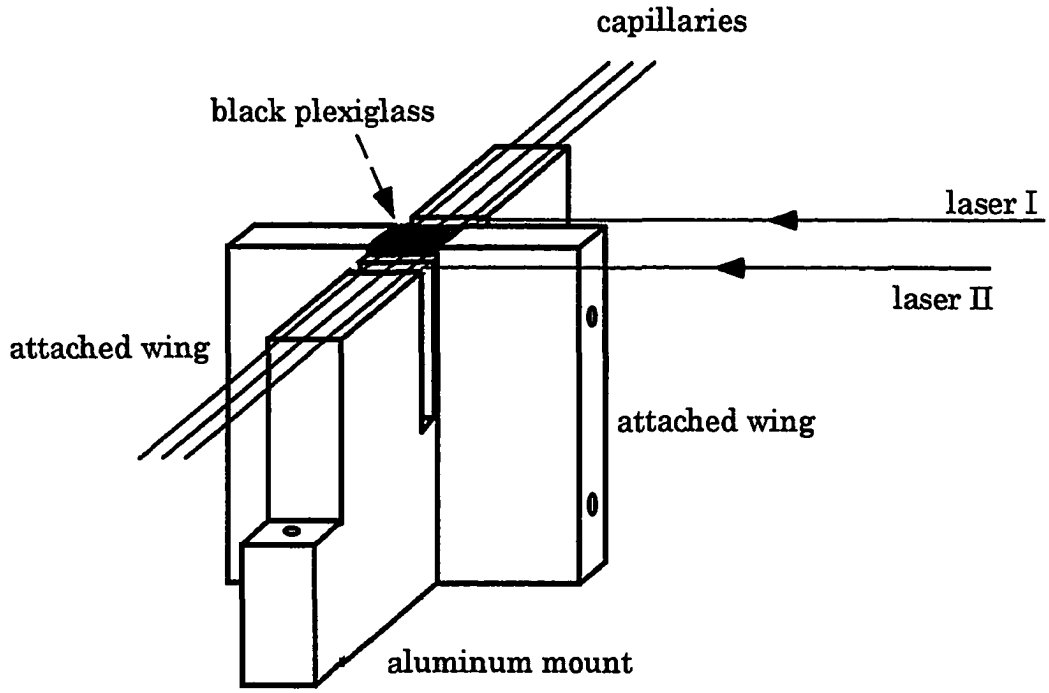


Figure 2. Capillary mount for two-window detection with two different laser beams for excitation in refractive-index-matching liquid.

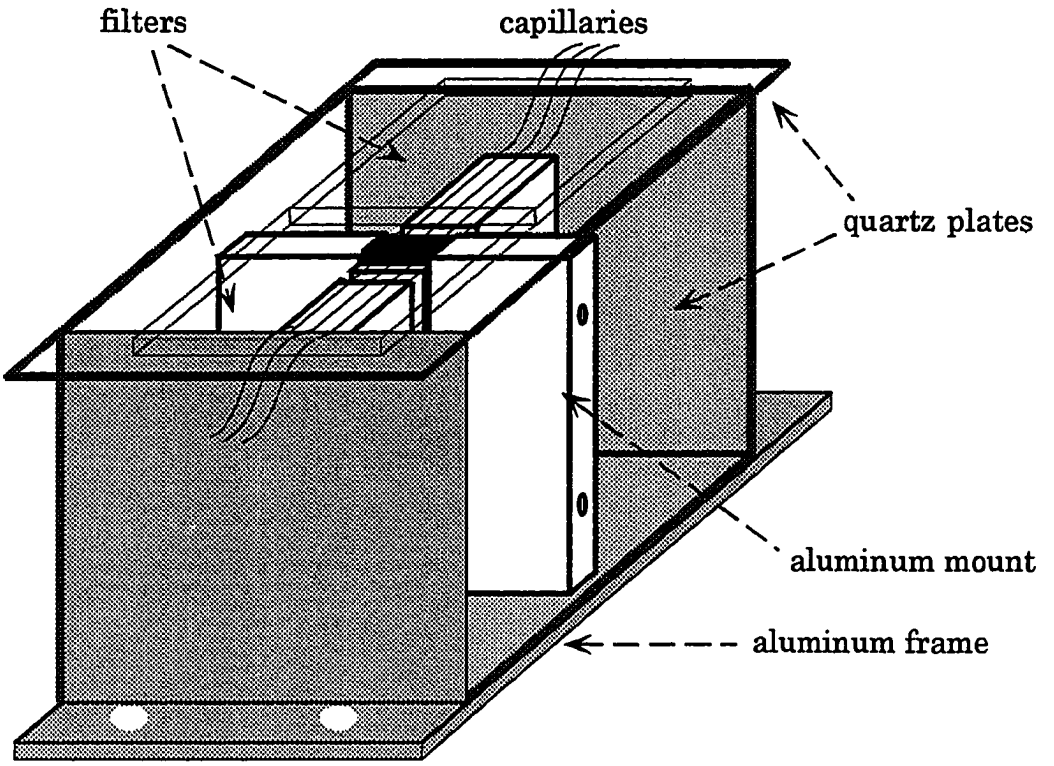


Figure 3. Liquid cell for immersing the capillary mount in Fig. 2 in refractive-index-matching liquid.

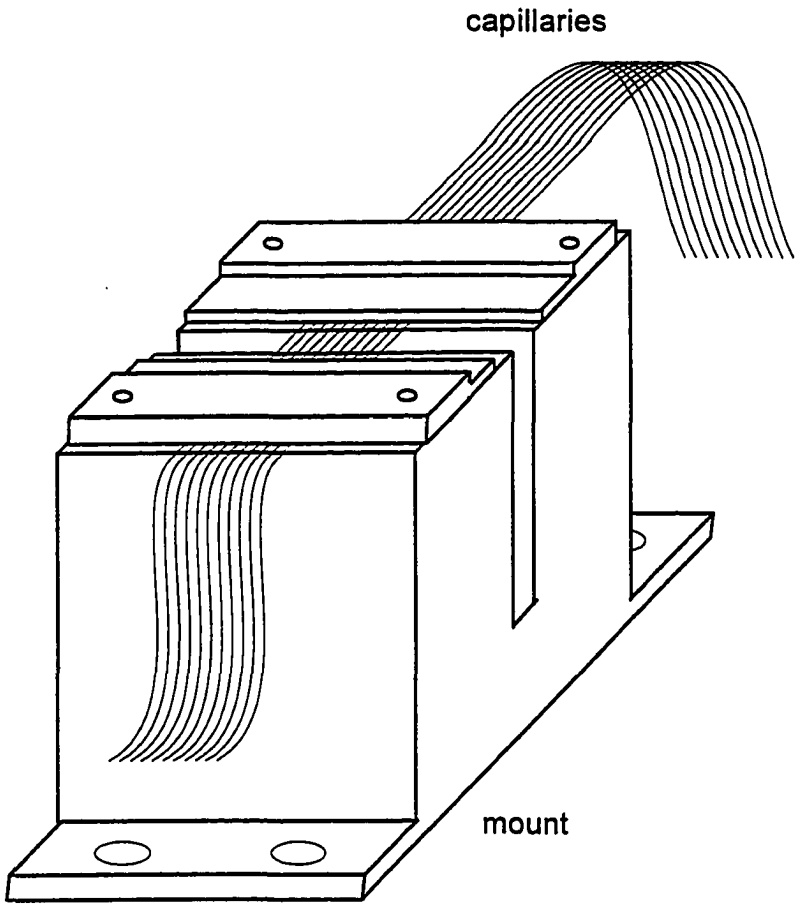


Figure 4. Capillary mount for one-laser two-color detection.

focusing field of the CID. The top of the mount is polished shiny to provide a flat surface for packing the capillaries. Two aluminum bars of 2"(L) x 1/2"(W) x 3/16" (H) with their bottom polished are used to press down the capillary array. The bars are fixed on the mount by two screws on each side so that the capillary array is secured. One third of the edges of the bars facing the laser path are thinned by half so that they do not obstruct the solid angle for fluorescence collection.

Excitation An Argon ion laser (Uniphase, San Jose, CA, Model 2213-150ML) with 488-nm and 514-nm emission or a Krypton laser (Coherent, Santa Clara, CA, Model Innova 300 FReD) operated at 568 nm is used as the excitation source. The laser beams are focused onto a capillary or a capillary array with 10-cm f.l. planar-convex lens (Oriel, Stratford, CT) or 10-cm f.l. cylindrical lens (Oriel). For excitation of a 100-capillary array, the laser beam is tilted by 4^0 relative to the capillary array plane. With a 10-cm cylindrical lens, the beam is focused into a sheet the plane of which is perpendicular to the capillaries.

Detection The SCM5000E CID system is used as the array detector. The setup and operation of the CID is similar to that in previous work (19). A wide angle camera lens (Nikon F1.4/28mm) is connected onto the CID lens mount. The 75-μm core of a capillary is focused onto one CID pixel. For detection of fluoroscein, an interference filter centered at 540 nm (Melles Griot) was used. For detection of DNA fragments, a cut-off filter is used.

RESULTS AND DISCUSSION

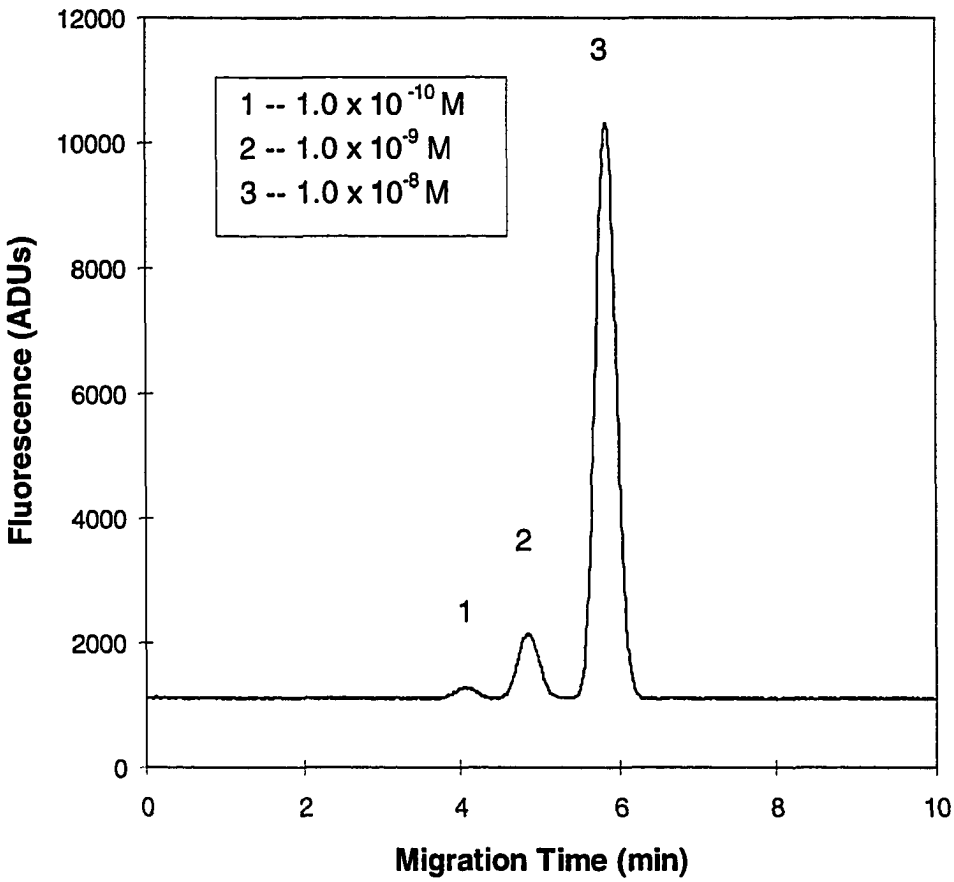
Capillaries In most CE applications, polyimide coating capillaries are used. A window is usually formed on a capillary by burning off the coating with flame or with boiling sulfuric acid. To build a 100-capillary array, burning windows on 100 capillaries one by one is a tedious process. After the coating is removed, a capillary becomes very fragile at the window. Burning off coating to create windows sometimes leaves two charred bumps on a capillary. These bumps prevent the capillaries to be packed evenly on a capillary mount, which will result in uneven distribution of excitation power among the capillaries.

It is attractive that a capillary can be used without having to remove coating to create a window. Several years ago, Yeung's group purchased transparent coated capillary for CE separation. Probably due to the poor quality of the coating of those early products, the use of these capillaries in CE separation and detection did not provide advantages over conventional capillaries. Recently, Clark *et al.* (21) made essential use of a UV-transparent capillary in a CE-laser fluorometry. According to the manufacturer of the UV-transparent capillaries (Polymicro Technologies), second generation coating materials have been used for coating the fused-silica capillaries. We ordered a batch of 150 μm o.d., 75 μm i.d. UV-transparent capillary for use in multiplexed capillary electrophoresis. Unfortunately, the coating of this batch capillary produced fluorescence comparable to that of 10^{-9} M fluorescein under 488-nm

excitation. A replacing batch was made after the manufacturer put efforts to control the purity of the coating materials. The fluorescence background of the replacing batch capillary was reduced by one order, which is marginal for DNA sequencing. Further purification of the coating material is possible but requires additional investment and thus the price will be higher. The manufacturer also experimented with two visible-transparent coating materials and sent the coated capillary samples to this lab for examination. Again, probably due to contamination during production or the difficulty of purification, these two coating materials are not better than the UV-transparent coating. The manufacturer has not put further efforts to produce purer coating materials up to this time.

At the time the experiments documented in this chapter were underway, only the first batch of the UV-transparent capillary with fluorescence background equivalent to that of 10^{-9}M fluoroscein was available. The fluorescence background of the capillary coating distributes over a wide wavelength range. By using an interference filter centered at 540nm (Melles Griot), the fluorescence background was discriminated in favor of the fluorescence of fluoroscein. A detection limit of $1x10^{-11}$ M was still achieved by using the UV-transparent capillary (Fig.5). Therefore, for running fluorescein to test crosstalk and excitation efficiency, UV-transparent capillaries were used

Figure 5. Electropherogram of detection of three different concentrations of fluorescein sample during one run. Each sample was injected at 9 kV for 20 sec and run for one minute before injecting next sample. 10 mW 488-nm beam was focused with 10-cm f.l. lens to illuminate the UV-transparent capillary. Capillary was 35-cm total, 20-cm effective length. The exposure time of the CID was 500 ms. The CID was operated at gain 150.


12000
10000
8000
6000
4000
2000
0
Fluorescence (ADUs)
1 -- 1.0 x 10^-10 M
2 -- 1.0 x 10^-9 M
3 -- 1.0 x 10^-8 M
1
2
3
0
2
4
6
8
10
Migration Time (min)

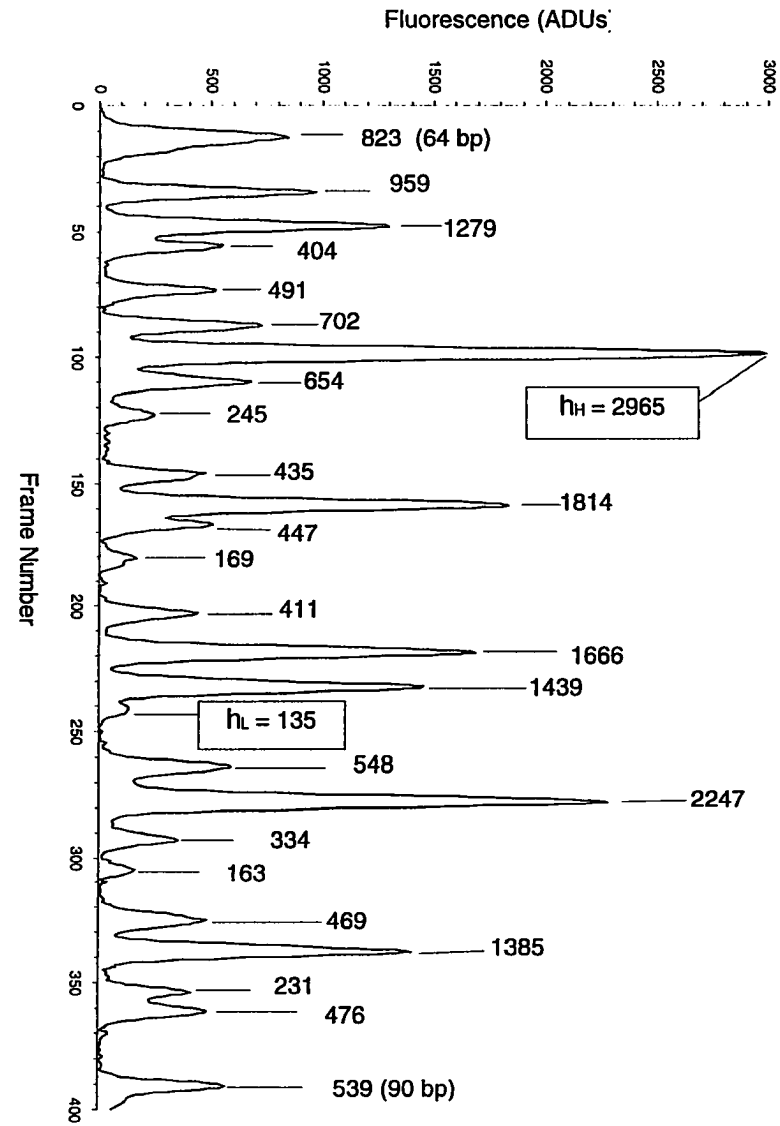
without burning off a window. For separation of DNA, a capillary with polyimide coating was used with a window created.

When a capillary array is built, one problem that should be handled with caution is that the contact medium on which capillaries are mounted affects the migration of DNA fragments. If a capillary is mounted in direct contact with an aluminum block which is grounded, migration time is 25% to 50% longer than that when the capillary is mounted on a plexiglass mount. After the aluminum block is floated by placing it on a plexiglass plate thicker than 2 cm, the retardation of migration of DNA fragments is eliminated. The exact reason of the effect has not been well understood. One explanation is that, under the influence of the electric field around the capillaries, opposite charges are induced to accumulate on the top surface of the aluminum block which is in contact with the capillaries, and the bottom surface of the aluminum block. If the aluminum block is grounded, the charges on the bottom surface will be dissipated into the ground. There is no more opposite charges to counter-balance the charges on the top surface of the aluminum block. The net charges under the capillaries produce an interfering electric field that changes the potential difference between the detection window and the injection end. The migration time is thus changed. If the aluminum block is floated, no net charges are accumulated on the block. The effect is eliminated.

Examination of crosstalk The crosstalk in multiplexed CE detection is related to reflected fluorescence from adjacent capillaries in a capillary array, and the focusing quality of a camera lens. Taylor and Yeung (10) reduced crosstalk among capillaries in a 10-capillary array down to 1% by placing a spacer between two adjacent capillaries. In the work, the capillary array was imaged onto a CCD imager through the camera extension of a binocular microscope. Placing a spacer between two capillaries is a straightforward and effective way to reduce crosstalk. But it also reduces the capillary density on an array. Lu and Yeung (17) minimized crosstalk by immersing the capillary array in refractive-index-matching liquid such as water. The crosstalk was small in water even without spacers between adjacent capillaries. This is partially ascribed to the reduction of scattered light after refractive index matching. Proper imaging geometry further reduces crosstalk. In the system, one or multiple capillaries are focused onto a CCD imager with a Canon F1.4/24 mm wide angle camera lens. Further reduction of crosstalk is achieved by only focusing the core of the capillary onto one pixel. The walls of two adjacent capillaries are focused onto the alternative pixel the signal of which is void, so that the scattered fluorescence from adjacent capillaries onto the outside wall of the capillary is spatially blotted. The principle of this focusing geometry resembles that of the spatial filter which is used in the Microfluor technology (22).

Quantitative assessment of acceptable crosstalk level for DNA sequencing is critical to orient the effort of instrumentation. The average signal strength in DNA sequencing is about one order higher than that of 1 $x10^{-10}$ M fluorescein in order to produce accurate base-calling results. Fig. 6 represents part of the electropherogram of separation and detection of a PGEM/U DNA sample. The peaks in the figure correspond to the fragments ranging from 64 bp to 90 bp. Peak heights are labeled on top of the peaks. It can be seen that the peak height difference in a PGEM/U DNA sequencing sample is large. The relative standard deviation of the peak heights is 86% in the range. The heights of the highest (H) and the lowest (L) peaks are printed in text boxes. Peak H is 22 times as large as peak L. If peak H causes 5% crosstalk in the adjacent capillaries, the crosstalk signal is comparable to the smallest peak in the same range in that capillary. The crosstalk signal could introduce an extra peak, or impose on a peak. The baseline noise is 12.5 ADUs. The S/N for peak L is thus 10. In order that the crosstalk from peak H does not affect the peaks with a similar size as peak L in adjacent capillaries, the crosstalk must be controlled under 2.5%, because 2.5% crosstalk from peak H produces a crosstalk signal in adjacent capillaries with S/N of 5, which is the marginal S/N of a detectable peak. Preferably, the crosstalk should be controlled under 1.5%, which produces a crosstalk signal of S/N of 3 in adjacent capillaries.

Figure 6. The red channel electropherogram of separation and detection of PGEM/U DNA sample using the two-color detection scheme (18). Only the fragments between 64 bp and 90 bp are shown here. Capillary was 360 μm o.d., 75 μm i.d., bare fused-silica with 60 cm total length and 50 cm effective length. The capillary was treated with 0.1 M HCl for 3 h before use. Excitation laser power was 20 mW (488 nm) at the detection window. Sample injection was at –12 kV for 20 s. Electrophoresis was run at –15 kV. Exposure time was 200 ms in the time period shown here.


Fluorescence (ADUs)
0
500
1000
1500
2000
2500
3000
Frame Number
0
50
100
150
200
250
300
350
400
823 (64 bp)
959
1279
404
491
702
654
245
h_H = 2965
435
1814
447
169
411
1666
1439
h_L = 135
548
2247
334
163
469
1385
231
476
539 (90 bp)

The following experiments were carried out to obtain a quantitative estimation of crosstalk for further consideration of optical setup.

To determine the crosstalk, a group of four 150-μm o.d., 75-μm i.d. UV-transparent capillaries were packed side by side tightly onto the capillary mount shown in Fig. 3. The capillary coating was not removed. The capillaries were illuminated with a 10-mW 488-nm beam focused with a 10-cm f.l. lens (Oriel). Fluorescence was collected through an interference filter centered at 540-nm. Cross-talk was tested in air, water and ethylene glycol respectively. The liquid cell with the capillary mount was placed on a translation stage with 5-μm resolution. The position of the capillary array was adjusted so that the liquid core of each capillary is centered at a pixel. This can be verified by looking at the pattern of the signals in the pixel corresponding to the core of a capillary and the two adjacent pixels corresponding to the capillary wall. If it shows that the center pixel has the largest signal and the two adjacent pixels have the same signal strength, the capillary is centered at a pixel. The second step is to adjust the camera lens to get the best focusing. In this experiment, the focusing ring on the camera lens was tuned to scan through a range of focusing positions around the best focusing position. Crosstalk was measured at each position. The smallest crosstalk corresponded to the best focusing. It was also found that when the camera lens aperture was reduced, crosstalk was reduced. The crosstalk was therefore tested at both aperture F1.4 and aperture F2.

Examination of crosstalk by injecting 1 x 10^{-10} M fluorescein produces the impression that there is no crosstalk between the capillaries (Fig. 7). This is misleading because the crosstalk signal from such a low concentration has been comparable to the baseline noise and can not be observed. When a higher concentration of sample is injected, for example, a DNA sequencing sample, the crosstalk may show up. Since the LOD is 1 x 10^{-11} M fluorescein in this case, to measure the cross-talk accurately, a fluorescein sample of 5 x 10^{-9} M or 1 x 10^{-8} M was injected. Fig. 8-13 show the electropherograms of detecting 5 x 10^{-9} M or 1 x 10^{-8} M fluorescein in the capillaries when the capillaries were immersed in water or ethylene glycol, or just in air with the camera lens aperture at F1.4 or F2. Crosstalk is calculated by measuring the peak height of the fluorescein signal in a selected capillary and its crosstalk signal in an adjacent capillary, as indicated in the figures. The results are summarized in Table 1.

The refractive index of fused silica is 1.46. In the order of air, water, and ethylene glycol, the refractive indices of a capillary wall and the immersing medium match better. The crosstalk thus decreases. The results also indicate that after the camera lens aperture is decreased by one stop, the crosstalk significantly decreases, especially in liquid. In air, the decrease of crosstalk due to lens aperture reduction is not as obvious as that in liquid. This indicates

Figure 7. Detection of 1 x 10^{-10} M fluorescein in four capillaries in an array in air. The camera lens aperture was set at F1.4. Other details are described in the text.

