1999

A new sieving matrix for DNA sequencing, genotyping and mutation detection and high-throughput genotyping with a 96-capillary array system

Qiufeng Gao
Iowa State University

INFORMATION TO USERS

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

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

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

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

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

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

A new sieving matrix for DNA sequencing, genotyping and mutation detection and high-throughput genotyping with a 96-capillary array system

by

Qiufeng Gao

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

Major: Analytical Chemistry

Major Professor: Edward S. Yeung

Iowa State University

Ames, Iowa

1999

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Qiufeng Gao

has met the dissertation requirement of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

To my family

TABLE OF CONTENTS

ABSTRACT

Capillary electrophoresis has been widely accepted as a fast separation technique in DNA analysis. In this dissertation, a new sieving matrix is described for DNA analysis, especially DNA sequencing, genetic typing and mutation detection. A high-throughput 96 capillary array electrophoresis system was also demonstrated for simultaneous multiple genotyping.

We first evaluated the influence of different capillary coatings on the performance of DNA sequencing. A bare capillary was compared with a DB-wax, an FC-coated and a polyvinylpyrrolidone dynamically coated capillary with PEO as sieving matrix. It was found that covalently-coated capillaries had no better performance than bare capillaries while PVP coating provided excellent and reproducible results.

We also developed a new sieving matrix for DNA separation based on commercially available poly(vinylpyrrolidone) (PVP). This sieving matrix has a very low viscosity and an excellent self-coating effect. Successful separations were achieved in uncoated capillaries. Sequencing of M13mp18 showed good resolution up to 500 bases in treated PVP solution.

Temperature gradient capillary electrophoresis and PVP solution was applied to mutation detection. A heteroduplex sample and a homoduplex reference were injected during a pair of continuous runs. A temperature gradient of 10 °C with a ramp of 0.7 °C/min was swept throughout the capillary. Detection was accomplished by laser induced fluorescence detection. Mutation detection was performed by comparing the pattern changes between the homoduplex and the heteroduplex samples. High throughput, high detection rate and easy operation were achieved in this system.

We further demonstrated fast and reliable genotyping based on CTTv STR system by multiple-capillary array electrophoresis. The PCR products from individuals were mixed with pooled allelic ladder as an absolute standard and coinjected from a 96-vial tray. Simultaneous one-color laser-induced fluorescence detection was achieved by using a CCD camera. The allele peaks for the unknown sample were identified by comparing the normalized peak intensities of the mixtures to those of the pooled ladder by using a straightforward algorithm. An extremely high level of confidence in matching the bands was indicated with negligible cross-talk (< 0.89%) between adjacent capillaries.

CHAPTER 1. GENERAL INTRODUCTION

Dissertation Organization

This dissertation consists of six chapters. It begins with a general introduction of the background. Recent research and progress are also discussed in the first chapter. In the following chapters, our discovery and development in DNA sequencing, genotyping and mutation detection with capillary electrophoresis are described. Chapter two is a manuscript of a poster for a conference. Chapters three, four, and five are manuscripts published in or prepared for the submission to a scientific journal. The last chapter summarizes the significance and perspective of the work.

Introduction

"We shall assume the structure of the gene to be that of a huge molecule, capable only of discontinuous change, which consists in a rearrangement of the atoms and leads to an isomeric molecule. The rearrangement may affect only a small region of the gene, and a vast number of different rearrangements may be possible"

Erwin Schrödinger

Ever since the era of molecular biology of gene began in 1945, people have realized the essential role of the gene in the life science. The gene is the unit of inheritance. Each gene is a nucleic acid sequence that carries the information representing a particular polypeptide. A gene is a stable entity, but is subject to occasional change in sequence. The genetic material of all organisms and many viruses is deoxyribonucleic acid (DNA). DNA analysis has

become an important way to study the evolution and polymorphism of life forms. It also provides a way to understand the cause of some diseases, such as cancer and AIDS, and helps to find cures to them.