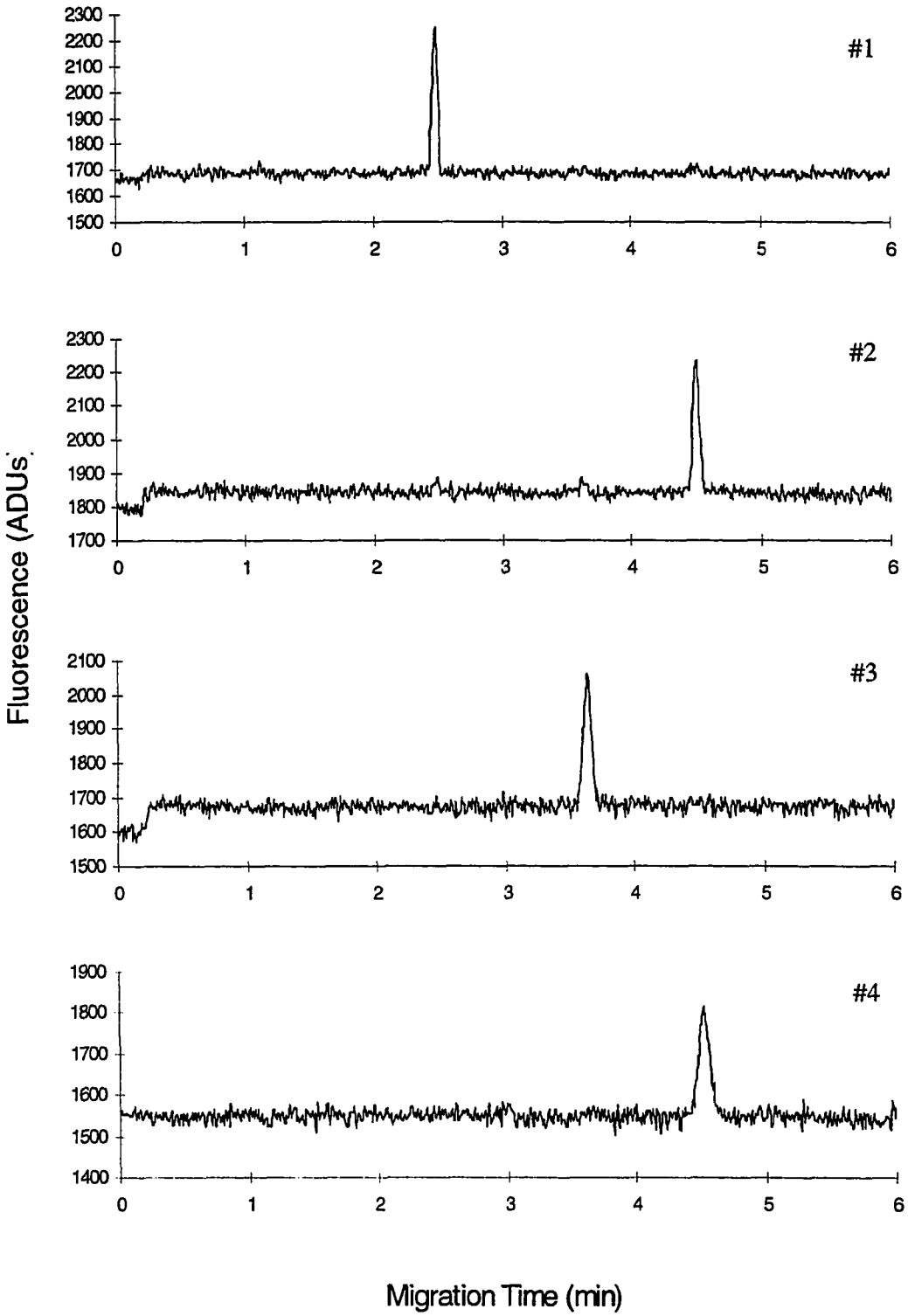

#1
#2
#3
#4
Fluorescence (ADUs)
Migration Time (min)

Figure 8. Detection of 5×10^{-9} M fluorescein in four capillaries in an array in air. The camera lens aperture was set at F1.4. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The crosstalk (H_c/H_p) is 4.9%.

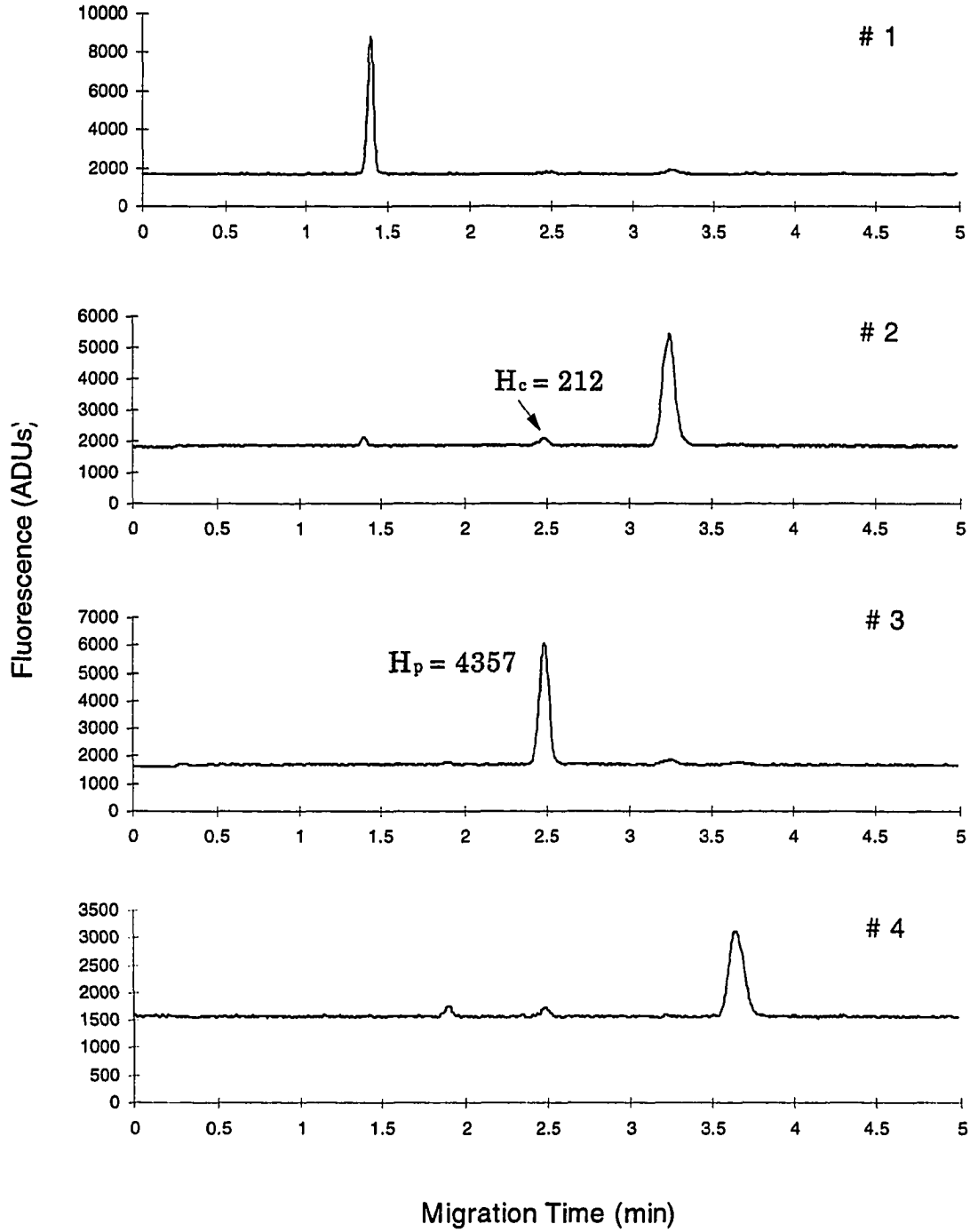

Fluorescence (ADUs)
# 1
# 2
Hc = 212
# 3
Hp = 4357
# 4
Migration Time (min)

Figure 9. Detection of 5 x 10^{-9} M fluorescein in four capillaries in an array in air. The camera lens aperture was set at F2. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The crosstalk (H_c/H_p) is 3.5%.

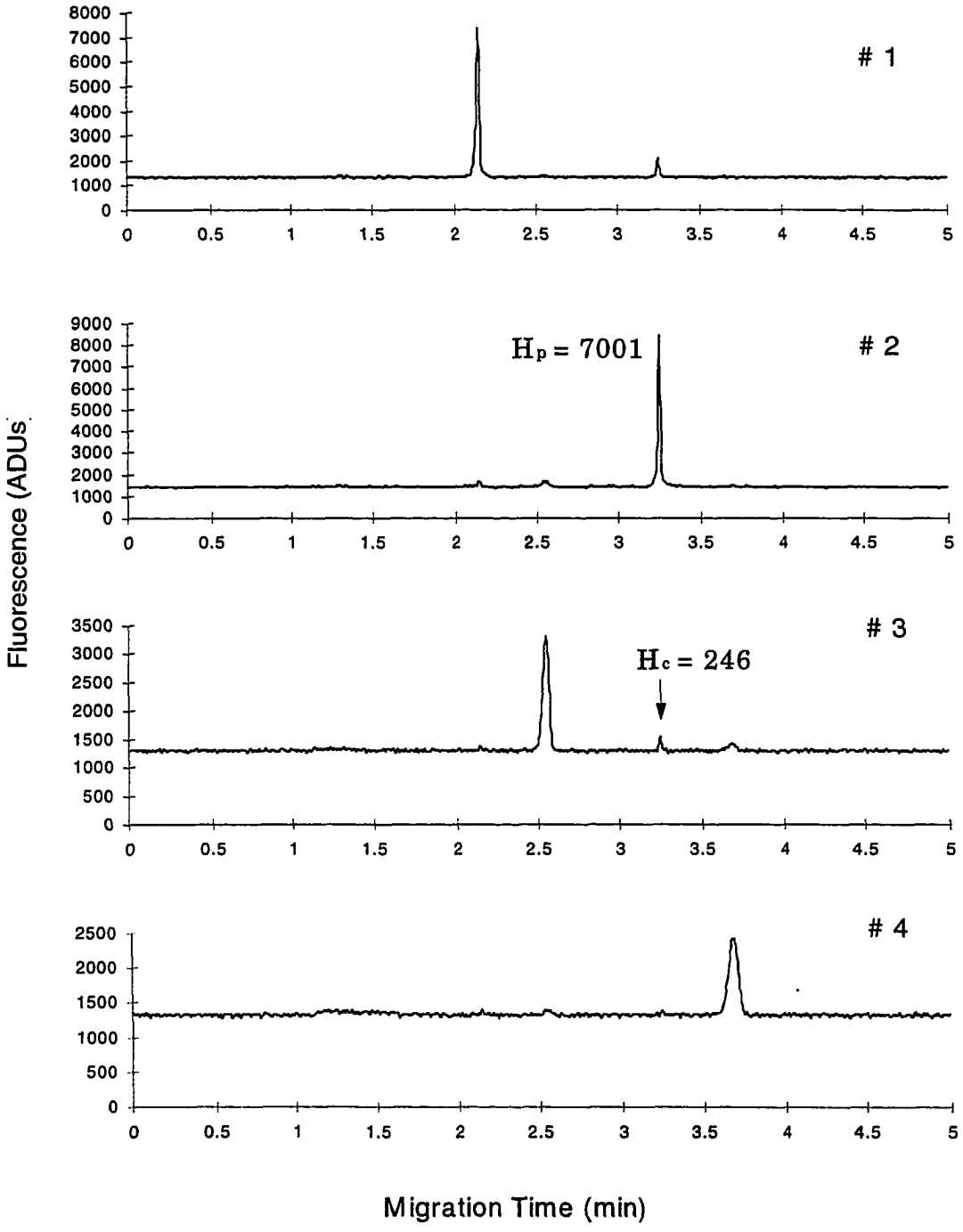

Fluorescence (ADUs)
# 1
# 2
Hp = 7001
# 3
Hc = 246
# 4
Migration Time (min)

Figure 10. Detection of 5 x 10^{-9} M fluorescein in three capillaries in an array in water. The camera lens aperture was set at F1.4. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The crosstalk (H_c/H_p) is 4.4%.

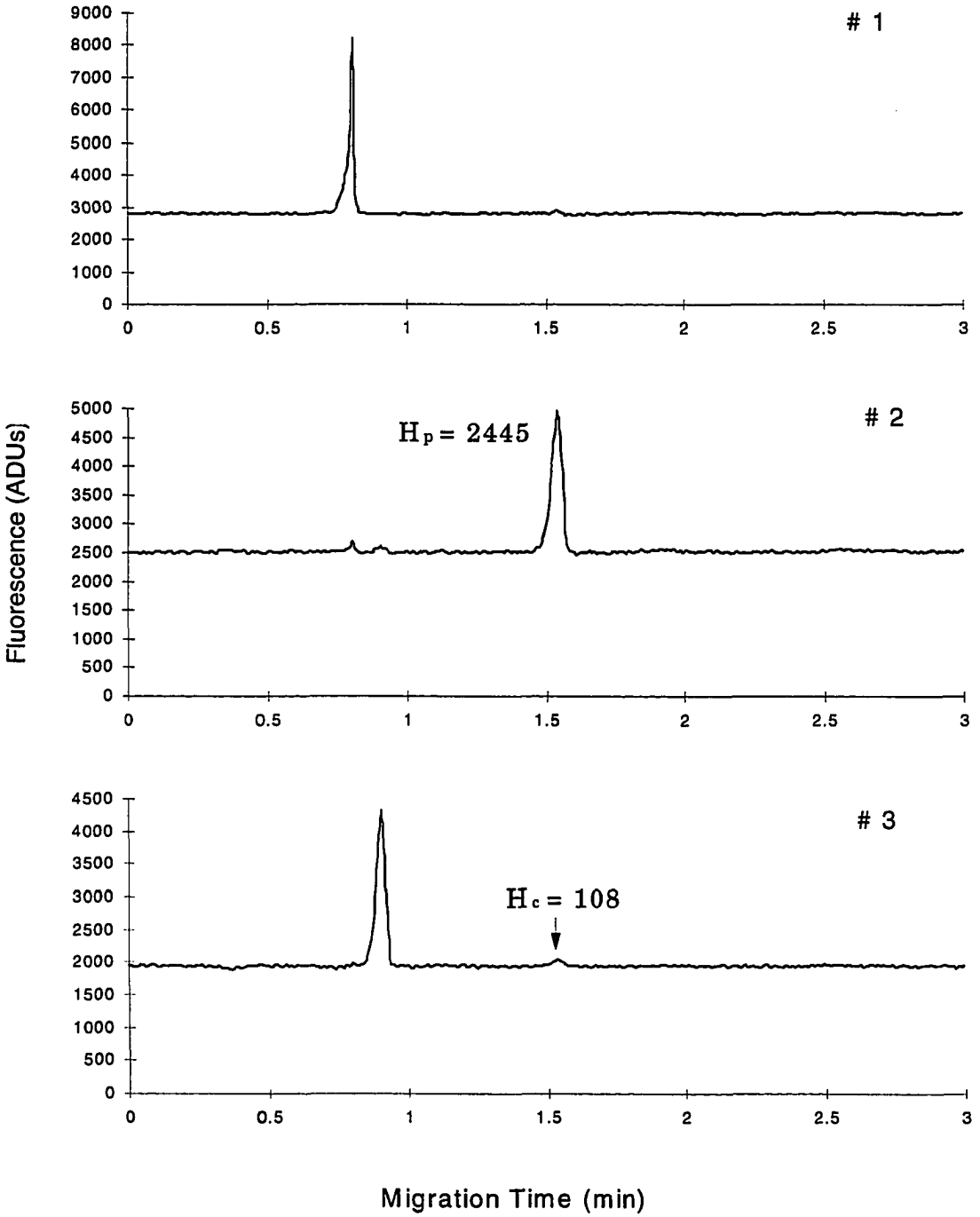

9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1
0
0.5
1
1.5
2
2.5
3
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Hp = 2445
# 2
0
0.5
1
1.5
2
2.5
3
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Hc = 108
# 3
0
0.5
1
1.5
2
2.5
3
Fluorescence (ADUs)
Migration Time (min)

Figure 11. Detection of 5 x 10^{-9} M fluorescein in three capillaries in an array in water. The camera lens aperture was set at F2. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The crosstalk (H_c/H_p) is 1.0%.

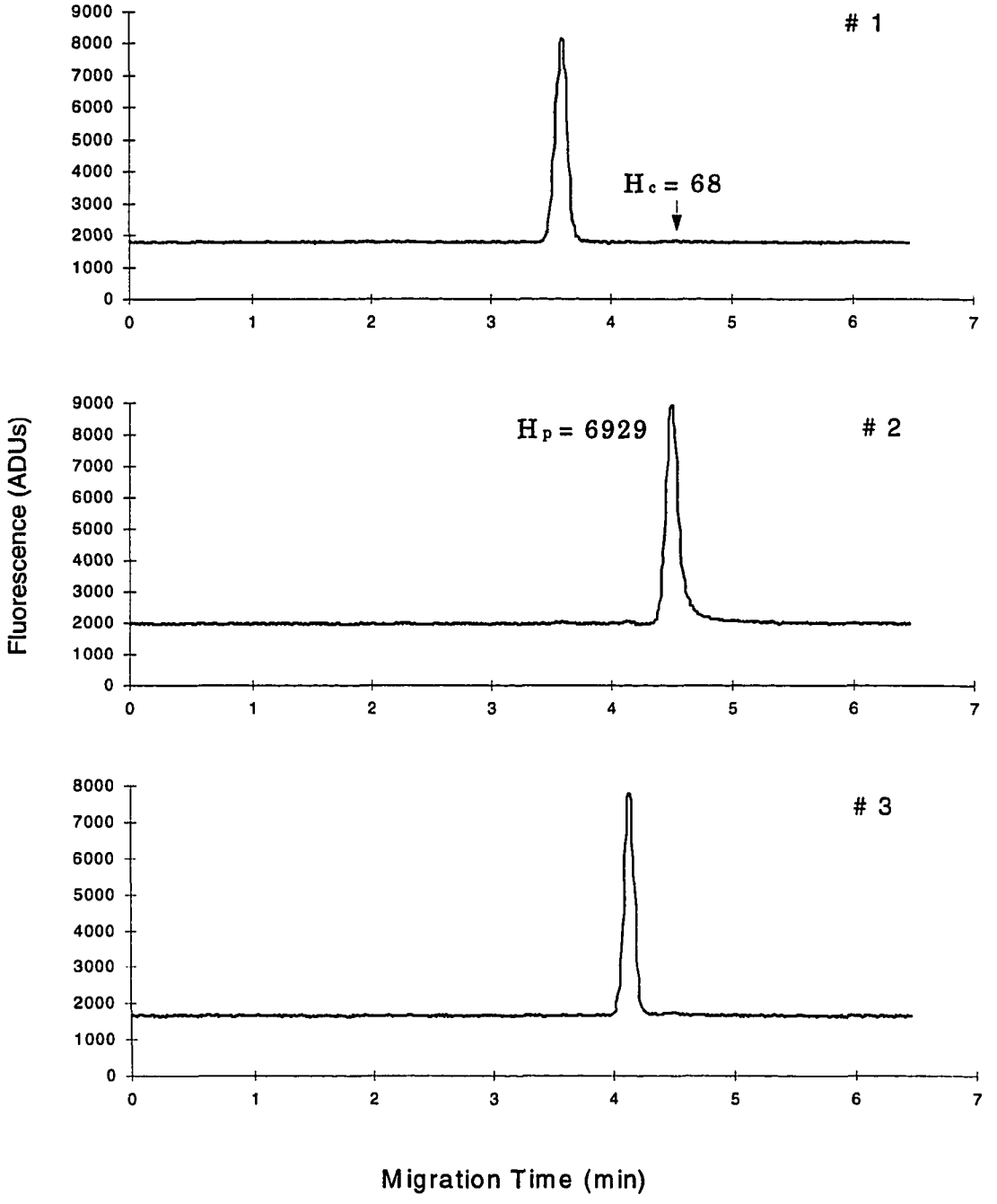

Fluorescence (ADUs)
# 1
Hc = 68
# 2
Hp = 6929
# 3
Migration Time (min)

Figure 12. Detection of 5 x 10^{-9} M fluorescein in four capillaries in an array in ethylene glycol. The camera lens aperture was set at F1.4. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The crosstalk (H_c/H_p) is 3.5%.

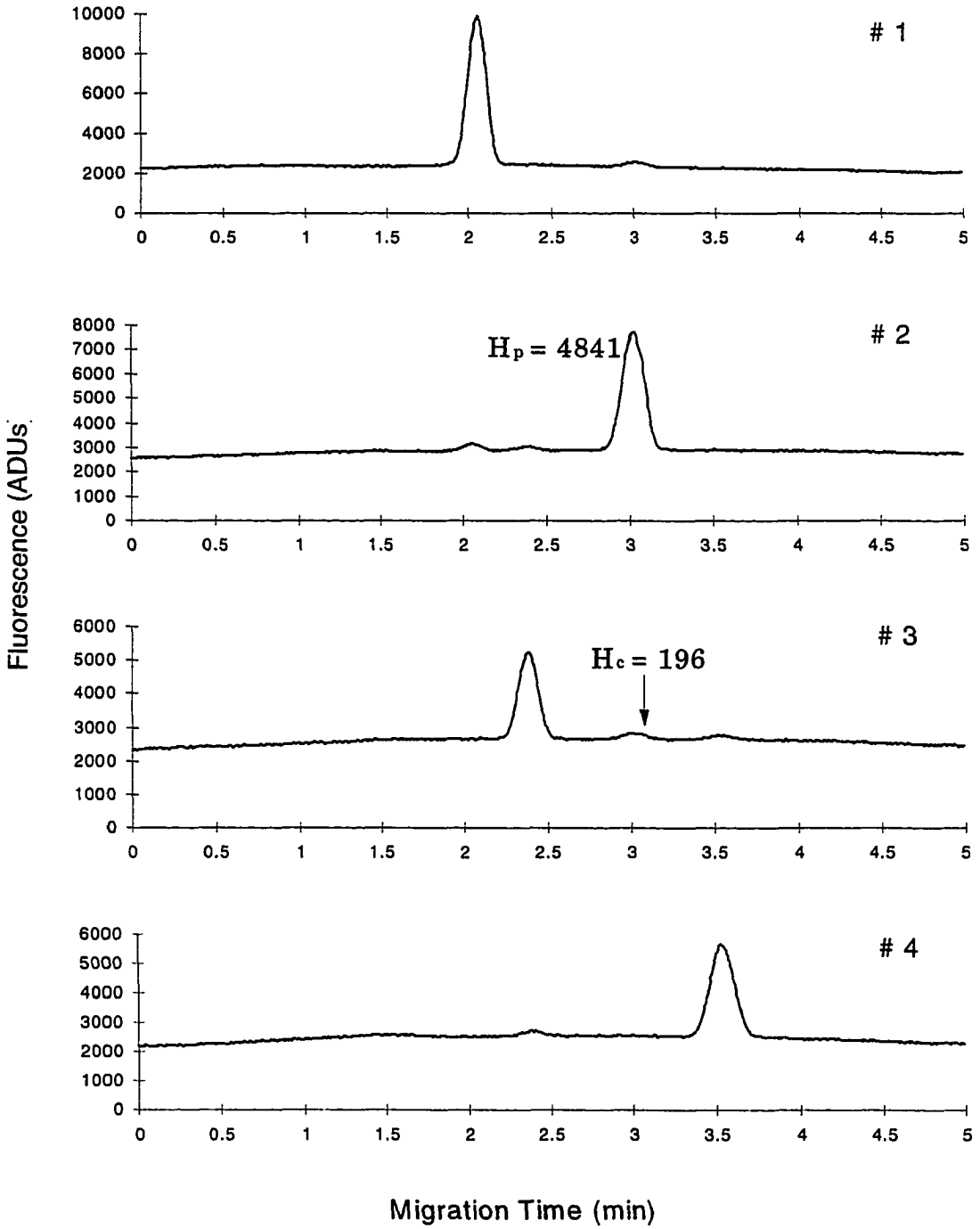

Fluorescence (ADUs)
# 1
# 2
Hp = 4841
# 3
Hc = 196
# 4
Migration Time (min)

Figure 13. Detection of 1 x 10^{-8} M fluorescein in four capillaries in the array immersed in ethylene glycol. The camera lens aperture was set at F2. H_p is the height of the fluorescein peak in a capillary, and H_c is the crosstalk signal caused by this peak in an adjacent capillary. The fluorescein signals in capillary #1 to #3 are off the 14-bit scale. The signal in #4 capillary and its crosstalk signal in #3 capillary is used to calculate the crosstalk. The crosstalk (H_c/H_p) is 1.1%.

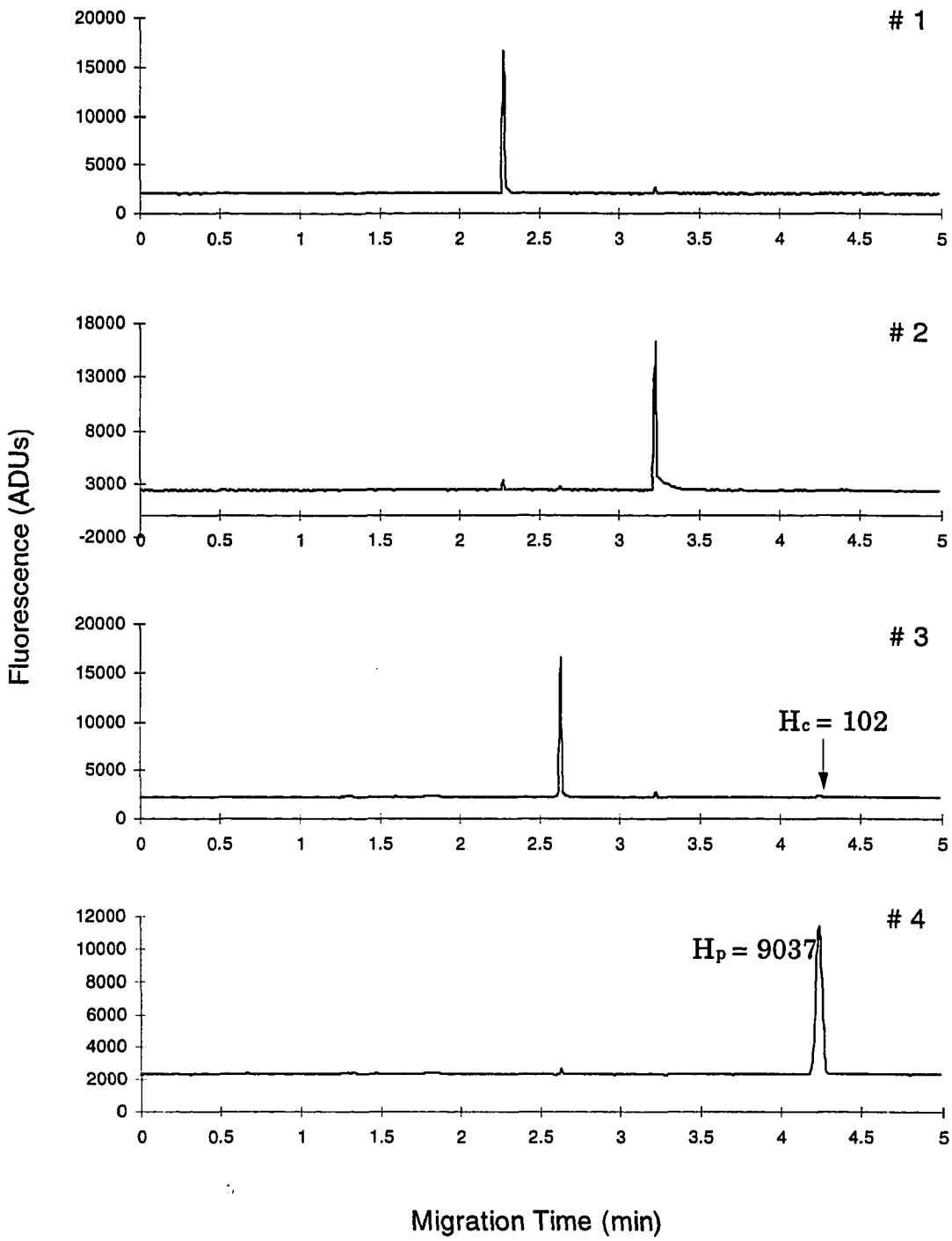

Fluorescence (ADUs)
# 1
# 2
# 3
# 4
Hc = 102
Hp = 9037
Migration Time (min)

Table 1. Crosstalk in different immersing media and at different camera lens aperture.

Camera lens aperture	Immersing media		
	Air (R = 1.0)	Water (R = 1.33)	Ethylene glycol (R = 1.41)
F1.4	4.9%	4.4%	4.1%
F2	3.8%	1.0%	1.1%

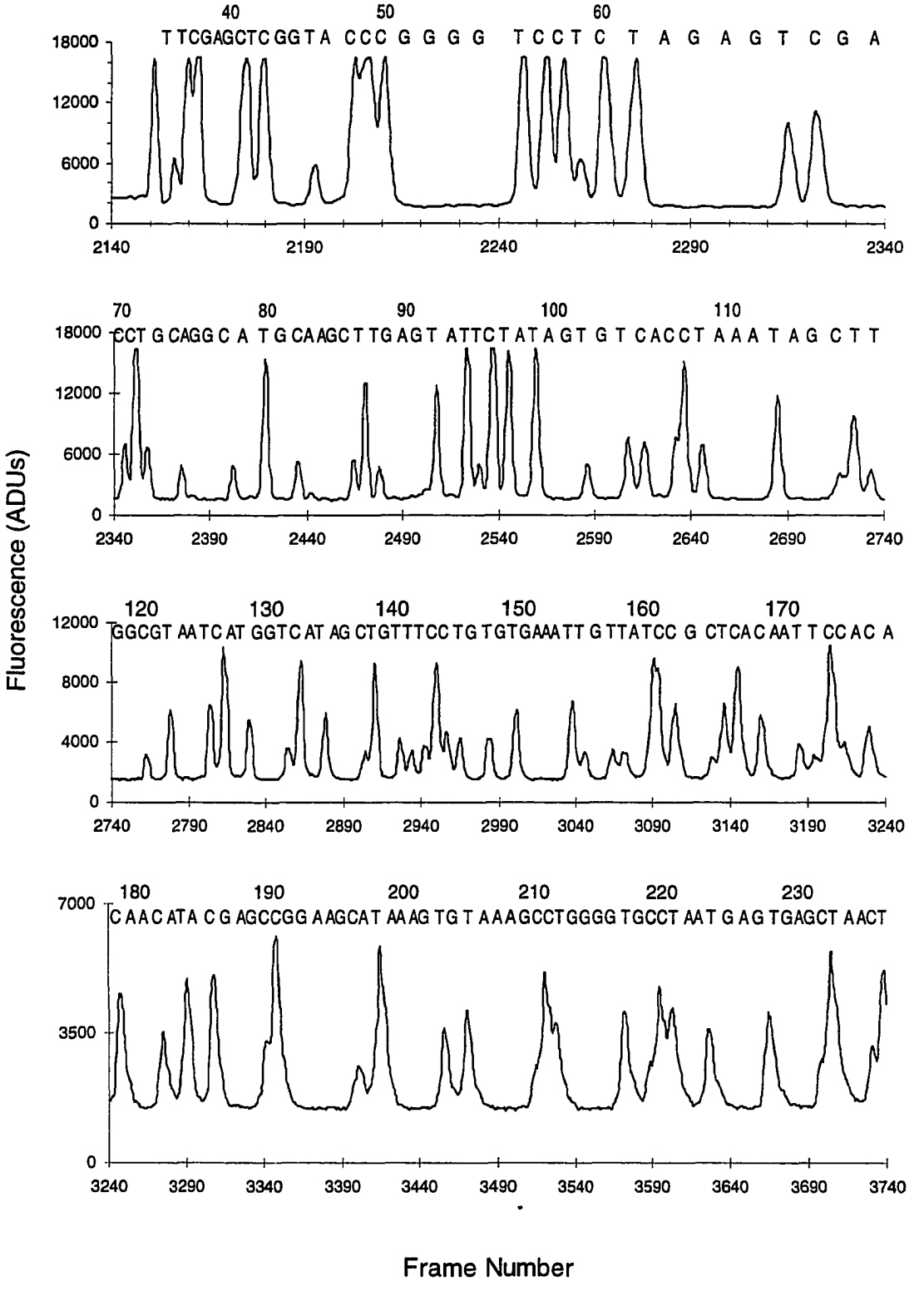
that crosstalk in air is primarily caused by reflected fluorescence between the capillary walls. Reduction of lens aperture can not eliminate this type of crosstalk. After the capillaries are immersed in liquid, the reflection of fluorescence between the capillary walls is greatly reduced. Crosstalk caused by unclear focusing becomes predominant. If the lens aperture is not decreased, immersing capillaries in refractive-index-matching liquids can not help. After the lens aperture is reduced, the field depth increases, focusing becomes sharper, and crosstalk decreases. Further decrease of lens aperture is undesirable because light collection efficiency will be lost. It is also unnecessary because the crosstalk has been reduced to such an extent that it will not cause interference. Ethylene glycol has closer refractive index to that of fused silica. Crosstalk in

ethylene glycol should be smaller than that in water. But the results indicate that the crosstalk in both cases is comparable at lens aperture F2. This suggests that other factors such as focusing have become the limiting factor. Although ethylene glycol may not help to further decrease crosstalk, it may help to incorporate laser rays into the capillary array better than water because the refraction and reflection of laser light at the interface of capillary wall and ethylene glycol is smaller.

Krypton ion laser as excitation source The Krypton ion laser with 568-nm output was tested as the excitation source for detecting DNA fragments in a sequencing sample prepared by the Sanger chemistry with ABI PRISM dye-labeled terminators.

Fig. 14 shows the elctropherogram of separation and detection of a PGEM/U DNA sequencing sample by using 568-nm line as the excitation light source. For the purpose of examination of the excitation efficiency of the 568-nm line for the four sets of fragments, only fragments within 240 bp are included in the figure. Bases are identified in the electropherogram by comparing the pattern of C and T fragments with the known sequence of the sample. Two OG590 glass filter was used for collecting fluorescence. The OG590 filters can not completely cut off 568 nm if the laser beam is shined directly onto the filter. If the capillary was illuminated in air, the scattered laser light almost saturated

Figure 14. Detection of PGEM/U sample separated in a capillary with 568-nm excitation. The capillary was immersed in ethylene glycol to reduce scattered laser light. A 20-mW (at detection window) 568-nm beam was focused with a 10-cm f.l. lens to illuminate the core of the capillary. Two OG590 glass filters were used for fluorescence collection. The capillary was 45 cm total, 35 cm effective length. Sample injection was at 9 kV for 20 sec. Electrophoresis was at 9 kV. Exposure time was 500 ms.


40 50 60
TTCGAGCTCGGTA CCC G G G G TCCTC T A G A G T C G A
18000
12000
6000
0
2140 2190 2240 2290 2340
70 80 90 100 110
CCT G CAGG C A T G CAAGC T TG AG T ATTCT AT A G T G T CACCT AAA T A G C T T
18000
12000
6000
0
2340 2390 2440 2490 2540 2590 2640 2690 2740
120 130 140 150 160 170
GGCGT AATC AT GGTC AT AG CTGTTTCC TG TGTGAAATT G TTATCC G CTCAC AAT T CC AC A
12000
8000
4000
0
2740 2790 2840 2890 2940 2990 3040 3090 3140 3190 3240
180 190 200 210 220 230
C AAC ATA C G AGCCGG AAGCAT AAAG TG T AAAGCCTGGGG TGCCT AAT G AG TGAGCT AACT
7000
3500
0
3240 3290 3340 3390 3440 3490 3540 3590 3640 3690 3740
Fluorescence (ADUs)
Frame Number

the CID detector. After the capillary was immersed in ethylene glycol, the scattered light was greatly reduced. The 568-nm line can excite C and T fragments efficiently but not for G and A. Close examination of Fig. 13 reveals that both G and A may or may not show up. Even if a G or an A shows up, the intensity is much smaller than that of a C or a T. Therefore, the electropherogram clearly shows all the peaks of C and T while the gaps or tiny peaks belong to G and A. The intensity of C and T fragments does not have a fixed ratio. This is because Taq polymerase was used to prepare the PGEM/U sample and thus the incorporation efficiency varied significantly.

In addition to the advantage of high excitation selectivity of the 568-nm line for detecting the C and T sets of DNA fragments in the PRISM-dye-prepared DNA sequencing sample, the results here indicate another advantage. Another advantage is that the scattering light from this 568-nm line is smaller due to its longer wavelength. Unlike in the situation in which the 488-nm or 514-nm line is used for excitation, the Raman scattering in water produced by the 568-nm line is shifted out of the fluorescence spectrum of C and T to longer wavelength, where the quantum yield of the CID detector is lower. Therefore, the 568-nm line causes a very low background.

Since the 568-nm line is highly selective for C an T fragments, this line together with the 488-nm line may potentially be used to construct a highly accurate base-calling system. Two detection windows need to be used, one for

each line. Since the system of one detection window with only 488-nm excitation has been found to be adequate for base-calling (18), use of the 568-nm line is not necessary for current application.

Excitation configuration Ueno and Yeung demonstrated simultaneous excitation of 100 capillaries by using an expanded gaussian beam (12). In this system, the laser beam is expanded to the size that can cover the cross width of the capillary array and then focused into a line with a cylindrical lens. The line intersects across the capillary array. The disadvantage of this excitation geometry is that each capillary only receives a very small portion of the total light output. Lu and Yeung (17) used a tightly focused beam to illuminate the capillary array with a side-entry geometry. Detection limit was improved by 100-fold with the same laser power. This side-entry excitation geometry requires much less laser power because all of the laser light passes through each capillary. However, this side-entry geometry has not been used to illuminate a 100-capillary array. The success of this excitation geometry greatly depends upon two factors. The first is that the capillary array must be packed extremely flat so that the laser beam will not be deviated by a capillary the core of which is out of the laser beam path. Secondly, the refractive indexes of the gel matrix, the capillary wall, and the surrounding medium outside of the capillaries should be matched as close as possible, so that the laser beam is not deviated by refraction at the multiple interfaces between the gel, the capillary

wall, and the surrounding medium. The two prerequisites are difficult to maintain in the case of DNA sequencing where a gel is used. Matching the refractive indexes very well among the gel and the capillary wall is difficult, because the gel composition is not easy to change for better matching of refractive indices without affecting separation efficiency. For example, it has been observed that when a gel containing several molar urea is filled into a capillary, the transmitted beam image is different from that when the gel does not contain urea. Because the capillary array needs to be packed manually at present stage, it is hard to achieve good uniformity across the whole 100-capillary array. If one capillary is out of the plane of the capillary array, this capillary will not be excited efficiently. If the refractive indices do not match, the laser beam will be deformed or deviated. Other capillaries after this one will not receive the same amount of laser light as previous ones.

To use laser output more efficiently without having to struggle with the stringent prerequisites, a compromised excitation configuration between the above two excitation geometry is investigated in this work. Specifically, a laser beam is focused into a sheet by using a cylindrical lens. The laser beam is tilted to certain degree in order that the tilted laser beam intersects with all the capillaries in the array. Then all of the capillaries receive a fresh part of light coming directly from the beam while receiving the light transmitted through previous capillaries. The light going through a capillary can be partially re-used

by following capillaries until the light goes out of the capillary array. The efficiency of use of laser output is improved. This is shown schematically in Fig. 15. Given a tilting angle θ, the size d that the laser beam needs to be expanded is calculated simply by

$$d = w \sin\theta,$$

where w is the width of the capillary array. If the refractive indices match, the fraction of light transmitting through a capillary that can be received by the next capillary depends only upon the tilting angle θ.

$$f = 1 - \sin\theta$$

It can be seen that when $\theta = 90^0$, the laser light passing through a capillary is totally missed by adjacent capillaries (12). When $\theta = 0^0$, all of the light passing through a capillary is received by next capillary (17).

Such a calculation is simplified. But it provides a rough estimation about how much a laser beam should be expanded and tilted to illuminate all of the capillaries while the maximum use of the laser output is maintained.

Experiments were carried out to test the excitation configuration. The Kr ion laser operated at 568-nm was used as the excitation source for detecting DNA samples. The size of the 568-nm beam was 1-mm. With a 10-cm f.l. cylindrical lens, the beam was focused to a 1-mm wide, about 75-μm thin sheet

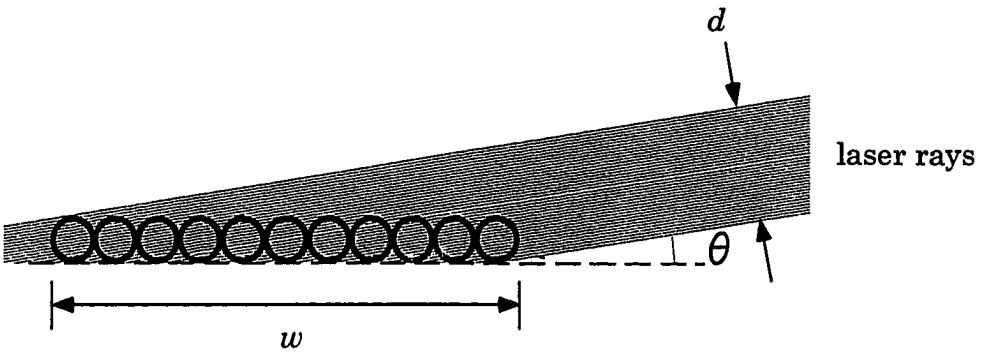


Figure 15. Diagrammatic illustration of the excitation configuration with a tilted laser beam. w -- the width of the capillary array. d -- the width of the laser beam. θ -- the tilting angle of the laser beam relative to the capillary array plane.

in a range of 2 cm centered at the focal point after the cylindrical lens. Fig. 16 shows the electropherogram obtained by exciting a single capillary with this sheet of laser beam. The sheet was shined perpendicularly across the capillary. The capillary was placed at the center of the beam. The laser power output was 240 mW. It is clear that the laser power incorporated into the capillary in this excitation mode was sufficient to detect the DNA fragments. The regular spikes on the baseline was due to the malfunction of the time base in the CID system. This problem was found later and avoided by using the host computer CPU clock for exposure time delay instead of using the time base in the CID controller. The spikes does not affect the evaluation of the excitation geometry here because we know how much the baseline noise should be based on previous normal results.

To test the feasibility of using this excitation geometry, three polyimide coating capillaries were setup on the mount shown in Fig. 3. A detection window was created on each capillary. The three capillaries were spaced evenly. The distance between the first and the third capillaries was 1.5 cm. This is the size of a 100-capillary array. The capillaries were immersed in ethylene glycol.

Fig. 17 and Fig. 18 represent the separation and detection of PGEM/U DNA sample simultaneously in the three capillaries with the laser beam tilted at 0^0 and 4^0 respectively. The DNA sample was injected into the three capillaries simultaneously from the same sample tube. Difference of injected amount among the capillaries was minimized, although there might be still

Figure 16. Detection of PGEM/U sample in one capillary with 568-nm excitation. The 240-mW 568-nm beam was focused with a 10-cm f.l. cylindrical lens into a 75 μm x 1 mm sheet. The sheet was shined perpendicularly across the capillary. The capillary was illuminated with the center part of the beam. The capillary was 45 cm total, 35 cm effective length. Sample injection was at 9 kV for 20 sec. Electrophoresis was at 9 kV.

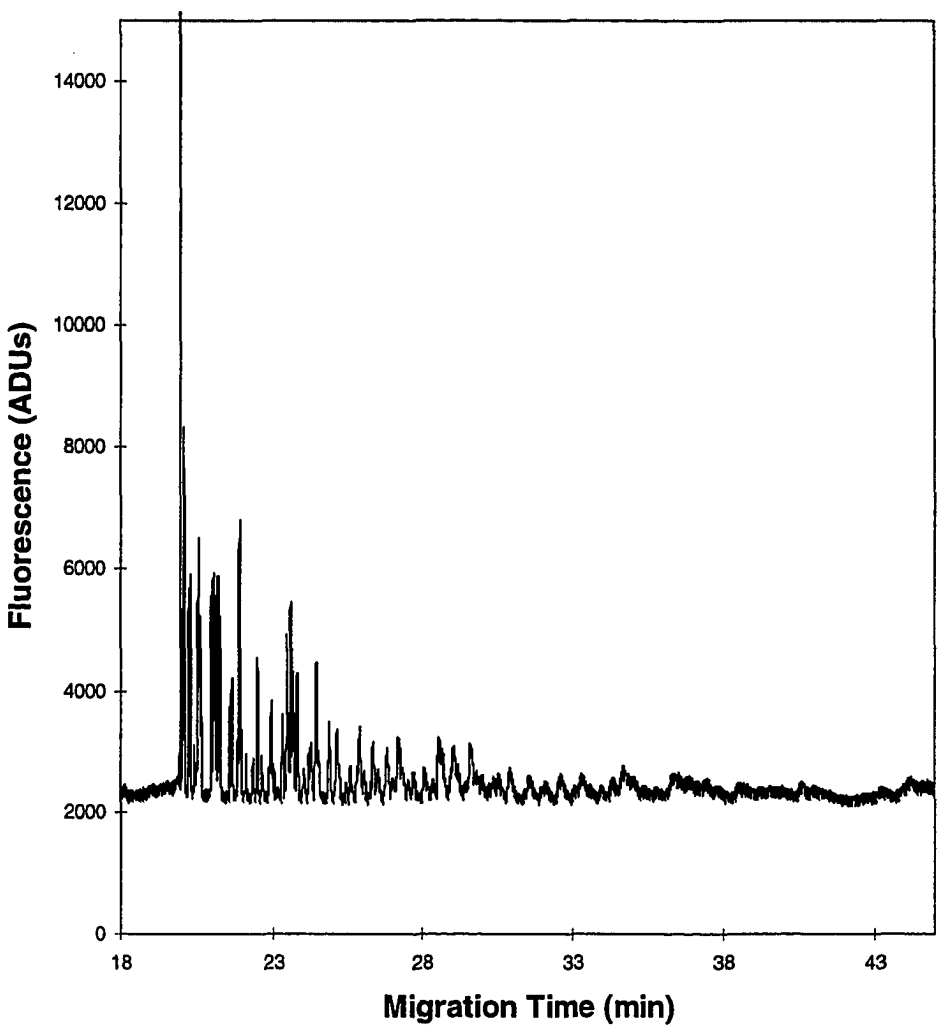

Fluorescence (ADUs)
14000
12000
10000
8000
6000
4000
2000
0
18
23
28
33
38
43
Migration Time (min)

Figure 17. Detection of PGEM/U sample in the three capillaries in an array with 568-nm excitation. The 240-mW 568-nm beam was focused with a 10-cm f.l. cylindrical lens into a 75 μm x 1 mm sheet. The beam was kept parallel to the capillary array plane. The sheet was shined perpendicularly across the capillaries. Capillary lengths, sample injection, and electrophoresis were the same as in Fig. 16.

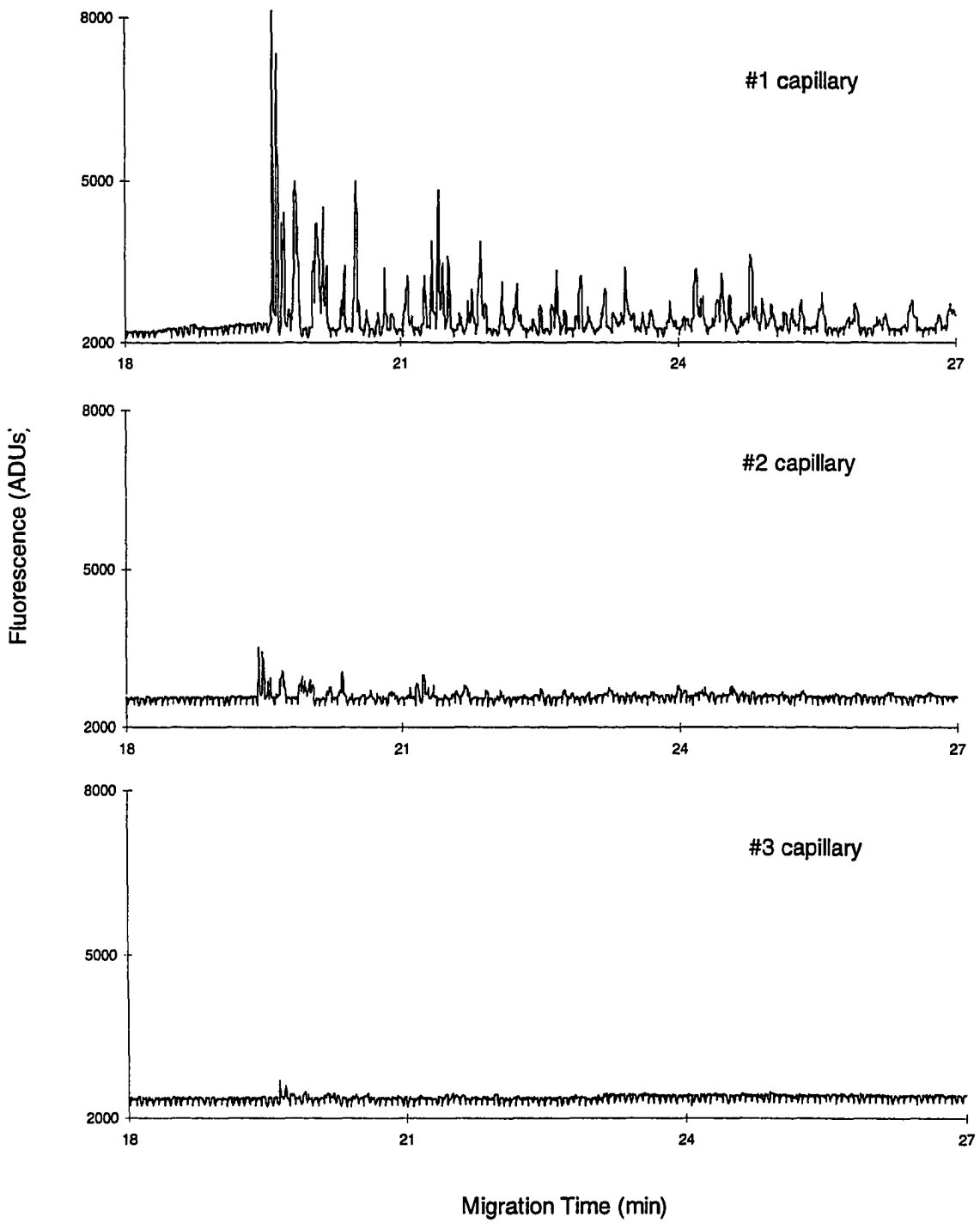
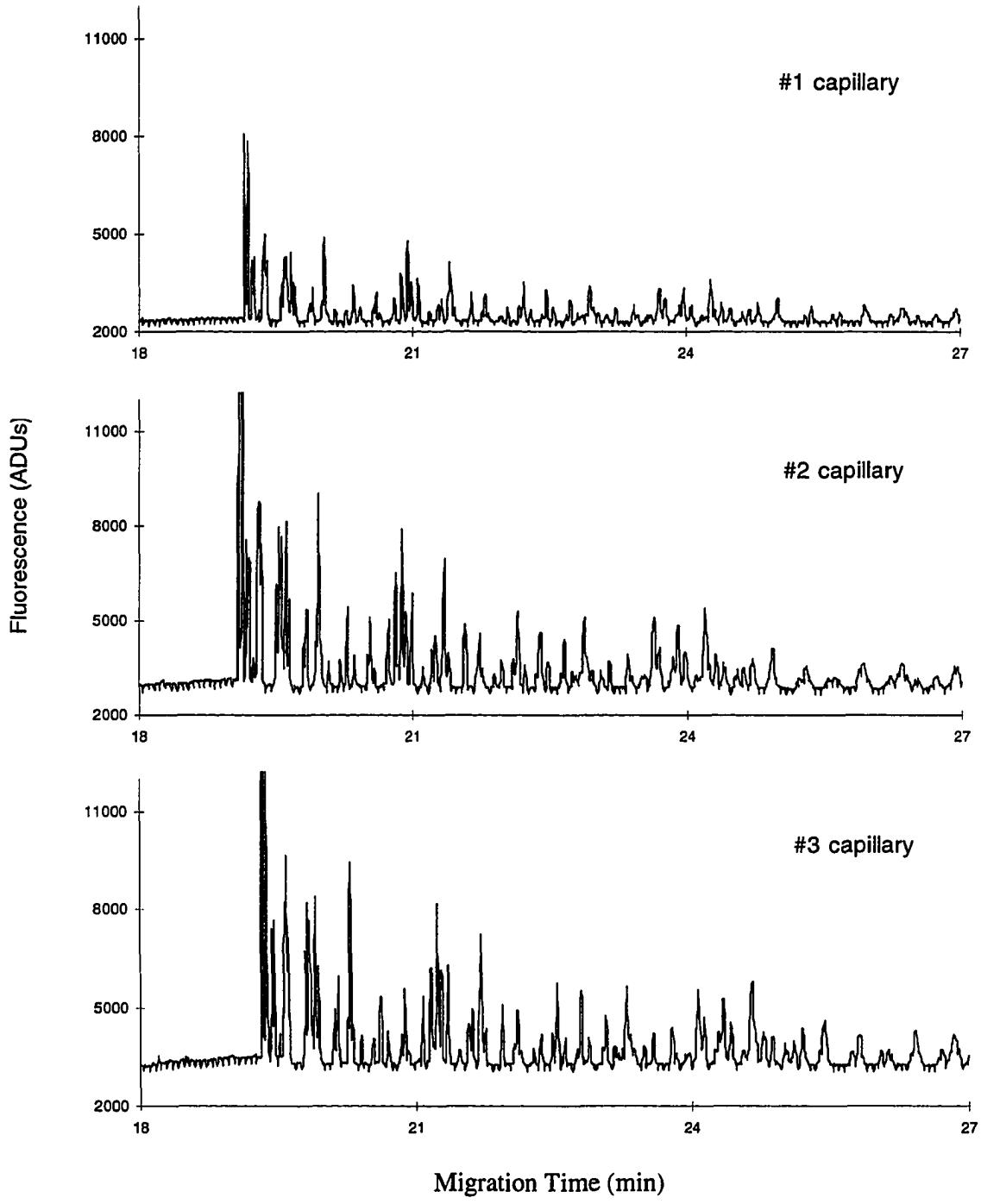

8000
5000
2000
#1 capillary
18
21
24
27
8000
5000
2000
#2 capillary
18
21
24
27
8000
5000
2000
#3 capillary
18
21
24
27
Fluorescence (ADUs)
Migration Time (min)

Figure 18. Detection of PGEM/U sample in the three capillaries in an array with 568-nm excitation. The 240-mW 568-nm beam was focused with a 10-cm f.l. cylindrical lens into a 75 μm x 1 mm sheet. The beam was tilted at 4^0 relative to the capillary array plane. The sheet was shined perpendicularly across the capillaries. Capillary lengths, sample injection, and electrophoresis were the same as in Fig. 16.