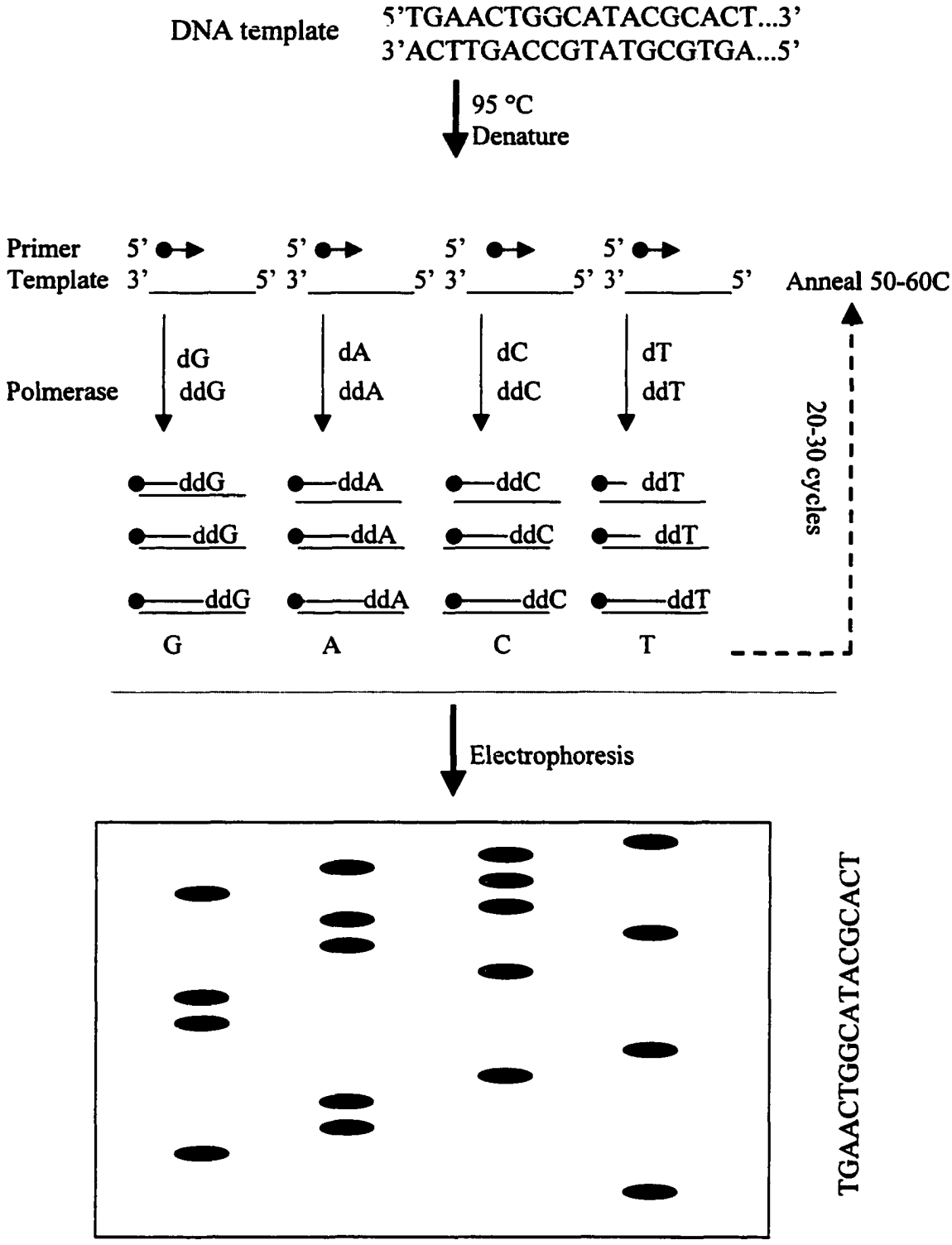
A gene consists of two DNA strands which are held together by weak hydrogen bonds and form a double helix. A DNA strand consists of a chemically linked sequence of nucleotides. Each nucleotide contains a heterocyclic ring of carbon and nitrogen atoms, a five-carbon sugar in ring form and a phosphate group. The nucleotides fall into four types: deoxyadenosine (A), deoxycytidine (C), deoxyguanosine (G), and deoxythymidine (T). The monomeric units of the nucleotides are linked together by phosphodiester bonds between the 5' phosphate groups and the 3' hydroxyl groups of adjacent sugars. The two DNA strands in a double helix are complementary to each other; an A is always paired with a T and a C is always paired with a G. The genetic information is reflected in the sequence of the nucleotides in which a DNA molecule is formed. DNA analyses are to explore the DNA sequence to find out useful information. Based on the goals and analytical methods, fundamental DNA analyses are focused on three major fields:

DNA Sequencing

The goal is to obtain the most detail information about the sequence of a DNA molecule, or most often, of a DNA fragment. The Sanger method is the most commonly employed method for DNA sequencing due to its ease of operation and automation. The key to this method is the Sanger reaction,[1] which generates a series of DNA fragments, the so-called Sanger ladder. The principle of this method is illustrated in Figure 1.