#1 capillary
#2 capillary
#3 capillary
11000
8000
5000
2000
18
21
24
27
Fluorescence (ADUs)
Migration Time (min)

about 10% variation which is normal in CE. But this variation does not affect the comparison of excitation configuration because the variation due to different beam-tilting angle is much larger. At 0^0, the first capillary deviated the light. The signal in the second and third capillaries was much weaker. Although the capillaries had been immersed in ethylene glycol for refractive index matching, because there was still some difference of refractive index between the gel, the capillary wall, and ethylene glycol, the rays passing through the first capillary were deviated and could not pass through the cores of the second and the third capillaries. After the beam was tilted at 4^0, another two capillaries were excited as efficiently as the first capillary. This clearly indicates that each capillary receives a fresh portion of light from the laser beam. The first capillary has about 40% weaker signal strength than the second and the third capillaries. This is probably because the first capillary was at the edge of the gaussian beam. Another part of reason is that unlike the second and the third capillaries, the first capillary does not receive extra light that transmits through previous capillaries. With the 240-mW laser power, the core of each capillary is estimated to receive about 1-mW laser power. Since the excitation efficiency of the 568-nm line for C and T fragments is high, the signal strength is impressive although the amount of laser incorporated into each capillary is merely 1-mW. The S/N was affected by the spikes caused by malfunction of the time base in the CID at that time. If the absolute signal strength in the figures is compared with

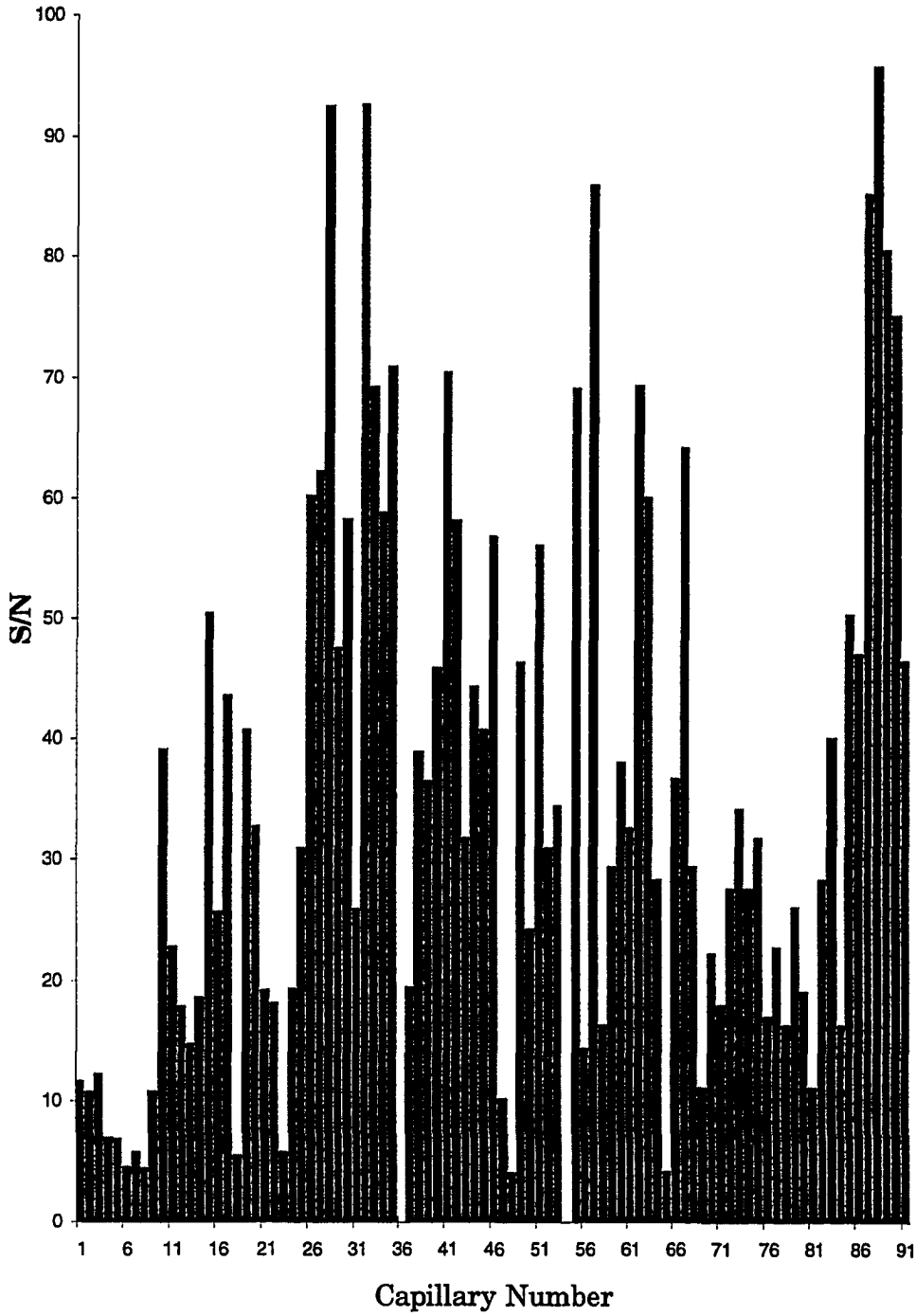
the normal baseline noise, which is usually about 15 ADUs, it can be seen that S/N ratio has been sufficient for quantitatively measuring each peak. If the signal strength in Fig. 17 and Fig. 18 is compared with that in Fig. 16, they are comparable.

Further experiment with this Kr ion laser was not available because the laser was returned to the manufacturer after the testing period was due.

The excitation geometry was further investigated with an actual 100-capillary array by using a 488-nm beam for excitation. The excitation and detection were carried out in air. To examine the evenness of excitation among the capillaries, fluorescein was used instead of DNA samples. The capillaries used in the 100-capillary array were UV-transparent capillaries without removing coating to create detection windows. The capillaries were set up on the mount shown in Fig. 4 that was made for one-beam-excitation two-spectral-channel detection system. After packing, the detection region of the capillaries was washed carefully with methanol and acetone to remove grease and particles. A 20-mW 488-nm line was focused with a 10-cm f.l. cylindrical lens and tilted at 4^0 to illuminate the capillary array. A $1x10^{-8}$ M fluorescein solution was injected into the capillaries simultaneously by dipping the injection end of the capillaries into the same sample vial and applying 6 kV for 20 sec. The capillaries were put back into the buffer vial for electrophoresis at 6kV. The exposure time was at 500 ms. The peak in each capillary was integrated and the S/N of the peak was

plotted vs. the capillary number (Fig. 19). Only peaks from the first 91 capillaries are included in the figure. The remaining 9 capillaries were missed by the detector because the subarray in the program was not set large enough to include the overhead at both sides of the image of the capillary array at that time. From Fig. 19, it can be seen that basically all the capillaries received certain amount of light. The average S/N is 35. The variation is 68%. The variation is unlikely only due to injection variation which is usually within 10%. The variation is largely due to the uneven distribution of laser power over the capillary array. Since the beam is tilted at 4^0, if a capillary is 10 μm higher than adjacent capillaries, it intersects the rays which would otherwise directly pass through the next capillary. In air, because of the difference of the refractive indices of the capillary and the air, the intersected rays are deviated. The next capillary receives less light. Quantitative estimation of this effect is complicated. But if the indices are matched, the situation is simplified. Even if the outstanding capillary intersects the rays that should go directly through other capillaries, the intersection does not deviate the rays. In this case, the unevenness of the capillaries will cause much less problem than in the case of air. The requirement of capillary packing will not be that stringent. Previously, when both 488-nm line and 514-nm line were used for detection in the two-window two-laser configuration, the scattered 514-nm light propagating in liquid caused high background in the 488-nm channel. Leaving the setup in air greatly

Figure 19. Excitation power distribution among the capillaries in a 100-capillary array in air. A 20-mW 488-nm beam was focused with a 10-cm f.l. cylindrical lens into a 75 μm x 1 mm sheet. The beam was tilted at 4^0 relative to the capillary array plane. The sheet was shined perpendicularly across the capillaries. The signal to noise ratio (S/N) of a fluorescein peak in a capillary indicates the excitation power incorporated into the capillary.


S/N
100
90
80
70
60
50
40
30
20
10
0
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
Capillary Number

reduced this interference because Raleigh and Raman scattering in air is much weaker than that in a liquid. Later optimization of the system leads to a simpler setup in which only the 488-nm line is used (18). It is possible again to immerse the capillary array in liquid to achieve better light coupling. Although Raman scattering in liquid is higher than in air, more laser power incorporated into capillaries, reduction of scattered laser light, and minimized crosstalk will provide more advantages to outweigh the disadvantage.

REFERENCES

1. (Joint U.S. Department of Energy and U.S. Department of Health and Human Services Report DOE/ER-0452P. Understanding Our Genetic Inheritance--The U.S. Human Genome Project: The Firt Five Years; Washington, DC, April 1990.
2. Smith, L. M.; Sanders, J. Z.; Kaizer, R. J.; Hughs, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. Nature 1986, 321, 674-679.
3. J. M. Prober, G. L. Trainer, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jemsen and K. Baumeister, Science **238**, 336 (1987).
4. W. Ansorge, B. S. Sproat, J. Stegemann, C. J. Schwager, Biochem. Biophys. Methods, 13, 315-323 (1986).

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

6. L. M. Smith, Nature, 349, 812-813 (1991).

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

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

9. H. Kambaraand and S. Takahashi, Nature, 361, 565-566 (1993).

10. J. A. Taylor and E. S. Yeung, Anal. Chem., 65, 956-960(1993).

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

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

13. D. C. Nguyen, R. A. Keller, J. H. Jett, and J. C. Martin, Anal. Chem., 59, 2158 (1987).

14. J. Zhao, D. Chen, and N. J. Dovichi, J. Chromatogr., 608, 117 (1992).

15. E. N. Fung and E. S. Yeung, Anal. Chem., 67, 1913 (1995).

16. H. T. Chang and E. S. Yeung, J. Chromatogr., in press.

17. X. Lu and E. S. Yeung, Appl. Spectrosc., 49, 605 (1995).

18. Q. Li and E. S. Yeung, Appl. Spectrosc., accepted.

19. Q. Li and E. S. Yeung, Appl. Spectrosc., 49, 825 (1995).

20. S. Hjerten, J. Chromatogr., 347, 191 (1985).

21. B. K. Clark, T. Vo-Dinh, and M. J. Sepaniak, Anal. Chem., 67, 680-683 (1995).

22. E. S. Yeung, P. Wang, W. Li, and R. W. Giese, J. Chromatogr., 559, 183 (1991).

CHAPTER 5. SIMPLE TWO-COLOR BASE-CALLING SCHEMES FOR DNA SEQUENCING BASED ON STANDARD 4-LABEL SANGER CHEMISTRY

A paper accepted by Applied Spectroscopy

Qingbo Li and Edward S. Yeung*

ABSTRACT

Two base-calling schemes for DNA sequencing are evaluated. Both are based on data collected from two broad-band emission channels derived from either one- or two-excitation channels. Standard 4-dye Sanger reaction products are used in conjunction with capillary electrophoretic separation in a polymer matrix. Data acquisition is compatible with high light-throughput imaging and minimal data storage. In one scheme, commercial chromatographic software provides peak recognition and peak heights. The peak-height ratios from the two channels provide base-calling accuracies of 99.3% and 97.1% through 330 bp and 350 bp, respectively. In another scheme, ratiograms are derived from the two channels. The resulting step-like functions permit calling bases even when successive peaks are not resolved. The base-calling accuracy is 99% through 340 bp. Because of simplicity in implementation, either scheme should be readily applicable to high-speed high-throughput DNA sequencing in capillary

arrays.

INTRODUCTION

Since the original demonstration of DNA sequencing[1] based on fluorescence labeling of the fragments from the Sanger reaction,[2] many variations have been developed. It is unlikely that CE will provide migration times that are reproducible enough among a group of capillaries to allow running the 4 set of fragments in separate capillaries.[3,4] So, one needs a color scheme to sort out the 4 bases run on one column. Because of difficulties in controlling the polymerase reaction and in detection signal-to-noise ratio (S/N), we consider the one-color 4-intensity scheme[5,6] to be least desirable at the present time.[7] The two-color two-intensity scheme is a hybrid between the above two methods.[5,8,9] It retains the advantage of a simpler optical arrangement, better light collection, and straightforward algorithm. However, uniform incorporation by the polymerase must also be assumed. So, the most robust technology is still the original 4-label scheme.[2] Furthermore, the initial demonstration of any alternative DNA sequencer should be based on dyes and protocols already familiar to biochemists (i.e. preferably the commercial protocols) so that side-by-side comparison and validation is possible without putting additional demands on the polymerase reaction.

Even with 4 dye labels, many optical schemes have been developed for calling bases. It should be noted that the 4 standard dyes (FAM, JOE, ROX, and TAMRA) are by no means spectrally distinct, either in excitation or in emission.[2] The present commercial instrument uses fairly narrow interference filters for emission and two laser wavelengths for excitation. Still, a complicated set of ratios have to be employed for base calling.

Recently, there is much interest in high-speed DNA sequencing based on capillary electrophoresis (CE), particularly when high-throughput is also achieved through multiplexing in a capillary array.[3,9-13] Simultaneous measurements in multiple capillaries put extreme demands on sensitivity, data storage and data analysis, even with the recent rapid advances in detectors and computer hardware. Therefore, any new base-calling scheme must also allow good light collection, be simple to implement as related to data acquisition, and must not require extensive calculations. A double-rotating-filter system analogous to the commercial hardware has been demonstrated.[7] Alternatively, the two laser beams can be made to be collinear and wedged prisms can be used to obtain 4 distinct images on an array detector.[10] A single laser can also be used, with minimal compromise in sensitivity.[6,14] Others have dispersed the fluorescence with a monochromator prior to array detection with one[15] or two[16] excitation wavelengths. Monochromator-based spectral identification of the labels in principle offers the best selectivity. In fact, on-the-fly spectra have been

acquired in CE with excellent sensitivity.[17,18] However, one needs to disperse the total fluorescence over many pixels to obtain a spectrum. This adds to the amount of raw data acquired and increases the acquisition time and the data work-up effort. Monochromators also do not have the favorable *f*-numbers for light collection that simple filters do.

We note that the so-called 2-color sequencing scheme developed at DuPont[19] is actually a 4-label method. The optics is simplified and the ratio-based base-calling algorithm is fairly straightforward.[7] One reason why this is not more widely used is that the 4 labels have emission bands that are very closely spaced. Even though the intensity ratios (used for base calling) are relatively independent of the incorporation rate of the polymerase reaction, spectral interference and a low S/N (low transmission of the bandpass filters) can lead to ambiguities. So, while there exist various proven base-calling schemes, there is room for improvement in terms of accuracy, speed and simplicity.

In this work, we demonstrate two straightforward and highly reliable base-calling algorithms derived from data acquired in two emission channels in a polymer-filled capillary.[13,20,21] In anticipation of applications in a capillary array, a charge-injection device (CID) camera[21] coupled through broad-band emission filters was used as the detector.

EXPERIMENTAL SECTION

Separation Fused-silica capillaries 45-60 cm long with 75 μm i.d., 150 μm or 360 μm o.d. (Polymicro Technologies, Inc., Phoenix, AZ) were used for separation. The inner wall of the capillary was coated with polyacrylamide or treated with 0.1 M HCl according to the procedure reported elsewhere.[22] The sieving matrix was prepared by dissolving 1.5% of 8-million MW poly(ethyleneoxide) (PEO) and 1.4% of 0.6-million MW PEO in running buffer, which was 1 x TBE with 3.5 M urea. DNA sequencing samples (PGEM/U) were prepared from the Sanger reaction according to standard protocols (ABI, DyeDeoxy Terminators and cycle sequencing with Taq polymerase) in the DNA Facility of Iowa State University. Capillaries were filled with a 5-ml syringe (Becton Dickinson & Co., Franklin Lakes, NJ) for 10-15 min by applying pressure at the syringe with a metal clamp. After a run was completed, the matrix was pushed out with compressed N_2 gas at 300 psi (within 2 min) or with a 100-μl syringe (Hamilton Company, Reno, NV) (within 30 s). Before flushing, the Teflon tubing which was used to connect the capillary to the pressure devices was first filled with water so that the capillary would not become clogged by dried particles. Two 20-ml glass sample vials containing the running buffer were used as buffer reservoirs. The matrix-filled capillary was pre-run for 10 min before injection. Injection was at –9 kV to –12 kV for 20 s or 15 s, depending upon the length of the capillary. Electrophoresis was run by applying

–9 kV to –15 kV at the injection end via a high-voltage power supply (Spellman, Plainview, NY).

Experimental setup The overall setup was similar to that reported before[21] but with a different capillary mount and new spatial arrangement of optical filters to accommodate either two-beam excitation or one-beam excitation. A capillary was mounted on an aluminum block of 1.6 cm (W) x 6.0 cm (L) x 6.5 cm (H) (Figure 1). Two 0.3-cm wide, 3-cm deep grooves were cut on the mount to form two optical channels. The deep grooves helped to reduce scattered light background. The two channels were placed 0.9 cm apart to reduce interference of scattered light between the two detection channels when the two-wavelength, two-beam excitation scheme was employed. The capillary was placed flat across the grooves and taped tight to it (Scotch tape, 3M, St. Paul, MN). The CID camera was set up above the capillary mount so that the capillary was oriented along the CID imager columns as before.[21] As many as 100 capillaries with 150 μm o.d. can be mounted in parallel on the mount.

An air-cooled Ar^+ laser (Uniphase, San Jose, CA, Model 2213-150ML) with multi-line emission was used for excitation. The 488-nm and 514-nm lines were separated with a glass prism. For two-wavelength, two-beam excitation, two 0.5-cm detection windows 1.2 cm apart were formed on the capillary by burning off the coating with boiling sulfuric acid. The detection windows were

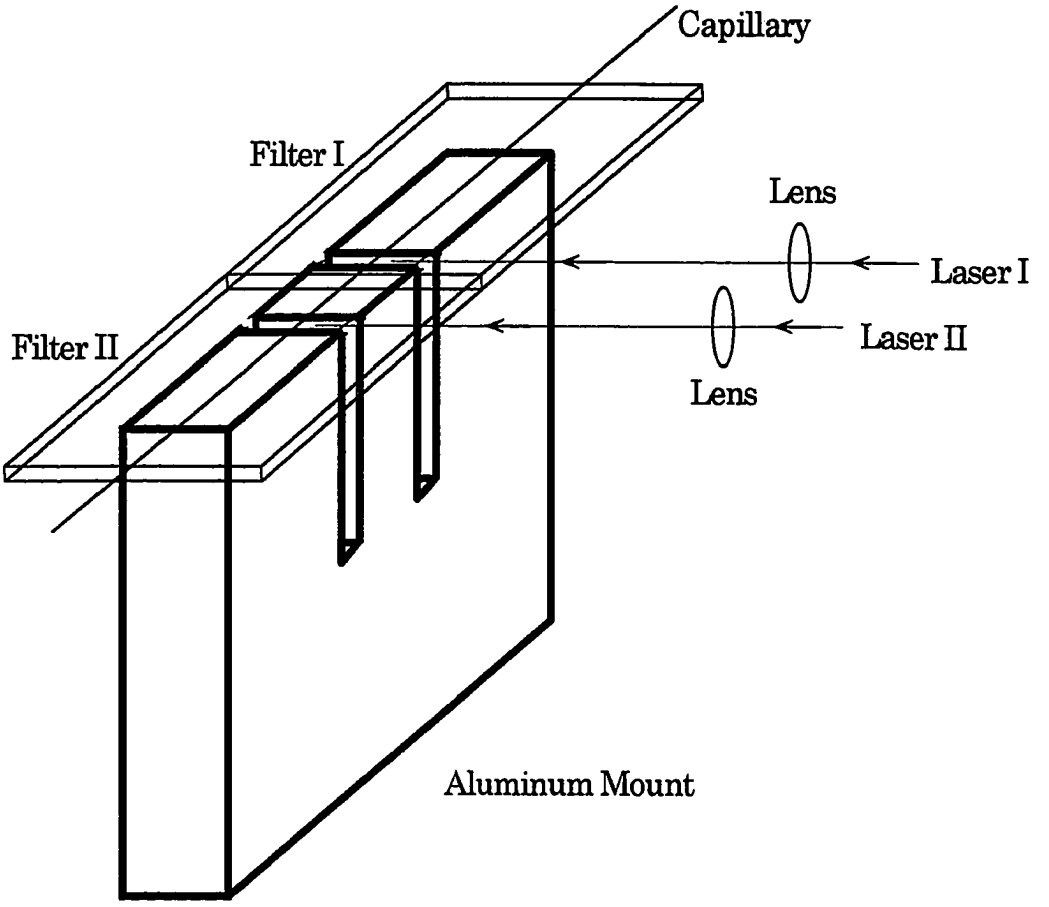


Figure 1. Schematic diagram of two-wavelength two-beam excitation and detection.

placed at the detection channels aligned with the grooves on the aluminum mount. The 488-nm beam and 514-nm beam were focused with 10-cm focal length lenses (Oriel, Stratford, CT) from the same side of the capillary mount. The edges of the two lenses were trimmed off to allow placing them side by side with the centers of the lenses in the light paths. The laser beams were perpendicular to the capillary and the CID camera. A 5 cm x 8 cm x 2 mm quartz plate was placed 4 mm horizontally above the capillary. On the quartz plate, a 600-nm high precision interference long pass filter (Ealing Electro-optics, South Natick, MA) and a RG610 glass filter (Schott Glass, Duryea, PA) was placed on top of the 514-nm excitation channel. A 0.5-inch diameter 488-nm Raman-edge filter (Physical Optics, Torrance, CA) was placed on top of the 488-nm excitation channel.

For one-wavelength, two-beam excitation, the 488-nm beam was split into two with a beam splitter (Melles Griot, Irvine, CA) to illuminate both detection windows. Beam focusing and filter setting were the same as that in the two-wavelength, two-beam excitation scheme.

For one-wavelength, one-beam excitation, only one 488-nm beam was focused onto the capillary. The image of the illuminated region was split into two by a tilted glass plate or an optical filter (Figure 2). To set up this system, the relative positions of the CID camera and the capillary were first aligned so that

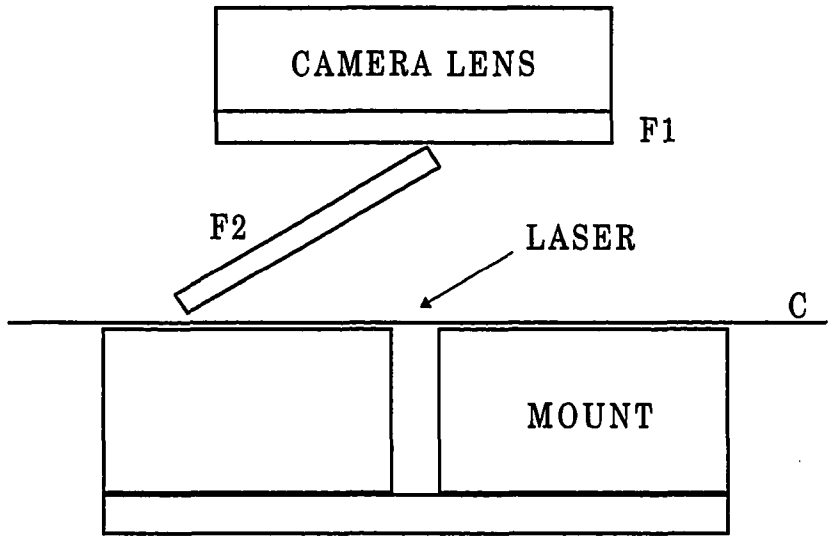


Figure 2. Optical arrangement to split a line of excited fluorescence in the capillaries (C) to 2 emission channels, one through a Raman-edge long-pass filter (F1) and the other through that plus a 610-nm long-pass filter (F2).

Ann Arbor, MI, Model Notch-Plus) was attached onto the camera lens (Nikon 28/1.4 AFD, filter size 72 mm). A 5 cm x 5 cm x 3 mm RG610 filter (Schott Glass) was tilted at 30_ relative to the focal plane of the camera. With the top the image of the detection window could be focused onto the center part of the imager. A 63 mm diameter 488-nm Raman-edge filter (Kaiser Optical Systems, edge of the RG610 filter parallel to the laser beam and facing toward the center of the field of view, the filter was translated until the top edge of the filter passed about 0.5 cm beyond the detection window. Light originating from the half of the field of view which was covered by the filter (red channel) was thus shifted toward the side of the filter. Emission was thus focused as one image which was 9 CID imager rows away from the original image. Light originating from the other half of the field of view (blue channel) was focused as before. The distance between the two images can be changed by adjusting the tilt angle of the RG610 filter.

The RG610 filter can serve dual purposes as both image displacer and optical filter. The RG610 filter also blocks the strongest scattered laser light, which forms a fan perpendicular to the capillary array. This is very important in reducing scattered light background at the blue channel. Although the scattered laser light hitting the RG610 filter produces some fluorescence, the fluorescing part of the RG610 filter is out of focus and does not cause significant background in the image area in the CID detector. For critical focusing, the

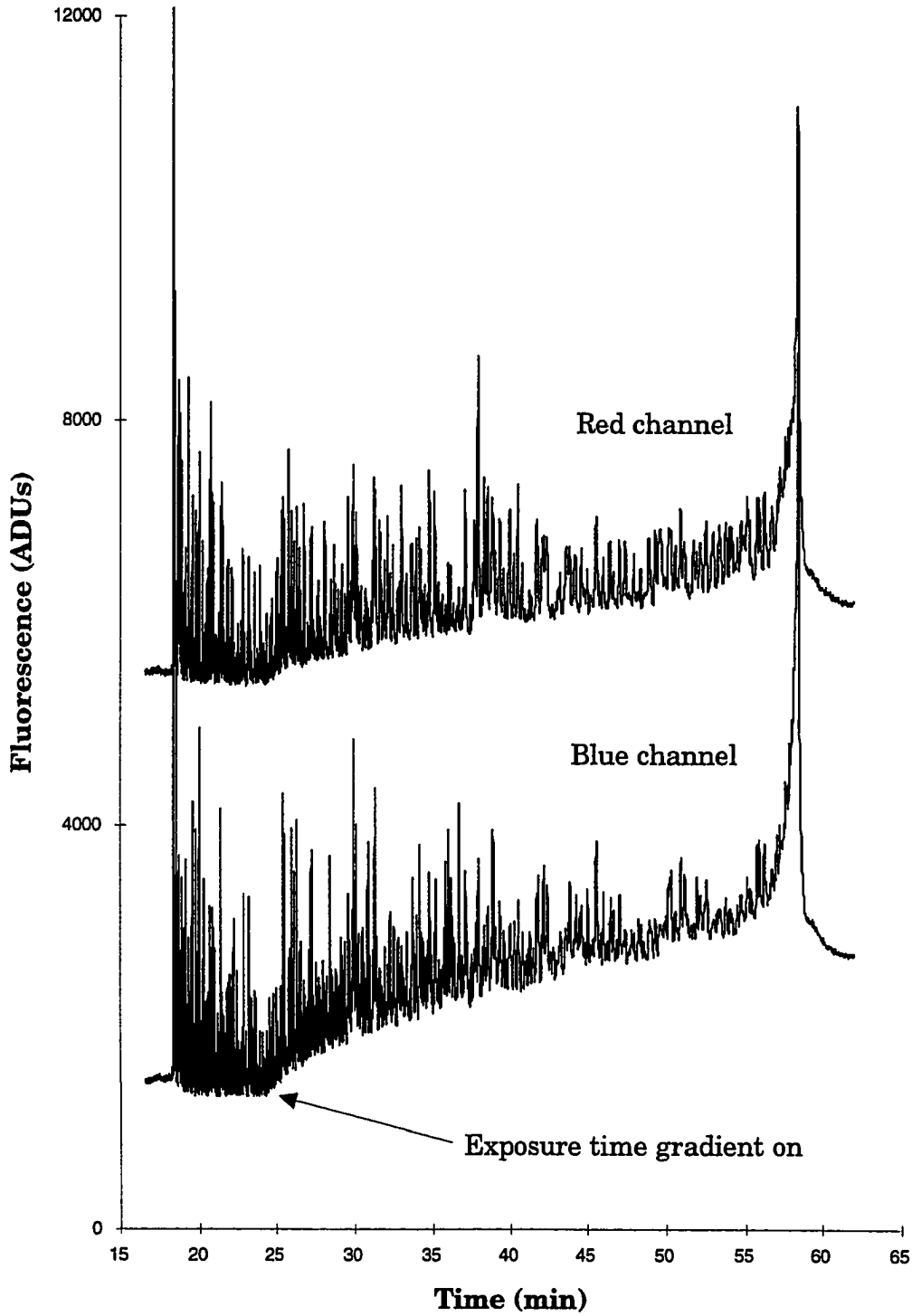
difference in effective optical pathlengths for the light in the two halves of the field of view can be compensated with a quartz plate. The quartz plate can be tilted at the same angle as the RG610 filter and covers the other half of the field of view in the same way as the RG610 filter.

CID camera operation and data analysis Details of CID operation have been described before[21] and will not be repeated here. Data analysis is carried out by using chromatographic software (CP, ChromPerfect, Justice Innovations, Palo Alto, CA) and QuattroPro for DOS (Borland, Scotts Valley, CA).

RESULTS AND DISCUSSION

Peak-height ratios The spectral properties of the 4 standard dye labels for the Sanger reaction provide the possibility that they may be distinguished with a peak-height ratio coding method when two long-pass filters are used for the two spectral channels. The procedure is straightforward, but has not yet been automated for on-line use. Electropherograms are generated right after a run is finished (Figure 3). For one capillary, two files are generated, each of which corresponds to one spectral channel. Integration results from standard chromatographic software are saved in ASCII files. Only migration times and peak heights need to be imported into a worksheet for base

Figure 3. Electropherograms of separation and detection of PGEM/U DNA sample by using one-wavelength, one-beam excitation and two-color detection scheme. Capillary was 360 μm o.d., 75 μm i.d., bare fused-silica with 60 cm total length and 50 cm effective length. The capillary was treated with 0.1 M HCl for 3 h before use. Excitation laser power was 20 mW (488 nm) at the detection window. Sample injection was at –12 kV for 20 s. Electrophoresis was run at –15 kV. Exposure time was 200 ms for 25 min and then increased by 350 μs per frame until the run was finished. When the last peak came out at 60 min, the exposure time was 1100 ms.

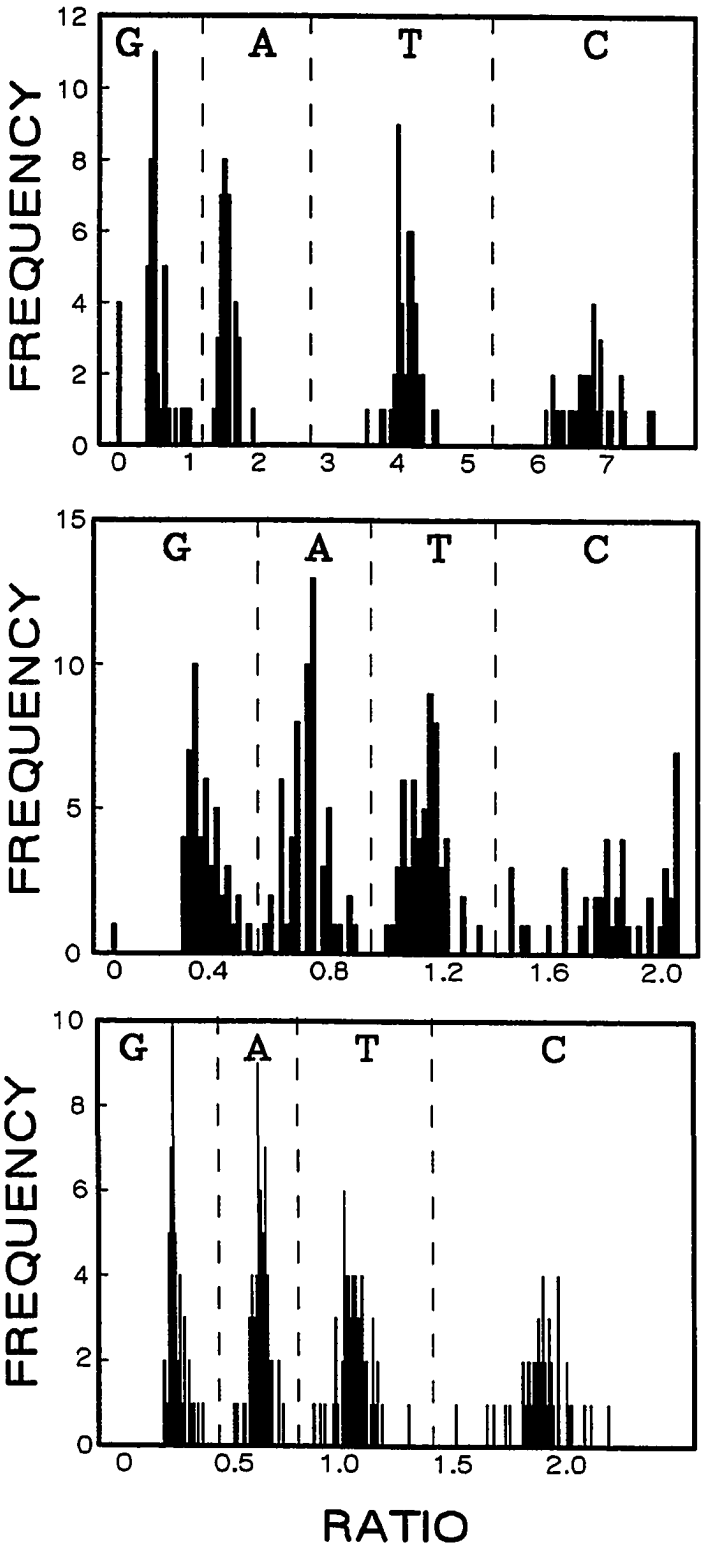

12000
8000
4000
0
Fluorescence (ADUs)
Red channel
Blue channel
Exposure time gradient on
15
20
25
30
35
40
45
50
55
60
65
Time (min)

calling. Migration time is used as the index to match the peaks of the same base at the two spectral channels.[21]

Different combinations of filters were examined for their abilities to discriminate the 4 labels from one another (Figure 4). Figure 4 also shows the importance of selecting the proper emission filters. The Raman-edge filter lets through more light (which is always desirable) and discriminates better between A and G compared to a 515-nm cutoff filter. On the other hand, a 610-nm cutoff filter discriminates better between C and T compared to a 600-nm filter and is chosen even though the latter lets through more light. The best combination turns out to be the 488-nm Raman-edge (RE) filter and RG610 filter. Both filters allow all bases to be detected at both spectral channels. The peak-height ratio can thus be calculated for all bases. The major advantage of using long-pass filters is that optical throughput is maximized. This is important because the detection limit determines how far one can read the bases in a given sequencing run.

A histogram is built from the peak-height ratios. In order to establish the criteria to call bases with good certainty, only the peaks within good separation resolution range are used in this calibration step. Peaks within 250 bp length certainly meet this requirement. The peaks fall into four distinct clusters (Fig. 4). Three ratios (dotted lines) are determined from the locations of the gaps

Figure 4. Clustering of peak-height ratios for the 4 standard dye labels. Dashed lines are demarcation points for base calling. Top: (514-nm excitation/610-nm long-pass emission) vs. (488-nm excitation/Raman-edge long-pass emission); middle: (488-nm excitation/600-nm long-pass emission) vs. (488-nm excitation/515-nm long-pass emission); and bottom: (488-nm excitation/610-nm long-pass emission) vs. (488-nm excitation/Raman-edge long-pass emission).


G
A
T
C
FREQUENCY
RATIO
0
1
2
3
4
5
6
7
0.4
0.8
1.2
1.6
2.0
0.5
1.0
1.5
2.0
12
10
8
6
4
2
15
5

between the four clusters, namely R_1, R_2, R_3 (Table I). A formula is set in the spreadsheet to call all the bases automatically with $G < R_1$, $R_1 < A < R_2$, $R_2 < T < R_3$, and $R_3 < C$ (Table II, last column). Standard deviations of the peak-height ratios are calculated for all of the same kind of bases within the readable base-pair range (Table I). The boundaries of the four clusters, i.e. R_1, R_2, R_3, are thus defined with certainty at least as high as 99.7% ($\pm 3\sigma$). As long as the optical setup does not change, the base-calling criteria, which is simply the R_1, R_2, R_3 values, will not change. A series of unknown DNA samples can then be sequenced in the calibrated system up to the readable length. An example of a typical spreadsheet output for peak-height ratios is shown in Table II.

When an electropherogram is analyzed with CP, integration starts at the very front edge of the first peak and ends at the trailing edge of the last peak. In this case, CP determines the correct baseline in the entire range without the need for manual intervention. The rising baseline due to exposure time gradient[21] does not affect the ability of CP to choose the proper baseline. But CP does not detect all of the peaks automatically, missing some partially resolved peaks ($R < 1.0$). This is the primary cause of base-calling errors. Presumably, better peak-finding software will avoid such difficulties.

In two-wavelength, two-beam excitation, base calling is based upon differences in fluorescence spectra as well as in absorption spectra. It also provides better S/N for the 2 rhodamine labels (Figure 4, top). DNA fragments are

Table I. Base-calling criteria for peak-height ratios (red channel divided by blue channel) derived from Figure 4, bottom

Cluster	R_{avg}	σ	3σ	Cluster range
G	0.27	0.043	0.130	0.14 ~ 0.40
A	0.63	0.040	0.120	0.51 ~ 0.75
T	1.06	0.068	0.21	0.85 ~ 1.26
C	1.89	0.197	0.59	1.30 ~ 2.48

R_{avg}—average peak-height ratio of bases up to 250 bp.

σ—standard deviation of peak-height ratio of bases up to 250 bp.

From these clusters, the base-calling criteria can be defined with $R_1 = 0.46$, $R_2 = 0.79$, and $R_3 = 1.28$.

Table II. Worksheet for base calling

t_m		Blue channel		Red channel			Called	Base
(min)	F	t'_m	Hb	t'_m	Hr	Hr/Hb	bases	No.
24.35	6612	55.38	617	55.38	655	1.06	T	150
24.38	6620	55.45	524	55.44	119	0.23	G	
24.48	6647	55.68	934	55.68	548	0.59	A	
24.55	6665	55.83	356	55.82	211	0.59	A	
24.62	6683	55.98	409	55.98	279	0.68	A	
24.66	6694	56.07	780	56.07	816	1.05	T	
24.72	6710	56.2	307	56.2	283	0.92	T	
24.77	6724	56.32	937	56.32	225	0.24	G	
24.84	6744	56.49	297	56.49	304	1.02	T	
24.89	6758	56.6	275	56.6	307	1.12	T	
24.94	6772	56.72	903	56.72	565	0.63	A	160
24.99	6784	56.83	950	56.83	1028	1.08	T	
25.02	6792	56.89	602	56.89	1188	1.97	C	
25.08	6809	57.03	615	57.03	1133	1.84	C	
25.13	6823	57.15	445	57.14	133	0.3	G	
25.21	6845	57.33	159	57.32	311	1.95	C	
25.25	6856	57.43	947	57.43	977	1.03	T	
25.31	6871	57.55	881	57.55	1633	1.85	C	
25.35	6883	57.65	2862	57.65	1768	0.62	A	
25.4	6896	57.76	584	57.76	1070	1.83	C	
25.44	6908	57.86	2484	57.86	1561	0.63	A	170

t_m—actual migration time.

F—frame number.

t'_m—migration in CP time scale.

Hb, Hr—peak heights for blue (b) and red (r) channels from CP software.

excited twice when they go through the two detection windows. Excitation and light collection efficiency are easy to optimize. However, scattered 514-nm light introduces extra background at the 488-nm channel where the RE filter is used because the RE filter does not block 514-nm light. Careful alignment and proper setup need to be exercised in order to minimize the interference. It is necessary to physically isolate the two channels by placing a shield between them (Fig. 1). The physical isolation is especially important when the capillary array is immersed in refractive-index matching liquid[20] because Rayleigh and Raman scattering is much stronger in liquid than in air. Scattered laser light not only distributes around the outside of the capillary, but also propagates within the capillary by internal reflection due to the difference of refractive index between the matrix, the capillary wall and air. It is found to be beneficial to leave several mm of polyimide coating between the two detection windows. The coating absorbs scattered laser light that propagates from one detection channel along the capillary to the other detection channel.

We found that all of the miscalled bases in the two-wavelength, two-beam excitation mode are features where the adjacent peaks were not resolved and were thus assigned inaccurate peak heights or are missed entirely by the standard chromatographic software. However, visual inspection of the raw data clearly revealed shoulders or unusually broad peaks corresponding to $1.0 > R >$

0.5 in the separation. It is expected that more sophisticated peak-recognition programs or deconvolution programs can resolve this problem in the future.

We then realized that there are two aspects of this base-calling scheme that need improvement. Matching of the migration times by normalization to the relative distance traveled depends on having uniform velocities (temperature, matrix homogeneity) along the entire capillary. The two laser beams also produce scattered light that can interfere with each other (514 nm laser line transmitted by the Raman-edge filter) and decrease the S/N. So, we tried to use the 488 nm laser line by itself. All 4 standard dye labels absorb at 488 nm, although not all 4 do so equally efficiently.[6,14,15] In our first attempt, we still use the two window approach[13,21] but we split the output from one laser to favor the red channel (passing through the 610 nm filter) to compensate for the lower absorption of those 2 dye labels. The results are essentially the same as those obtained by using 2 wavelengths in excitation (Figure 4). The advantages are lower stray light (no 514 nm present) and simplicity in using a single-line laser. Matching the migration times from the 2 channels is still necessary, however.

Our ultimate solution is to use one laser and one window in excitation. Superficially, this seems inadequate for calling 4 dye labels. In Figure 2, we show an elegant solution that splits the image into 2 emission channels with maximum light throughput. There is then no need to convert the time scales of

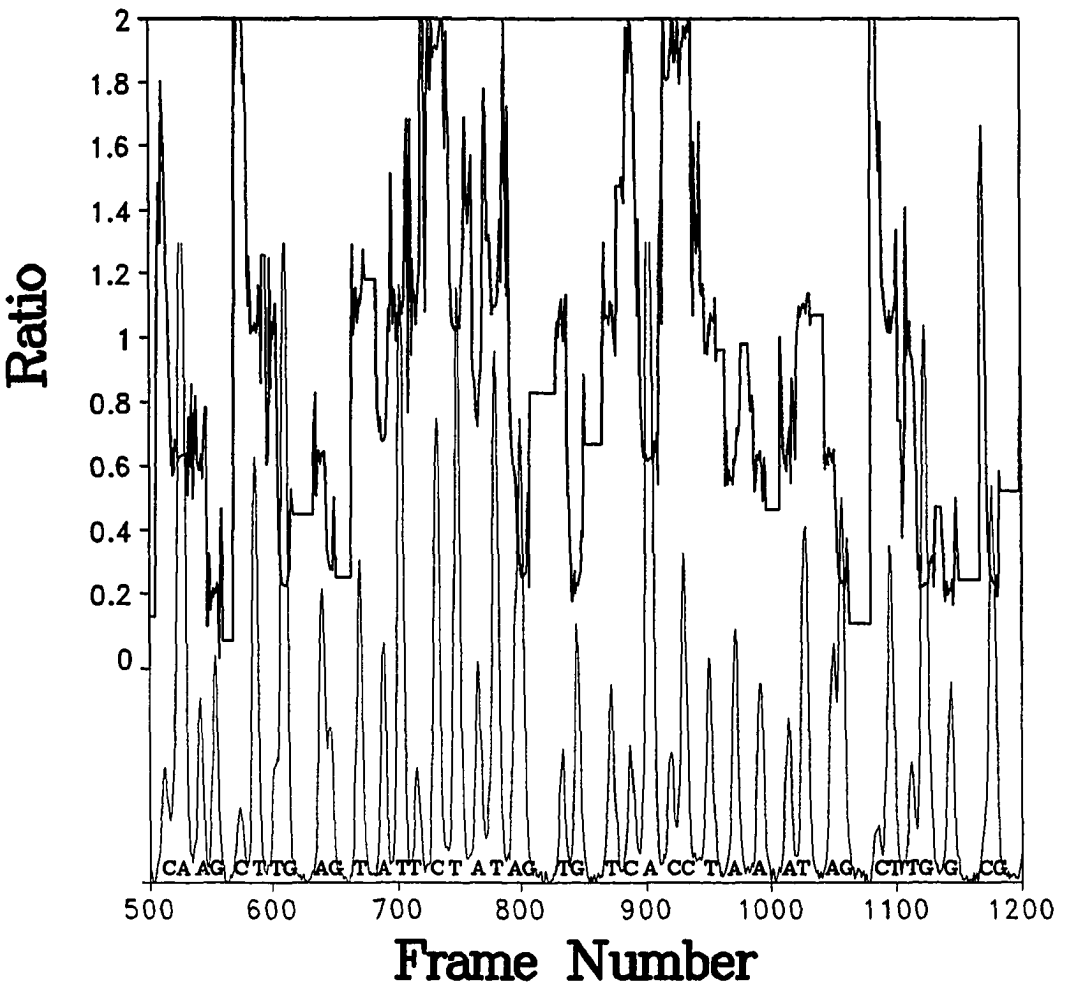
the two electropherograms obtained at the two spectral channels in order to match the peaks of a DNA fragment at the two spectral channels. This simplifies data analysis and improves base-calling accuracy. When exposure-time gradient is applied,[21] the one-beam excitation scheme also eliminates the chances of peak mismatch caused by nonuniform distribution of temperature or voltage along a capillary. This is because peak matching in the two-beam excitation scheme relies on the assumption that a DNA fragment migrates along a capillary with constant velocity. More importantly, the one-beam excitation scheme provides the freedom of operating the system at any gradient mode necessary to enhance resolution or detectability, such as exposure-time gradient, temperature gradient and voltage gradient, or combinations of these gradient methods.

By using the one-wavelength, one-beam excitation mode, within 330 bp, base-calling accuracy is 99.3%, i.e. only two errors occurred. One G was missed among the four Gs between 51 bp and 54 bp. The other error occurred at 317 bp where A was miscalled as T. Beyond 250 bp, extra care is needed because resolution and S/N is reduced. When partially resolved peaks are split by software, the position of splitting is critical. Some of the peaks are resolved in one spectral channel but unresolved in the other. When base calling proceeds to 353 bp, base-calling accuracy decreases to 97.1%. Although S/N is still sufficient for integration beyond 353 bp, it is not practical to call bases further in Figure 3

because that is roughly the limit of confidence for the Taq-catalyzed Sanger reaction.

Ratiograms Very recently, we devised another novel approach to base calling based on the one excitation laser/two emission wavelength data described above. However, instead of relying on software to identify and to determine the peak heights at each channel, we generated a "ratiogram", which is simply the ratio of signals from the 2 channels calculated point by point at each data interval (Figure 5). Such ratiograms have been used in liquid chromatography to determine peak purity when using diode array detectors or rapid-scan multiwavelength detectors. Software is in fact included in several commercial instruments. The idea is that the ratio of intensities at two independent wavelengths is independent of concentration (which varies across the peak), and can be used to sort out the unresolved components in the merged peaks. As long as the overall S/N is good (Figure 3), even peaks with resolution $R < 0.5$ can in many cases be identified by noting the ratios at the leading edge and at the falling edge of the merged peak. We still need the usual electropherogram to determine where the ratios are meaningful and where the signals are at the noise level and are therefore meaningless. However, peaks need not be resolved by the chromatography software and errors in determining peak heights (when unresolved) are avoided.

Figure 5. Ratiogram (top, heavy line) base calling for a portion of Figure 3 (bases 92 to 132). The horizontal lines are artifacts to prevent division errors when S/N is too low (at the peak valleys) to determine a meaningful ratio. The raw signal from the 488-nm excitation/Raman edge emission channel is plotted below (light line).


Ratio
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
500
600
700
800
900
1000
1100
1200
Frame Number

An actual example of such a "ratiogram" is plotted in Figure 5 on top of the electropherogram obtained through the RE filter, which should record all peaks regardless of the label. The called bases are typed below, as derived from the same criteria defined by Table I. The accuracy is 99% through 340 bases. Note the feature at marker 800. It is clearly broader than the surrounding features, indicating an overlapping set of fragments. The ratiogram clearly shows that the leading edge is "A" and the trailing edge is "G" in character. So, even though this feature led to base-calling error in the peak-height scheme using chromatography software, it was correctly called in this novel scheme. Other noteworthy portions are the regions around markers 600 and 1180, where partially resolved features are correctly called in a similar fashion.

We found that defining the baseline for the peaks is critical to base-calling accuracy. All 3 errors (< 340 bp) occurred in one stretch around 260 bp where our simple baseline-selection algorithm resulted in negative values for some of the fluorescence intensities in the 488-nm cutoff channel. Similarly, there was a series of errors around 350 bp where there was insufficient background subtraction. Refinement in the software should allow accurate base-calling to > 400 bp, since the peak resolution there is still better than 0.5.

The implication is that we have demonstrated the conceptual framework for a simple but highly accurate base-calling scheme that is independent of concentration (incorporation rate by the polymerase), is useful even for very poorly

resolved peaks (potentially permitting base calling to larger fragments and/or sacrificing some resolution for speed in the separation), requires minimal computation (improving speed and decreasing effort for data processing), and is compatible with the high light-throughput optics for excitation/emission described in previous sections.

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, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent and L. E. Hood, Nature **321**, 674-679 (1986).
2. F. Sanger, S. Nicklen and A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467 (1977).
3. X. C. Huang, M. A. Quesada and R. A. Mathies, Anal. Chem. **64**, 967-972 (1992).
4. T. T. Lee and E. S. Yeung, Anal. Chem. **63**, 2842-2848 (1991).

5. D. Chen, H. R. Harke and N. J. Dovichi, Nucl. Acids Res. **20**, 4873-4880 (1992).

6. R. Tomisaki, Y. Baba, M. Tsuhako, S. Takahashi, K. Murakami, T. Anazawa and H. Kambara, Anal. Sci. **10**, 817-820 (1994).

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

8. M. C. Ruiz-Martinez, J. Berka, A. Belenkii, F. Foret, A. W. Miller and B. L. Karger, Anal. Chem. **65**, 2851-2858 (1993).

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

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

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

12. J. A. Taylor and E. S. Yeung, Anal. Chem. **65**, 956-960 (1993).

13. K. Ueno and E. S. Yeung, Anal. Chem. **66**, 1424-1431 (1994).

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

15. A. E. Karger, J. M. Harris and R. F. Gesteland, Nucl. Acids Res. **19**, 4955-4962 (1991).

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

17. H. A. Fishman, R. N. Zare, J. V. Sweedler, J. B. Shear and R. H. Scheller, Anal. Chem. **63**, 496-502 (1991).

18. A. T. Timperman, K. Khatib and J. V. Sweedler, Anal. Chem. **67**, 139-144 (1995).

19. J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jensen and K. Baumeister, Science **238**, 336-341 (1987).

20. X. Lu and E. S. Yeung, Appl. Spectrosc. **49**, 605, 1995.

21. Q. Li and E. S. Yeung, Appl. Spectrosc. **49**, 825, 1995.

22. E. N. Fung and E. S. Yeung, Anal. Chem., **67**, 1913, 1995.

GENERAL CONCLUSION

Capillary electrophoresis is highly suitable for DNA sequencing in both aspects of separation and detection. The high-degree miniaturization of the CE instrument makes it possible to achieve both high separation speed and high detection sensitivity. More interestingly, it provides the possibility of constructing a highly multiplexed system without tremendous expansion. For example, a 100-capillary array only occupies 1.5 cm in width. The array system can be run by using only one laser, one detector, and one desktop computer, as one would do with a single capillary system. Therefore, with a limited expansion, several orders of improvement in throughput can be achieved.

The optical setup and the array detector are among the most important undertakings in developing a highly multiplexed CE instrument. Both high light throughput of the optical system and high sensitivity of the array detector are maintained in order to avoid using more complicated and thus more expensive excitation source. The sampling rate of the array detector is also very important for the detection system to catch up with the high speed of separation.

In the multiplexed CE system developed here, a simple yet elegant optical detection system has been established to provide a highly accurate base-calling scheme for DNA sequencing based on standard Sanger chemistry. The charge injection device is operated in random access mode to achieve high sampling rate

for high-speed separation and to reduce data volume. The exposure time gradient mode has been demonstrated to be highly effective for compensating for the decrease of peak intensity for longer DNA fragments in a Sanger reaction product.

Combination of the high-throughput detection system and the high-speed separation system produces unprecedented DNA sequencing speed. Based on the 100-capillary multiplexed CE technology, a single instrument can be used to sequence the whole human genome within two years. With some further development, the system can be readily scaled up to 1000-capillary. Sequencing the whole human genome will only take two months with a single instrument.

The multiplexed CE system developed here is not limited to DNA sequencing. Its inherent high speed and high throughput will surely produce a revolutionary impact on other biological applications. More and more mysteries of human development, life and disease will be revealed with an unprecedented speed with the aid of this technology.

LITERITURE CITED

1. A. Tiselius, Trans. Faraday Soc., 1937, 33, 524-535.
2. S. Raymond and L. Weintraub, Science, 130, 711-713 (1959).
3. R. Virtenen, Acta Polytech. Scand., 123, 1 (1974).
4. F. E. P. Mikkers, *et al*, J. Chromatogr., 169, 11 (1979).
5. J. W. Jorgenson and K. D. Lukacs, Anal. Chem., 53, 1298 (1981).
6. S. Hjerten, J. Chromatogr., 270, 1 (1983).
7. S. Terabe, *et al*, Anal. Chem., 56, 113 (1984).
8. S. Hjerten and M. Zhu, J. Chromatogr., 346, 265 (1985).
9. A. Cohen and B. L. Karger, J. Chromatogr., 397, 409 (1987).
10. E. S. Yeung, Optical detectors for capillary electrophoresis, in Advances in Chromatography, vol. 35, P. R. Brown and E. Gruhka, Eds., Marcel Dekker, Inc.: New York, 1995.
11. S. F. Y. Li, Capillary Electrophoresis, Elsevier: Amsterdam, 1992.
12. X. Huang, J. A. Luckey, M. J. Gordon, and R. N. Zare, Anal. Chem., 61, 766 (1989).
13. T. Kaneta, S. Tanaka, and H. Yoshida, J. Chromatogr., 538, 385 (1991).
14. P. Bocek, M. Deml, J Pospichal, and J. Sudor, J. Chromatogr., 470, 309 (1989).
15. S. J. Hjerten, J. Chromatogr., 347, 191-198 (1985).

16. C. S. Lee, W. C. Blanchard, C. T. Wu, Anal. Chem., 62, 1550 (1990).

17. N. K. Fung and E. S. Yeung, Anal. Chem., 67, 1913 (1995).

18. D. N. Heiger, A. S. Cohen, B. L. Karger, J. Chromatogr., 516, 33-48 (1990).

19. J. Sudor, F. Foret, and P. Bocek, Electrophoresis, 12, 1056-1058 (1991).

20. H. Pulyyaeva, D. Wheeler, M. M. Garner, and A. Crambach, Electrophoresis, 13, 608-614 (1992).

21. D. Tietz, A. Aldroubi, H. Pulyaeva, T. Guszczynski, M. M. Garner, and A. Crambach, Electrophoresis, 13, 614-616 (1992).

22. M. Chiari, M. Nesi, M. Fazio, and P. G. Righetti, Electrophoresis, 13, 690-697 (1992).

23. A. Guttman, B. Wanders, and N. Cooke, Anal. Chem., 64, 2348-2351 (1992).

24. M. C. Ruiz-Martinez, J. Berka, A. Belenkii, F. Foret, A. W. Miliier, and B. L. Karger, Anal. Chem., 65, 2851-2858 (1993).

25. N. Best, E. Arriaga, D. Chen, and Norman J. Dovichi, Anal. Chem., 66, 4063-4067 (1994).

26. (Joint U.S. Department of Energy and U.S. Department of Health and Human Services Report DOE/ER-0452P. Understanding Our Genetic Inheritance--The U.S. Human Genome Project: The Firt Five Years; Washington, DC, April 1990.

27. H. T. Chang and E. S. Yeung, J. Chromatogr., in press.

28. Q. Li and E. S. Yeung, Applied Spectrosc., accepted.

29. R. A. Wallingford and A. G. Ewing, Anal. Chem., 59, 678 (1987).

30. R. Smith, C. Barrinaga, and H. Udseth, Anal. Chem., 60, (1988).

31. Y. Xue and E. S. Yeung, Anal. Chem., 65, 1988 (1993).

32. S. A. Soper, E. B. Shera, J. C. Martin, J. H. Jett, J. H. Hahn, H. L. Nutter, and R. A. Keller, Anal. Chem., 63, 432 (1991).

33. J. Zhao, D. Chen, and N. J. Dovichi, J. Chromatogr., 608, 117 (1992).

34. H. Swerdlow, J. Zhang, D. Chen, H. R. Harke, R. Grey, S. Wu, N. J. Dovichi, and C. Fuller, Anal. Chem., 63, 2835 (1991).

35. E. S. Yeung, P. Wang, W. Li, and R. W. Giese, J. Chromatogr., 559, 183 (1991).

36. X. C. Huang, M. A. Quesada, and R. A. Mathiews, Anal. Chem., 64, 967-972 (1992).

37. J. A. Taylor and E. S. Yeung, Anal. Chem., 64, 1741-1744 (1992).

38. B. L. Hogan and E. S. Yeung, Anal. Chem., 64, 2841-2845 (1992).

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

40. J. A. Taylor and E. S. Yeung, Anal. Chem., 65, 956-960(1993).

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

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

43. T. N. Buican, M. J. Smyth, and H. A. Crissman, Applied Optics, 26, 5311-5316 (1987).

44. D. W. Tank, M. Sugimori, J. Connor, and R. R. Llinas, Science, 242, 773-777 (1988).

45. E. Betzig, J. H. Trauyman, T. D. Harris, J. S. Weiner, and R. L. Kostelak, Science, 251, 1468-1470 (1991).

46. E. Ishikawa, S. Hashida, T. Kohno, and K. Hirota, Clin. Chim. Acta, 194, 51-72 (1990).

47. R. M. McCamman, D. Weinrich, and H. J. Borys, J. Neurochem., 21, 473-476 (1993).

48. W. Tan, Z. Shi, S. Smith, D. Birnbaum, and R. Kopelman, Science, 258, 778-781 (1992).

49. J. A. Janowski, T. J. Schroeder, R. W. Holz, and R. M. Wightman, J. Biol. Chem., 267, 18329-18335 (1992).

50. R. T. Kennedy, R. L. St. Claire, J. G. White, J. W. Jorgenson, Mikrochim. Acta, II, 37-45 (1987).

51. B. R. Cooper, J. A. Jankowski, D. J. Leszczyszyn, M. R. Wightman, and J. W. Jorganson, Anal. Chem., 64, 691-694 (1992).

52. J. B. Chien, R. A. Wallingford, A. G. Ewing, J. Neurochem., 54, 633-638 (1990).

53. T. Lee and E. S. Yeung, Anal. Chem., 64, 3045-3051 (1992).

54. T. Lee, S. J. Lillard and E. S. Yeung, Electrophoresis, 14, 429-438 (1993).

55. Q. Li and Yeung, J. Capillary Electrophoresis, 1, 55-61 (1994).

56. Q. Xue and Yeung, J. Chromatogr. A, 661, 287-295 (1994).

57. Q. Xue and Yeung, Anal. Chem., 66, 1175-1178 (1994).

58. S. Lillard and E. S. Yeung, submitted to J. Chromatogr..

59. W. Tan and E. S. Yeung, Anal. Biochem., 226, 74-79 (1995).

60. Z. Rosenzweig and E. S. Yeung, Anal. Chem., 66, 1771-1776 (1994).

61. N. H. Cheung and E. S. Yeung, Appl. Spectrosc., 47, 882-886 (1993).

62. N. H. Cheung and E. S. Yeung, Anal. Chem., 66, 929-936 (1994).

63. H. T. Chang and E. S. Yeung, Anal. Chem., 67, 1079-1083 (1995).

64. T. F. Lee, The Human Genome Project, Cracking the Code of Life, Plenum Press: New York, 1991.

65. D. A. McGregor, PhD Thesis, Iowa State University, 1993, p.2-4.

66. E. Marshall, Science, 267, 783-784 (1995).

67. E. Marshall, Science, 267, 1270-1271 (1995).

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

69. Q. Li and E. S. Yeung, Appl. Spectrosc., 49, 825-833 (1995).

70. R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir, Science, 147, 1462-1465 (1965).

71. F. Sanger, S. Nicklen, and A. R. Coulson, Proc. Nat. Acad. Sci., 74, 5463 (1977).

72. A. Maxam and W. Gilbert, Proc. Nat. Acad. Sci., 74, 560-564 (1977).

73. E. Y. Chen, The Efficiency of Automated DNA Sequencing, in Automatic DNA Sequencing, in Automated DNA Sequencing and Analysis, M. D. Adams, C. Fields, and J. C. Venter, Eds, Acadamic Press: San Diego, 1994.

74. L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood, Nature, 321, 674 (1986).

75. W. Ansorge, B. S. Sproat, J. Stegemann, C. Schwager, J. Biochem. Biophys. Meth., 13, 315 (1986).

76. J. M. Prober, G. L Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza,M. A. Jensen, and K. Baumeister, Science, 238, 336 (1987).

77. S. A. Soper, L. M. Davis, F. R. Fairfield, M. L. Hammond, C. A. Harger, J. H. Jett, R. A. Keller, B. L. Marrone, J. C. Martin, H. L. Nutter, E. B. Shera, and D. J. Simpson, Proc. SPIE, 1435, 168-178 (1991).

78. H. G. Hansma and P. K. Hansma, SPIE, 1891, 66-70 (1993).

79. C. H. Chen, S. L. Allman, K. Tang, and R. B. Jones, SPIE, 1891, 94-101 (1993).

80. M. D. Eggers, M. E. Hogan, R. K. Reich, J. B. Lamture, K. L. Beattie, M. A. Hollis, D. J. Ehrlich, B. B. Kosicki, J. M. Shumaker, R. S. Varma, B. E. Burke, A. Murphy, and D. D. Rathman, SPIE, 1891, 113-126 (1993).

81. S. Fung, S. L. Woo, R. P. Haugland, S. M. Menchen, and C. R. Connell, U. S. Patent #4,855,225 (August, 8, 1989).

82. D. Chen, H. R. Harke and N. J. Dovichi, Nucl. Acids Res. 20, 4873-4880 (1992).

83. R. Tomisaki, Y. Baba, M. Tsuhako, S. Takahashi, K. Murakami, T. Anazawa and H. Kambara, Anal. Sci. 10, 817-820 (1994).

84. D. Chen, H. P. Swerdlow, H. R. Harke, J. Zhang, and N. J. Dovichi, SPIE, 1435, 161-167 (1991).

85. S. L. Pentoney, J. K. D. Konrad, and W. Kaye, Electrophoresis, 13, 467-474 (1992).

86. X. C. Huang and R. A. Mathies, Appilacation of Capillary Electrophoresis to DNA Sequencing, in Automated DNA Sequencing and Analysis, M. D. Adams, C. Fields, and J. C. Venter, Eds, Acadamic Press: San Diego, 1994.

87. S. Tabor and C. C. Richardson, J. Biol. Chem., 265, 8322-8328 (1990).

88. W. Ansorge, B. Sproat, J. Stegemann, C. Schwager, and M. Zenke, Nucleic Acids Research, 15, 4593-4602 (1990).

89. J. A. Luckey, H. Drossman, A. J. Kostichka, D. A. Mead, J. D'Cunha, T. B. Norris, and L. M. Smith, Nucleic Acids Research, 18, 4417-4421 (1990).

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

91. A. E. Karger, J. M. Harris, and R. F. Gesteland, Nucleic Acids Research, 19, 4955-4962 (1991).

ACKNOWLEDGMENTS

This work was performed at Ames Laboratory under Contract No. W-7405-Eng-82 with the United States Department of Energy. The United States government has assigned the DOE Report number IS-T1758 to this thesis.

I would say that I made the right decision to come to the Chemistry Department of Iowa State University, except that the winter in Iowa is a little bit hard to handle. When I first arrived here, I was inspired by different projects in the department. At the first contact with Dr. Yeung when I was looking for a group to join, Dr. Yeung described the CE instrument in a very plain way. He told me that a single cell can be analyzed in a capillary as thin as a human hair. I became very curious and obviously was attracted by the beauty of this tiny instrument (in fact it is not tiny if the laser is included). I could not let it go. I felt lucky that I was still able to join the group, although it had been crowded at that time.

The view opened wider after I got into the field. Unavoidably, I felt confused at the beginning. I was guided through the jungle step by step. More important than knowing what to do, I learn how to learn. It is fortunate that I work for Dr. Yeung who always conceives of and plans for future scientific exploration and challenge. The disciplines that he teaches me will always be

beneficial in my career life. He deserves my respect and gratitude more than just a word of thanks.

I would like to thank Drs. Houk, Johnson, Kostic, and Kintanar for being my POS committee members and helping me through the study program.

Every one I worked with in Yeung's research lab deserves my gratitude both professionally and personally, although I cannot mention all the names here.

First, I thank Sheri Lillard and Chris Smith for providing English check on my thesis.

Thomas Lee and David McGregor, my previous office mates, deserve my thanks for helping to orient me in the group. Yongjun Xue and his wife Hung Tian took care of me when my family was not here in the first year. They shared the joy with me when my son was born in China. The celebration party in their apartment is always a nice memory.

Thanks goes to Zeev Rosenzweig who always had a good chat at lunch time when we were in the same office. He shared his thoughts with me, and comforted me on my bad days.

Leslie Waits and Chris Smith deserve my thanks for share of knowledge of vacuum techniques and hand-on help whenever I need.

Thanks also goes to my peers H. T. Chang, Sheri Lillard, Qifeng Xue and Sarah Chang for helping, learning from, and sharpening each other. Of course, we have enjoyed so many parties together.

Special thanks goes to Weihong Tan who provides understanding, encouragement, and friendship. Challenging questioning will never be an insult if it comes out frankly. The only insult he ever gives me is that he showed me around how to go to a swimming pool although he just learned how to swim while I learned that more than twenty years ago. I would better not blame him because he enjoys his new sports so much.

When I need a gel, Eliza Fung is there. I felt guilty for getting a gel with just a yell. In turn, she always tries to bride my son with candies so that my son will force me to work for her. But my son is on my side. Sorry, Eliza.

Last, I thank my wife, Shuangyan, for years of support and companion, and my son, Haobo, for giving us so much joy of life.

APPENDIX A: OVERVIEW OF THE SOFTWARE

The software package has been developed over the last one and a half year according to the needs of experiments. There is not a clear-cut date on which the software is completed. The programs are organized here in such a format that will make the software package easier to understand, maintain, use, and improve. In every module, there is a change history note. Any modification should be recorded under the change history note for reference. This is very important for a group of users to use and further develop the software.

The software is developed by using the CIDTEC standard function library. It is unnecessary to repeat the descriptions about the use of these functions which have been available in the user's manual. Before one begins to read the "C" codes listed here, it is suggested that he/she at least glimpse at the SCM5000E Software User's Manual for the description of each function in the CIDTEC standard function library. In order to save time, it is also recommended that one read the Chapter 2 in this dissertation in order to follow the main frame of the programs. The flow charts in the Chapter 2 will be helpful.

The user-developed modules listed in the Appendices B-E do not represent the complete package that can be run on their own. The necessary auxiliary header file, modules, and function libraries are described as in the following.

Header file *scm_mce.h*: This header file must be included in the modules *mce_lib.c, img_lib.c, cal.c, mce.c*, and *img.c*. The header file is basically the same as the header file *scm_dos.h* provided in the SCM5000E Software User's Manual except that the following structure definition is added by the user:

```
typedef struct {
        char        name[5];     // subarray name
        unsigned    X;           // subarray X coordinate
        unsigned    Y;           // subarray Y coordinate
        unsigned    W;           // subarray width
        unsigned    H;           // subarray height
        unsigned    N;           // subarray read number
        unsigned    G;           // subarray read gain
        long        waitTime;    // subarray exposure time
}subArray;
```

Function library *mce.lib*: The user-defined function library is built with the following modules:

mce_lib.c: developed by the user

img_lib.c: developed by the user

dat_lib.c: developed by the user

error.c: provided by CIDTEC

Modules *mce_lib.c, img_lib.c,* and *dat_lib.c* can be divided into smaller ones in order to make maintenance and modification more flexible, if a user feels it is necessary.

Programs *cal.c, mce.c, and img.c*: The objects of these modules must be linked with the following libraries in order to produce the corresponding executable files:

atdaq_c.lib:	provided by National Instruments
scm_dosl.lib:	provided by CIDTEC
mce.lib:	provided by the user

APPENDIX B: FUNCTION LIBRARY *MCE.LIB*

The library includes the functions in the modules *mce_lib.c, img_lib.c,* and *dat_lib.c* which are developed by the user, and those in the module *error.c* which is provided by CIDTEC. Module *error.c* can be found in the SCM5000E Software User's Manual.

```
/*****************************************************************
                "MCE_LIB.C" -- SCM Application Module
                      User-defined Functions
              Developed by  QINGBO LI, ED YEUNG'S GROUP, ISU
                              May, 1994
*****************************************************************
CHANGE HISTORY:
      - NOV94, add function hms(), QL.
*****************************************************************

FUNCTIONS INCLUDED IN THIS MODULE:

      FUNCTION        DESCRIPTION
      initSCM()       initialize SCM
      ldCALparam()    load calibration data to SCM
      rSUBparam()     load subarray parameters into memory
      readSCM()       read subarray data from SCM to host base memory
      readSCM2XMS()   read subarray data from SCM to XMS
      AVGndro()       calculate the averaged readout of NDRO
      fileName()      create a file name to save data
      saveBIN()       save data in binary format
      saveASCII()     save data in ASCII format
      saveCP()        save data in ChromPerfect format
      hms()           calculate increment for exposure time gradient

*****************************************************************/
```

```
/* INCLUDE FILES */

#include                    <stdio.h>
#include                    <malloc.h>
#include                    "scm_mce.h"    // SCM header file

/* DEFINITIONS */

/* GLOBAL VARIABLE DECLARATIONS */

int                         exRSLT;          // result of function call
calDATA                     offcDATA[1];      // data for OFFC cal
calDATA                     offpDATA[512];     // data for OFFP cal
calDATA                     offrDATA[2];      // data for OFFR cal

/* FUNCTION PROTOTYPES */
int   initSCM      (void);
int   ldCALparam (void);
int   rSUBparam  (subArray *sub, unsigned subNum, char *source);
int   readSCM      (subArray subRead, unsigned *buffer);
int   readSCM2XMS      (subArray subRead, int XMShdl);
int   AVGndro      (subArray subRead, unsigned *buffer,
                        unsigned *AVGread);
int   fileName     (char *rootname, char *subName, char *extname,
                        unsigned pixel, char *filename);
int   saveBIN      (unsigned *buffer, unsigned count,
                        char *filename);
int   saveASCII    (unsigned *buffer, unsigned frameNo,
                        char *filename);
int   saveCP       (unsigned *buffer, int points, float sf,
                         char *fileName, char *sampleName, char *methodFile);
void  delay        (unsigned long n);
float hms          (unsigned *etg, unsigned *f);


/*****************************************************************
initSCM(): initializes the SCM after power-on or reset
*****************************************************************
INPUT PARAMETERS:
      - none
RETURN PARAMETERS:
      - if no error:  scmOK.....0
```

```
        - if error:     scmERROR:...1
*****************************************************************/
int     initSCM(void)
{
        scmRSLT = scm_CLReLOG();                // clear SCM error
        if(scm_CKerror(CLReLOG)) return(scmERROR);

        scmRSLT = scm_INITcomm();           // initialize SCM communications
        if(scm_CKerror(INITcomm)) return(scmERROR);

        scmRSLT = scm_RSTscm();         // reset SCM
        if(scm_CKerror(RSTscm)) return(scmERROR);
                                        // download RAM based 286 code.
        scmRSLT = scm_LD286PRG("SCM286.bin", 0x8000, 0x1000, 0x0000,
                                        0x1000, 0xF000);
        if(scm_CKerror(LD286PRG)) return(scmERROR);
        printf("\t\t...RAM based 286-Code loaded\n");

        scmDFLT.actCOL = 512;                   // override default row settings
        scmDFLT.actROW = 512;                   // override default row settings
        scmDFLT.colREF[3]=7.95;         // override default inj. voltage setting
                                        // download SCM operation parameters
        scmRSLT = scm_LDdefPARAMS(scmDFLT);
        if(scm_CKerror(LDdefPARAMS)) return(scmERROR);
        printf("\t\t...Defaults Loaded\n");

        return(scmOK);
}


/*****************************************************************
ldCALdata(): download calibration data to SCM
*****************************************************************
INPUT PARAMETERS:
        - none
RETURN PARAMETERS:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*****************************************************************/
int     ldCALdata(void)
{
unsigned                Pstat;                  // PASS/FAIL status
```

```
unsigned        i;              // index
FILE            *fp;            // file pointer

        fp=fopen("cal.cfg","r");        // open calibration data file
                                // read and download OFFC calibration data
fscanf(fp,"%u\t%u\n",&offcDATA[0].Bdata,& offcDATA[0].stat);
        printf("\t\t...loading OFFC\n");
        scmRSLT = scm_LDoffcCAL(offcDATA, &Pstat);
        if(scm_CKerror(LDoffcCAL)) return(scmERROR);
        if(!Pstat)
        printf("\t\t\t\t...SCM OFFC Calibration Failed\n");
                                // read and download OFFC calibration data
        for(i=0; i<scmDFLT.actROW; i++)
fscanf(fp,"%u\t%u\n",&offpDATA[i].Bdata, &offpDATA[i].stat);
        printf("\t\t...loading OFFP\n");
        scmRSLT = scm_LDoffpCAL(offpDATA, scmDFLT.actROW, &Pstat);
        if(scm_CKerror(LDoffpCAL)) return(scmERROR);
        if(!Pstat)
        printf("\t\t\t\t...SCM OFFP Calibration Failed\n");
                                // read and download OFFR calibration data
        for(i=0; i<2; i++)
fscanf(fp,"%u\t%u\n",&offrDATA[i].Bdata, &offrDATA[i].stat);
        printf("\t\t...loading OFFR\n");
        scmRSLT = scm_LDoffrCAL(offrDATA, &Pstat);
        if(scm_CKerror(LDoffrCAL)) return(scmERROR);
        if(!Pstat)
        printf("\t\t\t\t...SCM OFFR Calibration failed.\n");

printf("\t\t...Calibration Data Loading Complete\n");
        fclose(fp);
        return(scmOK);
}


/*****************************************************************
 rSUBparam(): read subarray parameters from hard disk
*****************************************************************
INPUT PARAMETERS:
        - sub --- the array of pointers to subArray structures
        - num --- number of subarrays to define
        - source --- string of root and file name containing parameters to read
RETURN PARAMETERS:
```

```
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*****************************************************************/
int     rSUBparam(subArray *sub, unsigned num, char *source)
{
FILE        *fp;    // file pointer
unsigned    i;      // index
char        header[80];  // string

if( (fp=fopen(source,"r")) != NULL)
        {
        fgets(header, 80, fp);        // skip the header
for(i=0; i<num; i++) fscanf(fp,"%s\t%u\t%u\t%u\t%u\t%u\t%u\t%li\n",
        &sub[i].name,       // get subarray name
        &sub[i].N,          // get NDRO numbers
        &sub[i].X,          // get subarray starting X coordinate
        &sub[i].Y,          // get subarray starting Y coordinate
        &sub[i].W,          // get subarray width
        &sub[i].H,          // get subarray height
        &sub[i].G,          // get Gain for this subarray
        &sub[i].waitTime    // get exposure time for this subarray (not used)
                        );
        fclose(fp);         // close file
        return(scmOK);
        }
         else {
                printf("\t\t.....Loading subarray parameter failed.\n");
                return(scmERROR);
                }
}


/*****************************************************************
readSCM(): read subarray data from SCM to host base memory
*****************************************************************
INPUT PARAMETERS:
        - subRead --- subarray to read
        - buffer ---  host base memory block
RETURN PARAMETER:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*****************************************************************/
```

```
int     readSCM(subArray subRead, unsigned *buffer)
{
unsigned   j, k;          // index variable

        scmRSLT = scm_LDgain(subRead.G);          // load gain
        if(scm_CKerror(LDgain)) return(scmERROR);
                                        // load number of reads per pixel
        scmRSLT = scm_LDnumREADS(subRead.N);
        if(scm_CKerror(LDnumREADS)) return(scmERROR);
                                        // load subarray starting address
        scmRSLT = scm_LDsubSTART(subRead.X, subRead.Y);
        if(scm_CKerror(LDsubSTART)) return(scmERROR);
                                        // load subarray size
        scmRSLT = scm_LDsubSIZE(subRead.H, subRead.W);
        if(scm_CKerror(LDsubSIZE)) return(scmERROR);
                                    // read subarray data into host base memory
        scmRSLT = scm_SUBread64K(buffer,
                    subRead.W*subRead.H*subRead.N, scmNORMAL);
        if(scm_CKerror(SUBread64K)) return(scmERROR);

        return(scmOK);
}


/*****************************************************************
readSCM2XMS(): read SCM subarray data from SCM to XMS
*****************************************************************
INPUT PARAMETERS:
        - subRead --- subarray to read
        - XMShdl --- handle to the XMS block for data storage
RETURN PARAMETERS:
        - if no error:  scmOK.....0
        - if error:          scmERROR:...1
*****************************************************************/
int     readSCM2XMS(subArray subRead, int XMShdl)
{
unsigned   j, k;                  // index variables
                                        // load gain
        scmRSLT = scm_LDgain(subRead.G);
        if(scm_CKerror(LDgain)) return(scmERROR);
                                        // load number of reads
        scmRSLT = scm_LDnumREADS(subRead.N);
```

```
        if(scm_CKerror(LDnumREADS)) return(scmERROR);
                                        // load subarray start address
        scmRSLT = scm_LDsubSTART(subRead.X, subRead.Y);
        if(scm_CKerror(LDsubSTART)) return(scmERROR);
                                        // load subarray size
        scmRSLT = scm_LDsubSIZE(subRead.H, subRead.W);
        if(scm_CKerror(LDsubSIZE)) return(scmERROR);
                                        // read subarray
        scmRSLT=scm_SUBreadXMS(XMShdl,scmNORMAL);
        if(scm_CKerror(SUBreadXMS)) return(scmERROR);

        return(scmOK);
}



/*****************************************************************
SUBinjADJ(): Clear charge in a subarray
*****************************************************************
INPUT PARAMETERS:
        - subRead --- subarray to inject charge from
        - count ---  number of cycles to inject the charge
RETURN PARAMETER:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*****************************************************************/
int     SUBinjADJ(subArray subRead, unsigned count)
{
unsigned   k;           // index variable
                        // load subarray start address
        scmRSLT = scm_LDsubSTART(subRead.X, subRead.Y);
        if(scm_CKerror(LDsubSTART)) return(scmERROR);
                // load subarray size
        scmRSLT = scm_LDsubSIZE(subRead.H, subRead.W);
        if(scm_CKerror(LDsubSIZE)) return(scmERROR);
                // inject charge in the subarray
for(k=0;k<count;k++){
        scmRSLT = scm_SUBinject();
        if(scm_CKerror(SUBinject)) return(scmERROR);}

        return(scmOK);
}
```

```
/*************************************************************************
int AVGndro(): Calculate the average of NDRO readings for all pixals in a
subarray
*************************************************************************
INPUT PARAMETERS:
        - subRead ---- subarray
        - buffer - host base memory block containing original NDRO data
        - AVGread - averaged data
RETURN PARAMETERS:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*************************************************************************/
int     AVGndro(subArray subRead, unsigned *buffer, unsigned *AVGread)
{
unsigned long *temp;            // pointer to temporary memory block
int           i, k, m;          // index variable

                                // allocate temporary memory block for calculation
        if( (temp = (unsigned long *) (malloc(
                subRead.H*subRead.W*sizeof(long) ))) == NULL){
                printf("\n\t%s\n", ".....Memory allocation failed (temp)");
                free(temp);
                return(scmERROR);}
                // calculate the average readout for each pixel in the subarray
for(k=0;k<subRead.H;k++)
        for(m=0;m<subRead.W;m++)
        {
        temp[k*subRead.W+m]=0;
                for(i=0; i<subRead.N; i++)
        temp[k*subRead.W+m] += buffer[i+subRead.N*(k*subRead.W+m)];
        temp[k*subRead.W+m] /= subRead.N;
        AVGread[k*subRead.W+m] = (unsigned) temp[k*subRead.W+m];
        }
                free(temp);
                return(scmOK);
}


/*************************************************************************
 fileName(): Produce a file name to save an image or a subarray data
```

```
**********************************************************************
INPUT PARAMETERS:
        - rootname --- the directory to save data
        - subName --- the name of the subarray
        - extname --- the extension of the file name
        - pixel --- the pixel in the subarray for which data to be saved
RETURN PARAMETERS:
        - filename --- the name of the file to save dat
NOTE:
        The filename should be in the format such as g:\data\A0015.00r, where
        g:\data\ ia the root name, A is the subarray name, 15 is the pixel # in the
        subarray, and .00r is the extention name.
**********************************************************************/
int     fileName(  char *rootname,  char *subName, char *extname,
                        unsigned pixel, char *filename)
{
char ftemp[5];
        strcpy(filename,rootname);
        strcat(filename,subName);
        if(pixel<10)
        strcat(filename,"000");
        if(pixel<100 && pixel>=10)
        strcat(filename,"00");
        if(pixel<1000 && pixel>=100)
        strcat(filename,"0");
        itoa(pixel,ftemp,10);
        strcat(filename,ftemp);
        strcat(filename, extname);
}


/**********************************************************************
 saveBIN(): Save data onto hard disk in binary format
**********************************************************************
INPUT PARAMETERS:
        - buffer --- the memory block containing the data to save
        - count --- the number of points to save
        - filename --- the file name to save the data
RETURN PARAMETERS:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
**********************************************************************/
```

```
int     saveBIN(unsigned *buffer, unsigned count, char *filename)
{
FILE *fp;

if((fp = fopen( filename, "wb")) != NULL){
        fwrite(&buffer[0], sizeof(buffer[0]), count, fp);
        return(scmOK);
        } else
                {printf("Open file failed (saveBIN).\n");
                 return(scmERROR);}
}


/*************************************************************************
 saveASCII(): Save data onto hard disk in ASCII format
*************************************************************************
PARAMETERS:
        - buffer --- the memory block containing the data to save
        - frameNo --- the number of points in the memory block to save
        - filename --- the file name to save the data
RETURN PARAMETERS:
        - if no error:  scmOK.....0
        - if error:     scmERROR:...1
*************************************************************************/
int     saveASCII(  unsigned *buffer, unsigned frameNo, char *filename)
{
FILE *fps;
unsigned i, j;

        fps=fopen(filename, "w");
if( fps != NULL){
        for(i=0; i<frameNo; i++)
                fprintf(fps,"%u\t\n",buffer[i]); fclose(fps);
                return(scmOK);
                }else
                        {printf("Open ASCII file failed.\n");
                         return(scmERROR);}
}


/*************************************************************************
 delay(): dummy loop for time delay
```

```
**********************************************************************
PARAMETERS:
        - n --- the number of cycles for the dummy loop to repeat
RETURN PARAMETERS:
        - none
**********************************************************************/
void    delay(unsigned long n)
{
float a = 314.159;
unsigned long i;
        for(i=0;i<n;i++) {
                a /= 1000.0;
                a *= 1000.0;  }
}


/**********************************************************************
int saveCP(): save data in Chrom Perfect format
**********************************************************************
PARAMETERS:
        - buffer --- the memory block containing the data to save
        - points --- the number of points to save
        - sf --- sampling frequency
        - fileName --- the file name to save the data
        - sampleName --- the name of the sample
        - methodFile --- the name of the method file to analyse the data in CP
RETURN PARAMETERS:
        - if no error:  0
        - if error:     1
NOTE:
        When the function was written, the exact format of the Chrom Perfect raw
        data file was not known (the manufacturer reserved the right of not
        disclosing the format). The format was decoded by looking at each byte of
        the header in the binary format. The header defined here does not
        necessarily represent the actual header defined by the CP software
        manufacturer (Justice Innovation). The format defined in this function
        has proven to work correctly as long as the purpose of analyzing the data
        file with CP is concerned. For the exact format, please contact the Justice
        Innovation to see if you can get the format for free. If you know the exact
        format, the function may be written in a concise way.
**********************************************************************/
int     saveCP(unsigned *buffer, int points, float sf,
```

```
                    char *fileName, char *sampleName, char *methodFile)
{
void spacerDEF(int spacerSize, char *spacer);

typedef struct
{
char        spacerA[5];
char        sample[60];
char        spacerB[2];
char        time[34];
char        spacerC[42];
char        file[24];
char        spacerD[145];
float       freq;
char        spacerE[4];
int         pointNum;
char        spacerF[22];
}header;

struct ELEsize
{
int         spacerA;
int         sample;
int         spacerB;
int         time;
int         spacerC;
int         file;
int         spacerD;
int         freq;
int         spacerE;
int         pointNum;
int         spacerF;
};

struct ELEsize hs =
{     5,    60,   2,    34,   42,   24,   145,  1,    4,    1,    22 };
                                                                  //344 bytes
char        timebuf[16];
header      hdr;
int         i, id;
FILE        *fp;
float       *tempBuf;
```

```
        if((tempBuf=(float *) malloc((points*sizeof(float))  )) == NULL){
                printf("\t\t\t\t...tempBuf Memory Allocation Failed\n");
                free(tempBuf);
                return(1);}
        spacerDEF(hs.spacerA, hdr.spacerA);
        spacerDEF(hs.spacerB, hdr.spacerB);
        spacerDEF(hs.spacerC, hdr.spacerC);
        spacerDEF(hs.spacerD, hdr.spacerD);
        spacerDEF(hs.spacerE, hdr.spacerE);
        spacerDEF(hs.spacerF, hdr.spacerF);
        spacerDEF(hs.sample, hdr.sample);
        spacerDEF(hs.time, hdr.time);
        spacerDEF(hs.file, hdr.file);
        strcpy(hdr.sample, sampleName);
        _strdate( timebuf );
        strcpy(hdr.time, timebuf);
        _strtime( timebuf );
        strcat(hdr.time, "  " );
        strcat(hdr.time, timebuf);
        strcpy(hdr.file, methodFile);
        hdr.freq = sf;
        hdr.pointNum = points;
        for(i=0;i<hdr.pointNum;i++) tempBuf[i] = (float) buffer[i];
if( (fp = fopen( fileName, "wb" )) != NULL ){
        fwrite(&hdr, sizeof(hdr), 1, fp);
        fwrite(&tempBuf[0], sizeof(tempBuf[0]), hdr.pointNum, fp);
        } else
                {printf( "Write error in saveCP()" );
                 return(1);}
        fclose(fp);
        free(tempBuf);
        return(0);
}

void spacerDEF(int spacerSize, char *spacer)
{int j;
        for(j=0;j<spacerSize;j++)  spacer[j]=0;  }


/*********************************************************************
float hms(): calculate the exposure time gradient increment
```

```
**************************************************************************
INPUT PARAMETERS:
        - etg --- the array containing the exposure time control parameters
                  including the time to start exposure time gradient,
                  the time to stop the run, the initial exposure time,
                  and the final exposure time.
RETURN PARAMETERS:
        - exposure time increment in msec
        - the number of total frames to take
**************************************************************************/
float   hms(unsigned *etg, unsigned *f)
{
float t[6];
int i;
        t[0] = etg[1];
        t[1] = etg[3];
        t[2] = etg[2];
        t[3] = etg[4];
        for(i=0;i<2;i++)t[i]*=1000;
        t[4] = (2*(t[1]-t[0])-t[3]+t[2])/(t[3]+t[2]);
        t[5] = (t[3]-t[2])/t[4];
        t[4] += t[0]/t[2];
        *f=(unsigned)t[4];          // the total number of frames to take
        return (t[5]);              // the exposure time increment in msec
}
```

```
/*****************************************************************
                "IMG_LIB.C" --- SCM Application Module
                        User-defined Functions
                Developed by QINGBO LI, ED YEUNG'S GROUP, ISU
                              Dec., 1993
*****************************************************************
CHANGE HISTORY:

*****************************************************************

FUNCTIONS INCLUDED IN THIS MODULE:

        FUNCTION        DESCRIPTION
        disp()          display image
        initPALLETTE()  set up a linear grey scale;
        clearDISP()     clear screen
        setGRAPHmode()  set graphic mode

MACRO INCLUDED IN THIS MODULE:

        MACRO           DESCRIPTION
        RGB()           convert three 8-bit byte to a packed 24-bit word

*****************************************************************/

/* INCLUDE FILES */

#include <stdio.h>
#include <graph.h>

/* GLOBAL VARIABLE DECLARATIONS */
long int    pal[256];           // pallette


/* FUNCTION PROTOTYPE */
char  disp              (unsigned W, unsigned *image_buf, unsigned scale,
                         int XpSize, int YpSize);
void  initPALLETTE      (void);
void  clearDISP         (void);
int   setGRAPHmode      (void);

/*****************************************************************
```

```
char disp():  display image
*************************************************************************
INPUT PARAMETERS:
        - W --- width of the subarray to display
        - image_buf --- memory block containing the pixel data to diplay
        - scale --- image display scale
        - XpSize --- X direction magnification
        - YpSize --- Y direction magnification
RETURN PARAMETERS:
        - none
*************************************************************************/
char disp(unsigned W, unsigned *image_buf, unsigned scale,
         int XpSize, int YpSize)
{
  int  x,y,x1,x2,y1,y2,i,j;  // index variables
  int  intensity;

    for(y=0;y<1;y++)
    {
        y1=y*YpSize;        // set Y magnification
        y2=YpSize+y1;

        for(x=0;x<W;x++){
           intensity=(image_buf[y*W+x])>>scale;      // adjust scale
           _setcolor((short)intensity); // set color of clear
           x1=x*XpSize;                      // set X magnification
           x2=XpSize+x1;
           _rectangle( _GFILLINTERIOR, x1, y1, x2, y2 );  // display
           }
       }
}


/*************************************************************************
RGB(): converts three 8-bit byte to a packed 24-bit word
*************************************************************************
INPUT PARAMETERS:
        - r:   value for red color (0-255)
        - g:   value for grn color (0-255)
        - b:   value for blu color (0-255)
RETURN PARAMETERS:
        - 24-bit packed word (long)
```

```
*****************************************************************/
#define RGB(r, g, b)   (0x3F3F3FL & ((long)(b) << 16 | (g) << 8 | (r)))


/*****************************************************************
initPALLETTE(): initialize graphic display
*****************************************************************
INPUT PARAMETER:
        - none
RETURN PARAMETER:
        - none
*****************************************************************/
void    initPALLETTE(void)
{
unsigned        i;                              // index variables
short           red, grn, blu;                  // colors used to make pallette
                        // Set up a default linear grey scale palette
        for(i=0; i<256; i++) {
        red = grn = blu = i>>2;
        pal[i] = RGB(red, grn, blu);}
}


/*****************************************************************
clearDISP(): set graphic display to background color
*****************************************************************
DESCRIPTION:
        - initializes graphic display
        - called anytime display needs to be cleared
        - should be in graphics display mode before calling
INPUT PARAMETER:
        - none
RETURN PARAMETER:
        - none
*****************************************************************/
void    clearDISP(void)
{
        _clearscreen(_GCLEARSCREEN);
}


/*****************************************************************
```

```
setGRAPHmode(): set graphic display mode
*****************************************************************
DESCRIPTION:
        - puts display into graphics mode
        - should be in text display mode before calling
RETURN PARAMETERS:
        - error if not able to set mode ..... 1
        - no error ..... 0
*****************************************************************/
int     setGRAPHmode(void)
{
                // set to _MRES256COLOR mode: 320 x 200, 256 colors
        if(! _setvideomode(_MRES256COLOR))
                return(1);
        _remapallpalette(pal);                  // redefine the pallette
        _settextcolor(255);
        return(0);
}
```

```
/*************************************************************************
                "DAT_LIB.C" --- SCM Application Module
                        User-defined Functions
                Developed by QINGBO LI, ED YEUNG'S GROUP, ISU
                              Apr., 1995
*************************************************************************
CHANGE HISTORY:

*************************************************************************

FUNCTION INCLUDED IN THIS MODULE:

        FUNCTION        DESCRIPTION
        blc()           correct baseline for an electropherogram
        mvBox()         moving box smoothing

*************************************************************************/

/* INCLUDED FILES */

#include <stdio.h>
#include <malloc.h>

/* FUNCTION PROTOTYPES */

int     blc         (unsigned *buf, unsigned ptsNum, unsigned secSize,
                     unsigned winSize1, int smSwitch, unsigned winSize2);
int     mvBox       (unsigned *buf, unsigned ptsNum, unsigned winSize);


/*************************************************************************
int blc(): subtract baseline for an electropherogram
*************************************************************************
INPUT PARAMETERS:
        - buf --- the memory block containing data of electropherogram
        - ptsNum --- the number of data points
        - secSize --- the size of sections the eletropherogram divided
        - winSize1 --- the size of the moving box to smooth raw data
        - winSize2 --- the size of the moving box to smooth the minima
        - smSwitch --- the switch deciding whether to smooth minima or not
RETURN PARAMETERS:
        - no error ..... 0
```

```
        - baseline corrected electropherogram
        - if error ..... -1
*********************************************************************/
int     blc(unsigned *buf, unsigned ptsNum, unsigned secSize,
            unsigned winSize1, int smSwitch, unsigned winSize2)
{
int     i, j, k, parts; // index variables
float   *slope;         // the slope of the line connecting two minima
float   temp, tmp[2];   // variables for convenience
unsigned    *TMPbuf;    // array pointer
unsigned    *BL_ht;     //  pointer to the array containing minima
unsigned    *BL_pt;     // pointer to the array containing
                        // the positions of the minima

        if((TMPbuf=(int *) malloc(ptsNum*sizeof(int)) ) == NULL)
                    return(-1);
        for(i=0;i<ptsNum;i++) TMPbuf[i]= buf[i];      // swab data
        mvBox(TMPbuf, ptsNum, winSize1);              // smooth data
        parts=ptsNum/secSize;                   // number of sections to divide
                                            // buffers to store minima
        if((BL_ht= (int *) malloc((2+parts)*sizeof(int)) ) == NULL) return(-1);
        if((BL_pt= (int *) malloc((2+parts)*sizeof(int)) ) == NULL) return(-1);

        BL_pt[0]=0;
        BL_ht[0]=TMPbuf[0];
        BL_pt[1+parts]=ptsNum-1;
        BL_ht[1+parts]=TMPbuf[ptsNum-1];

        for(i=1;i<(parts+1);i++)        // search for minima in smoothed data
        {
          BL_ht[i]=32000;
          for(j=0;j<secSize;j++)
          {  k = j+(i-1)*secSize;
               if(TMPbuf[k]<BL_ht[i])
               { BL_ht[i]=TMPbuf[k];
                 BL_pt[i]=k; }
          }
        }
                    // smooth minima if chosen to do so
        if(smSwitch==1 && parts>winSize2)mvBox(BL_ht, parts+2, winSize2);
                    // allocate array to store slopes
        if((slope=(float *) malloc((parts+2)*sizeof(float)) ) == NULL)
```

```
                return(-1);
                // build baseline using minima and slopes
        for(i=0;i<(parts);i++)
         {
          if(BL_pt[i]==BL_pt[i+1])goto noConnection;
          tmp[0]=BL_ht[i];
          tmp[1]=BL_ht[i+1];
          slope[i]=(tmp[1]-tmp[0])/(BL_pt[i+1]-BL_pt[i]);
          for(j=0;j<(BL_pt[i+1]-BL_pt[i]);j++)
            {
             tmp[0]=(float)j;
             tmp[0]*=slope[i];
             TMPbuf[BL_pt[i]+j]=BL_ht[i]+((unsigned)tmp[0]);
             }
          noConnection:          // if two minima overlap, no line in between
          ;
          }
                                 // subtract baseline
        for(i=0;i<ptsNum;i++)
         {
          if(buf[i]>TMPbuf[i])
            buf[i]-=TMPbuf[i];
          else
            buf[i]=1;            // if a point is <= 0, set to 1.
         }
                                 // free memory blocks
        free(TMPbuf);
        free(slope);
        free(BL_ht);
        free(BL_pt);
        return(0);
}


/*****************************************************************************
int mvBox(): moving box smoothing
*****************************************************************************
INPUT PARAMETERS:
        - buf --- data to smooth
        - ptsNum --- number of data points
        - winSize --- moving box size, should be odd
```

```
RETURN PARAMETERS:
        - no error ..... 0
        - smoothed data
        - if error ..... -1
**********************************************************************/
int mvBox(unsigned *buf, unsigned ptsNum, unsigned winSize)
{
int i, j, NL, NR, *box;       // index variables
long sum;

   if( (box=(int *) malloc(winSize*sizeof(int))) == NULL) return(-1);
   NL = (winSize-1)/2;   // winSize should be odd
   NR = NL;
   for(i=NL;i<(ptsNum-NR);i++)
   {  sum=0;
      for(j=0;j<winSize;j++) box[j]=buf[i-NL+j];
      for(j=0;j<winSize;j++) sum += box[j];
      buf[i]=sum/winSize;}
   free(box);
   return(0);
}
```

APPENDIX C: PROGRAM *CAL.C*

```
/*****************************************************************
                    "CAL.C" -- SCM Application Module
                      SCM5000E Calibration Program
                 Developed by QINGBO LI, ED YEUNG'S GROUP, ISU
                                Dec., 1993
*****************************************************************
CHANGE HISTORY:

*****************************************************************/

/* INCLUDE FILES */
#include     <stdio.h>
#include     "scm_mce.h"              // SCM header file

/* DEFINITIONS */

/* GLOBAL VARIABLE DECLARATIONS */
int          exRSLT;                  // result of function call
calDATA      offcDATA[1];              // data for OFFC cal
calDATA      offpDATA[512];            // data for OFFP cal
calDATA      offrDATA[2];             // data for OFFR cal

FILE *fp;                             // file pointer

/* FUNCTION PROTOTYPES */
int  ckEXerr       (char *msg); //  get error message, in scm_.lib
int  initSCM       ();          // initialize SCM, in mce.lib
int  cal           ();          // calibrate SCM, in this module
int  main          (void);

/*****************************************************************
calSCM(): performs calibration on SCM & downloads data to SCM
*****************************************************************
DESCRIPTION:
            - performs OFFP & OFFR calibrations
            - downloads OFFP & OFFR data to SCM
RETURN PARAMETER:
```

```
                - if no error:          scmOK.....0
                - if error:             scmERROR...1
*****************************************************************/
int                     calSCM(void)
{
unsigned    Pstat;                      // load PASS/FAIL status
unsigned    i;                          // index

                                        // close shutter
                scmRSLT = scm_CTRLshutter(3, scmWAIT);
                if(scm_CKerror(CTRLshutter)) return(scmERROR);

                                        // calibrate OFFC
                printf("\t\t...calibrating OFFC\n");
                scmRSLT = scm_GEToffcCAL(200, (float)0.25, offcDATA,
                                scmDFLT.actCOL, scmDFLT.actROW);
                if(scm_CKerror(GEToffcCAL)) return(scmERROR);

                                        // load OFFC
                printf("\t\t...loading OFFC\n");
                scmRSLT = scm_LDoffcCAL(offcDATA, &Pstat);
                if(scm_CKerror(LDoffcCAL)) return(scmERROR);

                                        // save OFFC calibration data to hard disk
fprintf(fp,"%u\t%u\n",offcDATA[0].Bdata, offcDATA[0].stat);

                if(!Pstat)
                        printf("\t\t\t\t...SCM OFFC Calibration Failed\n");

                        // calibrate OFFP printf("\t\t...calibrating OFFP\n");
                scmRSLT = scm_GEToffpCAL(200, (float)0.25, offcDATA,
                                offpDATA, scmDFLT.actCOL, scmDFLT.actROW);
                if(scm_CKerror(GEToffpCAL)) return(scmERROR);

                        // load OFFP printf("\t\t...loading OFFP\n");
                scmRSLT = scm_LDoffpCAL(offpDATA, scmDFLT.actROW,
                                                &Pstat);
                if(scm_CKerror(LDoffpCAL)) return(scmERROR);

                        // save OFFP calibration data to hard disk
        for(i=0; i<scmDFLT.actROW; i++)
fprintf(fp,"%u\t%u\n",offpDATA[i].Bdata, offpDATA[i].stat);
```

```
            if(!Pstat)
                printf("\t\t\t\t...SCM OFFP Calibration Failed\n");

                // calibrate OFFR printf("\t\t...calibrating OFFR\n");
            scmRSLT = scm_GEToffrCAL(200, (float)0.25, (float)0.5,
                    offcDATA, offrDATA,
                    scmDFLT.actCOL, scmDFLT.actROW);
            if(scm_CKerror(GEToffrCAL)) return(scmERROR);

                // load OFFR printf("\t\t...loading OFFR\n");
            scmRSLT = scm_LDoffrCAL(offrDATA, &Pstat);
            if(scm_CKerror(LDoffrCAL)) return(scmERROR);

                // save OFFR calibration data
        for(i=0; i<2; i++)
fprintf(fp,"%u\t%u\n",offrDATA[i].Bdata, offrDATA[i].stat); if(!Pstat)
                printf("\t\t\t\t...SCM OFFR Calibration Failed\n");
            printf("\t\t...Calibration Complete\n");
            return(scmOK);
}


/*********************************************************************
int main(void): calibrate SCM when the CID is turned on or reset
*********************************************************************/
int main(void)
{                   // open file to save calibration data
        if( (fp=fopen("cal.cfg","w")) !=NULL){
            printf("\n\nInitializing\n");
            exRSLT=initSCM();
            if(exRSLT) goto errExit;
            printf("\nCalibrating\n");
            exRSLT=calSCM();
            if(exRSLT) goto errExit;
        } else {        printf("Error: open file failed\n");
                        goto errExit;                   }
        exit(0);
errExit:
        printf("Error occurs, calibration not completed\n");
        exit(-1);
}
```

APPENDIX D: PROGRAM *IMG.C*

```
/*****************************************************************
                    "IMG.C" -- SCM Application Module
                    Image Aquisitoin and Display Program
                  Developed by QINGBO LI, ED YEUNG'S GROUP, ISU
                                Dec., 1993
*****************************************************************
CHANGE HISTORY:
        - 06SEPT94, add offsets w_off, h_off , QL.
        - 16DEC94, add background subtraction, QL.
*****************************************************************/
#pragma check_pointer(off)

/* INCLUDE FILES */
#include      <process.h>
#include      <stdlib.h>
#include      <conio.h>
#include      <stdio.h>
#include      <malloc.h>
#include      <graph.h>
#include      "scm_mce.h"         // user's header file.
        // should be in the same directory unless you specify the route.

/* DEFINITIONS */
#define       subNum 1    // number fo subarray to read and display.
        // for subNum>1, some modification should be made to the program.

/* GLOBAL VARIABLE DECLARATIONS */
int           exRSLT;             // result of function call
long int      pal[256];           // pallette
unsigned      GRAPHmode =0;
calDATA       offcDATA[1];        // data for OFFC cal
calDATA       offpDATA[512];      // data for OFFP cal
calDATA       offrDATA[2];        // data for OFFR cal
struct        videoconfig screen;       // screen configuration
subArray      sub[subNum];              // array of subarray structure


/*****************************************************************
int main(): acquire and display images
```

```
**********************************************************************/
int main()
{
unsigned x,y,i,h,num,m;          // index variables
unsigned expTime;          // exposure time
unsigned scale, XpSize, YpSize;          // intensity scale and magnification
unsigned *Bbuf, *bkgBbuf, *AVGread, *AVGbkg;  // pointers to memory block
char    c;                  // character for acepting keyboard instruction
int     XMShdl, BKGhdl;  // handles for storing images
int     w_off, h_off;          // image display offset

        printf("\n\n\t\t***** SCM5000E Application Code ***** \n");
        printf("\t\t  ***  IMAGE FOCUSING SOFTWARE  ***\n");

        printf("\nInitializing SCM:\n");
        exRSLT = initSCM();
        if(ckEXerr("initSCM"))  goto errEXIT;

        printf("\n\nLoading calibration data to SCM:\n");
        exRSLT = ldCALdata();
        if(ckEXerr("ldCALdata"))   goto errEXIT;
                        // load subarray parameters
        scmRSLT = rSUBparam(sub, subNum,
                                        "f:\\qc25\\scm\\subparam.cfg");
        if(ckEXerr("rSUBparam"))   goto errEXIT;

        printf("\n\t\t...Reading SCM Subarray :\n");
              // set default values to the following parameters
        expTime=100;        // exposure time
        scale=4;            // display scale
        XpSize=1;           // X direction manification
        YpSize=1;           // Y direction magnification
        w_off = 0;          // X direction offset
        h_off = 0;          // Y direction offset
        goto Background;    // acquire background image

CHANGEgain:
    printf("The old gain is %i ; enter the new gain (cover lens): \n", sub[0].G);
        scanf("%i", &sub[0].G);
        goto Background;

CHANGEexposureTime:
```

```
    printf("The old exposure time is %i ms.\nEnter the new exposure time: ",
            expTime);
        scanf("%i", &expTime);
        goto acquire;

CHANGEstartpoint:
    printf("\nThe old start position is X= %d, Y= %d \n",sub[0].X,sub[0].Y);
        printf("Enter the new start position (cover lens): ");
        scanf("%d %d",&sub[0].X, &sub[0].Y);
        goto Background;

Subarraysize:
    printf("\nThe old subarray size is W= %d, H= %d \n",sub[0].W,sub[0].H);
        printf("Enter the new subarray size (cover lens): ");
        scanf("%d %d",&sub[0].W, &sub[0].H);
        goto Background;
                                    // acquire background image
Background:
        scmRSLT = scm_XMSmemALLOC((int far*)&BKGhdl,
                                (long) sub[0].H*sub[0].W*sub[0].N);
        if(scm_CKerror(XMSmemALLOC)) {
           printf("\t\t\t\t...Read Memory Allocation Failed\n");
           goto errEXIT;   }
                                // close shutter
        scmRSLT = scm_CTRLshutter(3, scmWAIT);
        if(scm_CKerror(CTRLshutter)) return(scmERROR);
                                // clear charge
        scmRSLT = scm_GLBinject(100);
        if(scm_CKerror(GLBinject)) return(scmERROR);
                                // delay for exposure time
        scm_WAITmsec(expTime);
                                // read background image
        scmRSLT = readSCM2XMS(sub[0], BKGhdl);
        if(ckEXerr("readSCM2XMS"))  goto errEXIT;

acquire:
        scmRSLT = scm_XMSmemALLOC((int far*)&XMShdl,
                                 (long) sub[0].H*sub[0].W*sub[0].N);
        if(scm_CKerror(XMSmemALLOC))   {
           printf("\t\t\t\t...Read Memory Allocation Failed\n");
           goto errEXIT;   }
                                // clear charge
```

```
        scmRSLT = scm_GLBinject(100);
        if(scm_CKerror(GLBinject)) return(scmERROR);
                                // open shutter
        scmRSLT = scm_CTRLshutter(0, scmWAIT);
        if(scm_CKerror(CTRLshutter)) return(scmERROR);
                                // expose
        scm_WAITmsec(expTime);
                                // close shutter
        scmRSLT = scm_CTRLshutter(3, scmWAIT);
        if(scm_CKerror(CTRLshutter)) return(scmERROR);
                                // read image
        scmRSLT = readSCM(sub[0], XMShdl);
        if(ckEXerr("readSCM"))  goto errEXIT;
                                // display image
        goto display;

CHANGEintensity:
        printf("The old intensity scale is %i\n", scale);
        printf("Enter the new intensity scale: ");
        scanf("%i",&scale);
        goto display;

CHANGEpixalsize:
        printf("The old pixel size (X x Y) is %i x %i\n", XpSize, YpSize);
        printf("Enter the new pixelSize:  ");
        scanf("%d %d",&XpSize, &YpSize);
        goto display;

CHANGEoffset:
        printf("The old offset (w_offset  h_offset) is %i x %i\n", w_off, h_off);
        printf("Enter the new offset:  ");
        scanf("%d %d",&w_off, &h_off);
        goto display;

display:
                                // set monitor for displaying image
        if(c != 'm')
        {
        initPALLETTE(0);
        setGRAPHmode();
        clearDISP();
         }
```

```
        // allocate memory block for transferring image data
    if((Bbuf=(unsigned int *) malloc(
        (sub[0].W*sub[0].N*sizeof(int)))) == NULL)
    {
        printf("\t\t\t...Read Memory Allocation Failed\n");
        free(Bbuf);
        return(scmERROR);
    }
        // allocate memory block for transferring background image data
    if((bkgBbuf=(unsigned int *) malloc(
        (sub[0].W*sub[0].N*sizeof(int)))) == NULL)
    {
        printf("\t\t\t...Read Memory Allocation Failed\n");
        free(bkgBbuf);
        return(scmERROR);
    }
        // allocate memory block for calculating NDRO of image
    if((AVGread=(unsigned int *) malloc(
        (sub[0].W*sizeof(int)))) == NULL)
    {
        printf("\t\t\t...Read Memory Allocation Failed\n");
        free(AVGread);
        return(scmERROR);
    }
        // allocate memory block for calculating NDRO of bkg image
    if((AVGbkg=(unsigned int *) malloc(
        (sub[0].W*sizeof(int)))) == NULL)
    {
        printf("\t\t\t...Read Memory Allocation Failed\n");
        free(AVGbkg);
        return(scmERROR);
    }
                // display background subtracted image
for(h=0;h<sub[0].H;h++)
 {              // read image data from XMS
   scmRSLT=scm_XMSmemCOPY(XMShdl,
                Bbuf,
                (long) sub[0].W*sub[0].N*h,
                (long) sub[0].W*sub[0].N,
                scmXMS2BASE);
   if(scm_CKerror(XMSmemCOPY)) goto errEXIT;
                // read bkg image data from XMS
```

```
scmRSLT=scm_XMSmemCOPY(BKGhdl,
                       bkgBbuf,
                       (long) sub[0].W*sub[0].N*h,
                       (long) sub[0].W*sub[0].N,
                       scmXMS2BASE);
if(scm_CKerror(XMSmemCOPY)) goto errEXIT;
                // calculate NDRO average
exRSLT = AVGndro(sub[0].W, sub[0].N, Bbuf, AVGread);
if(ckEXerr("AVGndro")) goto errEXIT;
                // calculate NDRO average
exRSLT = AVGndro(sub[0].W, sub[0].N, bkgBbuf, AVGbkg);
if(ckEXerr("AVGndro")) goto errEXIT;
                // subtract background
for(m=0;m<sub[0].W;m++)
 AVGread[m] =  (AVGread[m] + 64) - AVGbkg[m] ;
                // set display offset
_setvieworg(w_off*XpSize, (h+h_off)*YpSize );
                // display image
disp( sub[0].W, AVGread, scale, XpSize, YpSize );
}
        // if no instruction from keyboard, clear charge every 5 sec
do
 {   scmRSLT = scm_GLBinject(5);
       if(scm_CKerror(GLBinject)) return(scmERROR);
       scm_WAITmsec(5000);
  }
  while(!kbhit());
        // read instruction from keyborad
c = getch();
        // identify the instruction
if(c != 'm')            // reset screen if not a repeated image acquisition
{
_clearscreen( _GCLEARSCREEN);
_setvideomode( _DEFAULTMODE );
 }
                        // release memory blocks
free(Bbuf);
free(bkgBbuf);
free(AVGread);
free(AVGbkg);
                        // further identify the instruction
switch(c)
```

```
{
      case 'g':
            scmRSLT = scm_XMSmemFREE(XMShdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            scmRSLT = scm_XMSmemFREE(BKGhdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            goto CHANGEgain;
            break;
      case 'e':
            scmRSLT = scm_XMSmemFREE(XMShdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            goto CHANGEexposureTime;
            break;
      case 's':
            scmRSLT = scm_XMSmemFREE(XMShdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            scmRSLT = scm_XMSmemFREE(BKGhdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            goto CHANGEstartpoint; // change subarray position
            break;
      case 'a':
            scmRSLT = scm_XMSmemFREE(XMShdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            scmRSLT = scm_XMSmemFREE(BKGhdl);
            if(scm_CKerror(XMSmemFREE)) goto errEXIT;
            goto Subarraysize;          // change subarray size
            break;
case 'o':
        goto CHANGEoffset;              // change display offset
        break;
case 'p':
        goto CHANGEpixalsize;           // change magnification
        break;
case 'i':
        goto CHANGEintensity;           // change display scale
        break;
case 'm':
        scmRSLT = scm_XMSmemFREE(XMShdl);
        if(scm_CKerror(XMSmemFREE)) goto errEXIT;
        goto acquire;                   // acquire new image
        break;
case 'q':
```

```
                scmRSLT = scm_XMSmemFREE(XMShdl);
                if(scm_CKerror(XMSmemFREE)) goto errEXIT;
                scmRSLT = scm_XMSmemFREE(BKGhdl);
                if(scm_CKerror(XMSmemFREE)) goto errEXIT;
                goto interuptEXIT;              // quit the program
                break;
        default:
                goto display;       // refresh the screen
                break;
        }

errEXIT:                        // errors occur, quit the program
        scmRSLT = scm_CTRLshutter(3, scmWAIT);
        free(Bbuf);
        free(bkgBbuf);
        free(AVGread);
        free(AVGbkg);
        scmRSLT = scm_XMSmemFREE(XMShdl);
        exit(-1);

interuptEXIT:               // terminated by user
        scmRSLT = scm_CTRLshutter(3, scmWAIT);
        free(Bbuf);
        free(bkgBbuf);
        free(AVGread);
        free(AVGbkg);
        scmRSLT = scm_XMSmemFREE(XMShdl);
        exit(1);
}
```

APPENDIX E: PROGRAM *MCE.C*

```
/*****************************************************************
                    "MCE.C" -- SCM Application Module
                    Multiplexed CE Application Program
                Developed by QINGBO LI, ED YEUNG'S GROUP, ISU
                                May, 1994
*****************************************************************
CHANGE HISTORY
        - DEC94, incorporate function hms(), QL.
        - MAR95, reorganize the input sequence, QL.
        - 28April95, add baseline subraction, QL.
*****************************************************************/

/* INCLUDE FILES */
#include      <process.h>
#include      <stdlib.h>
#include      <conio.h>
#include      <stdio.h>
#include      <malloc.h>
#include      <time.h>
#include      <math.h>
#include      "scm_mce.h"

/* DEFINITIONS */
#define maxSubNum 16
/** maxSubNum is the maximun number of subarrays allowed in this program.
Can be set to higher value if necessary. There's no upper limit. The only limit is
the available memory.  **/

/* GLOBAL VARIABLE DECLARATIONS */
int           exRSLT;           // result of function call
calDATA    offcDATA[1];         // data for OFFC cal
calDATA    offpDATA[512];       // data for OFFP cal
calDATA    offrDATA[2];         // data for OFFR cal
unsigned   subNum;              // number of subarrays to read

/* ASCII STRING DECLARATIONS */
char        revMSG[] = {"V2.1 APRIL95QL"};

/* FUNCTIONS */
```

```
int    main        ();


/*************************************************************************
int main():  monitor fluorescence from the detection window of capillaries
*************************************************************************/
int         main()
{
subArray    sub[maxSubNum]; // array of subarray structures.
unsigned    maxFrameNo;       // maximun frame # allowed.
unsigned    winSize1;     // size of moving box for smoothing raw data in blc()
unsigned    winSize2;     // size of moving box for smoothing minima in blc()
unsigned    smSwitch;     // switch deciding whether smooth or not
unsigned    frame;        // frame number index
unsigned    frmDelay;     // frame # when exposure time gradient (ETG) on
unsigned    subSize;      // subarray size
unsigned    secSize;      // section size for searching minima in blc()
unsigned    i, m, n, p, max, avgIndx;  // index variables
unsigned    *Bbuf;  // pointer to memory block (MB) for reading subarray...
unsigned    *AVGread;  // pointer to MB for calculating NDRO
unsigned    *saveBuf;     // pointer to MB for sorting out data to save
unsigned    etg[8];       // array containing ETG parameters
char        stemp[100]; // string for temporary use
char        fname[40];  // filename to save data
char        rname[20];  // route name for saving data
char        extname[5]; // extension of the files to save data
char        sample[20]; // the name of the sample injected
char        method[10]; // the name of the CP method file to analyse data
FILE        *fps, *fp, *fmax;     // file poiters
int         XMShdl[16];       // XMS handles
int         blcSwitch[2];       // switches deciding whether correct baseline
time_t      rtime1, rtime2, iniETGtime;      // time variables
long        expTimeCounts;    // cycles to repeat the dummy loop in delay()
float       smpf;         // average sampling frequency
float       hCounts;   // cycles to repeat the dummy loop in delay() for ETG
strcpy(     rname, "g:\\data\\");
strcpy(     extname, ".00r");

        printf("\n\n\t\tSCM5000E Code -- %s\n", revMSG);
                        // read input parameters from hard disk
        if( (fp=fopen("f:\\qc25\\scm\\miscparm.cfg","r")) == NULL ){
```

```
        printf("\n**** Reading f:\\qc25\\scm\\miscparm.cfg failed.");
        exit(-1);}

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%s", &sample);

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%s", &method);

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &subNum);

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &etg[1]);     // the time when ETG begins
etg[1] *=60;

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &etg[2]);     // initial exposure time

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &etg[3]);     // the time to stop the run
etg[3] *= 60;

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &etg[4]);     // final exposure time

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%u", &etg[0]);     // cycles of delay() loop per msec

fgets(stemp, sizeof(stemp), fp);
sscanf(stemp, "%d", &blcSwitch[0]);     // baseline correction ?
blcSwitch[1]=0;
              // if decide to correct baseline, read in the parameters
if(blcSwitch[0]>0)
{
 fgets(stemp, sizeof(stemp), fp);
 sscanf(stemp, "%d", &secSize);

 fgets(stemp, sizeof(stemp), fp);
 sscanf(stemp, "%d", &winSize1);

 fgets(stemp, sizeof(stemp), fp);
```

```
        sscanf(stemp, "%d", &smSwitch);

        fgets(stemp, sizeof(stemp), fp);
        sscanf(stemp, "%d", &winSize2);
        }

      fclose(fp);     // close the input data file

                      // calculate the cycles of delay()  for exposure
      expTimeCounts = ((long)etg[2])*((long)etg[0]);
      hCounts = hms(etg,&maxFrameNo)*((float)etg[0]);
      maxFrameNo += 50;        // allow 50 overhead for maximun frame #

      printf("\nInitializing SCM:\n");
      exRSLT = initSCM();
      if(ckEXerr("initSCM"))     goto errEXIT;

      printf("\n\nLoading calibration data to SCM:\n");
      exRSLT = ldCALdata();
      if(ckEXerr("ldCALdata"))   goto errEXIT;

                           // read in subarray parameters
      scmRSLT = rSUBparam(sub, subNum,
                           "f:\\qc25\\scm\\subparam.cfg");
      if(ckEXerr("rSUBparam"))   goto errEXIT;

                           // allocate XMS blocks to save images
      for(i=0;i<subNum;i++)
      {
      scmRSLT = scm_XMSmemALLOC((int far*)&XMShdl[i],
                           ((long) (sub[i].H*sub[i].W))*((long) maxFrameNo));
      if(scm_CKerror(XMSmemALLOC))
         {
         printf("\t\t\t\t...Read Memory Allocation Failed\n");
         goto errEXIT;                                        }
       }

      printf("\n\nReading SCM subarrays:\n");
      printf("\t\t...press F2 to start\n");

                      // wait for F2 key to be pressed to begin data acquisition
hitagain:
```

```
        if( getch() == 0x3c)
        {
         printf("\t\t...echo F2\n");
         printf("\t\t...press F10 to stop at any time\n\n");
         goto startSCM;
         }
        else
         goto hitagain;
                              // begin data acquisition
startSCM:
                              // open shutter
     scmRSLT = scm_CTRLshutter(0, scmWAIT);
     if(scm_CKerror(CTRLshutter)) goto errEXIT;

    time(&rtime1);            // get system time
    time(&rtime2);            // get system time
    frmDelay = 0;             // reset frmDelay
                              // the loop of acquiring images
                              // if time is due, terminate image acquisition
    for(frame=0; (rtime2-rtime1) < etg[3]; frame++)
    {
     if(kbhit())
        if(getch() == 0x44)       // if F10 key is pressed, terminate the program
          goto DONE;
                                  // clear charge
     scmRSLT = scm_GLBinject(100);
     if(scm_CKerror(GLBinject)) goto errEXIT;
                                  // calculate the exposure time
     if((rtime2-rtime1) <= etg[1])
         {
          expTimeCounts=((long)etg[2])*((long)etg[0]);
          frmDelay = frame;
          }
        else
          expTimeCounts = ((long)etg[2])*((long)etg[0])
                        + ((long)hCounts)*((long)(frame-frmDelay));

     delay(expTimeCounts);        // delay for exposure

     printf("%ld sec\n",rtime2-rtime1); // show the elapsed time on screen
                              // read the subarrays sequentially
     for(i=0;i<subNum;i++)
```

```
    {
      subSize=sub[i].H*sub[i].W;     // calculate the subarray size
                    // allocate the MB according to the subarray size
      if((Bbuf=(unsigned int *) malloc(
      (subSize*sub[i].N*sizeof(int))  )) == NULL)
        {
          printf("\t\t\t\t...Bbuf Memory Allocation Failed\n");
          free(Bbuf);
          goto errEXIT;
          }
                    // allocate the MB according to the subarray size
      if((AVGread=(unsigned int *) malloc(
      (subSize*sizeof(int))  )) == NULL)
        {
          printf("\t\t\t\t...AVGread Memory Allocation Failed\n");
          free(AVGread);
          goto errEXIT;
          }
                          // read the subarray
      exRSLT = readSCM(sub[i], Bbuf);
      if(ckEXerr("readSCM"))  goto errEXIT;
                          // calculate the average of NDRO
      exRSLT = AVGndro(sub[i], Bbuf, AVGread);
      if(ckEXerr("AVGndro"))  goto errEXIT;
                          // put data into XMS
      scmRSLT=scm_XMSmemCOPY(XMShdl[i],
                           AVGread,
                           ((long) subSize)*((long) frame),
                           (long) subSize,
                           scmBASE2XMS);
      if(scm_CKerror(XMSmemCOPY))
          { printf("\n****** XMS full. Data aquisition terminated\n");
            goto DONE;
           }
                          // free memory blocks for resizing
      free(Bbuf);
      free(AVGread);
     }
     time(&rtime2);     // get system time
  }
                    // image acqusition finished. Begin to save data
DONE:
```

```
        scmRSLT = scm_CTRLshutter(3, scmWAIT);
        if(scm_CKerror(CTRLshutter)) goto errEXIT;
        printf("The time for %d frames is: %d\n", frame,(rtime2-rtime1));
        printf("\n\nSaving data, please wait.....\n\n");

/******
      Save data: at this point, the data are still in XMS, inadvertent
      termination of the program will cause the disaster of losing all of the data.
      Therefore, during data collection, make sure the electric power in the
      room is secured. Don't temper the instrument at this time. Avoid any
      electric surge in the room which could cause the CID to send back a false
      message to terminate the program. If the hard disk is full, the program
      will remind you to assign a new route to save the data. That will be a
      risk. Don't let that happen. Check available hard disk space before you
      run the program.
******/

secure:
              // calculate the average sampling frequency
   smpf =  ((float)frame)/((float)(rtime2-rtime1));
              // allocate MB for retreaving data from XMS
   if((Bbuf=(unsigned int *) malloc(
          (sizeof(int))  )) == NULL)
           {
             printf("\t\t\t\t...Read Memory Allocation Failed\n");
             free(Bbuf);
             goto errEXIT;
             }
             // allocate MB for sorting data to save
   if((saveBuf = (unsigned int *) malloc(
          (frame*sizeof(int))  )) == NULL)
           {
             printf("\t\t\t\t...Memory Allocation Failed\n");
             free(saveBuf);
             goto errEXIT;
             }
      strcpy(stemp, rname); // assign a file to save maxima in pixels
      strcat(stemp, "maxima.dat");
   fmax = fopen(stemp, "w"); // open the file to save the maxima in the pixels
                  // save the data in each subarray sequentially
   for(i=0;i<subNum;i++)
   {
```

```
    subSize=sub[i].H*sub[i].W; // calculate subarray size

/****
      Build electropherogram for each pixel: in this program, the data in every
      pixel in a subarray is saved. The useful data are sorted out later. In the
      future, if the focusing configuration is decided and not changed frequently,
      the pixels containing useful information can be sorted out here. Only the
      useful data will be saved in order to save time and space.
****/
                    // scan each pixel in the subarray sequentially
    for(m=0; m<subSize; m++){
                    // collect the intensity of this pixel in all frames
       for(n=0;n<frame;n++){
        scmRSLT=scm_XMSmemCOPY(XMShdl[i], Bbuf,
                    ( ((long) subSize) * ((long) n) + ((long) m) ),
                     1, scmXMS2BASE);
        if(scm_CKerror(XMSmemCOPY)) goto errEXIT;
        saveBuf[n] = Bbuf[0];
       }
        saveBuf[0] = saveBuf[1];
                                    // correct baseline
   if(blcSwitch[0]==1 && blcSwitch[1]==1)
        bsl(saveBuf,frame,secSize,winSize1,smSwitch,winSize2);
                                    // find out the maximun in the pixle
        max = 0;
        for(p=10;p<frame-10;p++)
          if(saveBuf[p]>max) max=saveBuf[p];
                            // produce the file name to save the data
        exRSLT=fileName(rname, sub[i].name, extname, m, fname);
                            // save the data in CP format (or other foramts)
        exRSLT=saveCP(saveBuf, frame, smpf, fname, sample, method);
        if(ckEXerr("saveCP"))
         {
          printf("\t\t.....Saving files failed\n");
          printf("\t\tEnter new file root name: ");
          scanf("%s", &rname);
          goto secure;
          }

        fprintf(fmax, "%s\t%d\n", fname, max);
       }
    }
```

```
    fclose(fmax);
    free(Bbuf);
    free(saveBuf);
                        // save the exposure parameters for later reference
        if(etg[1]==0)
          {
          etg[5]= 0;
          etg[6]= frame;
          if(etg[1]==etg[3])
                etg[7]= (etg[3]*1000)/frame;
          else
                etg[7]= etg[2];
          }
        else
          {
          etg[5]=frmDelay;
          etg[6]=frame;
          etg[7]= (etg[1]*1000)/frmDelay;
          }

          strcpy(stemp, rname);
          strcat(stemp,"exposure.dat");
          fp=fopen(stemp, "w");
            for(i=0;i<8;i++) fprintf(fp, "%d\n", etg[i]);
            fclose(fp);
          strcpy(stemp, rname);
          strcat(stemp,"freq.dat");
          fp=fopen(stemp, "w");
            fprintf(fp, "%d\n", frame);
            fprintf(fp, "%d\n", rtime2-rtime1);
            fprintf(fp, "%9.6f\n", smpf);
            fclose(fp);

        if(blcSwitch[0]==1 && blcSwitch[1]==0)
        {
           strcpy(rname,"g:\\blc\\");
           strcpy(extname,".01r");
           blcSwitch[1]=1;
printf("\n\nSaving baseline subtraction results. Please wait.....\n\n");
           goto secure;
          }
```

```
                    // free memory block before quitting the program
   for(i=0;i<subNum;i++)
    {
       scmRSLT = scm_XMSmemFREE(XMShdl[i]);
       if(scm_CKerror(XMSmemFREE)) goto errEXIT;
       }

       printf("\n\nProgram Complete\n");
       exit(0);

errEXIT:
       scmRSLT = scm_CTRLshutter(3, scmWAIT);
   for(i=0;i<subNum;i++)
       scmRSLT = scm_XMSmemFREE(XMShdl[i]);
       exit(-1);
}
```

APPENDIX E: FORMAT OF THE INPUT FILES

miscpara.cfg

PGEM/U	sample name
mce.met	method file for CP analysis
2	# of subarrays to read (subarray parameters are in *subparam.cfg*)
1	Tr (in min) at which exposure time gradient begins
1000	starting exposure time (in msec)
2	Tr (in min) at which the run will stop
1000	final exposure time (in msec) when the run stops
234	number of cycles of *delay()* to produce 1ms delay (depends on CPU)
-1	baseline correction switch. -1 -- No; 1 -- Yes.
100	section size for minima searching in function *blc()*
3	moving box size for raw data smoothing in function *blc()*
-1	minima points smoothing switch in function *blc()*
5	moving box size for smoothing minima points in function *blc()*

subparam.cfg and *imgparam.cfg*

name	N	X	Y	W	H	G	dummy
a	1	100	100	5	5	250	0
b	1	100	100	2	2	250	0