Figure 1. Sanger dideoxy reaction and cycle sequencing

(1) Template denaturing
(2) Primer annealing
(3) Primer extension with dNTP and termination by ddNTP
(4) Repeat above steps by 20-30 cycles
(5) Four-lane slab gel electrophoresis of the reaction products


DNA template
5'TGAACTGGCATACGCACT...3'
3'ACTTGACCGTATGCGTGA...5'
95 °C
Denature
Primer
Template
Polmerase
dG
ddG
dA
ddA
dC
ddC
dT
ddT
Anneal 50-60C
20-30 cycles
ddG
ddA
ddC
ddT
G
A
C
T
Electrophoresis
TGAACTGGCATACGCACT

The Sanger reaction is performed enzymatically. The DNA fragment to be sequenced is cloned into a plasmid or a vector first. Then plasmid or vector DNA is extracted and used as template. Reaction reagents (a primer, dNTP's, ddNTP's, and polymerase) are added to the reaction vial. In each reaction cycle, the template is first denatured into single-strand form. The primer then binds to the complementary part on the template. Under the catalysis of polymerase, the primer is enlongated by adding dNTP's to its 3' end following complementary rule. The enlongation may be stopped by incorporating a ddNTP to the 3' end because ddNTP is lack of the hydroxyl group at its 3' position. A series of DNA fragments with four terminators are generated in this fashion.

Traditionally, each reaction is involved with a certain type of ddNTP. Then products from four reactions are separated in a slab gel side by side. Small fragments move faster than large fragments. The migration time sequence of the fragments follows their size sequence, so that the DNA sequence of the template can be determined. The detection method used to be autoradiography. Either the primer or the dNTP's are labeled with a radioactive isotope, typically ^{32}P.

In 1986-1987, several groups developed fluorescent label to replace the traditional radioactive labels and replaced the autoradiography with laser induced fluorescence detection.[2-4] The fluorescent label not only reduced the radiation hazard caused by radioactive label, but also made the analysis more efficient because each sample can be analyzed in just one lane instead of four lanes if four-dye labeling is applied. In addition, the automation of reading electropherogram, inputting data into computer and performing analysis will be feasible with fluorescence detection. Now the four-dye, one-reaction scheme has become a standard method for DNA sequencing.

Genotyping

Human genome contains about 3 billion nucleotides. A human has 23 pairs of chromosomes. DNA fragments at identical positions on each pair of chromosomes are referred as alleles. It has been found that human genes have variations among people. An individual is said to be homozygous or heterozygous at a specific locus if the two alleles at that locus have identical or different sequences respectively. It has also been found that in the noncoding regions of human DNA, variable number of repeat sequence units may exist. The exact number of the repeat units may differ from person to person. Thousands of loci containing this type of repeat units have been found. It is believed there are more to be discovered as we have more knowledge about human genome sequence. From statistical point of view, it is possible to differentiate any two people by comparing the number of such repeats in each locus provided enough loci are studied. This technique has a significant application potential. It has been used in a wide range of medical and research applications,[5,6] but has revolutionized the field of biological identification, in particular forensic science and paternity testing.[7] With a broad definition, genetic typing may be involved with all kinds of polymorphisms, including single base-pair variations. In this thesis, the term "genotyping" refers to the technique of characterizing the polymorphism based on tandem arrays of repeated units.

Satellites,[8] minisatellites,[9,10] and microsatelites[11,12] are all classes of repetitive DNA in which the repeat sequence units are arranged in tandem. The three classes are distinguished on the basis of repeat unit length. Particularly, minisatellites and microsatellites are more commonly used for genotyping. Minisatellites typically consist of a 9 to 100 bp sequence repeated 10 to 1000 times. These units are also called variable number of tandem repeats

(VNTR). Minisatellites are highly polymorphic loci and have been ideal systems for use in individual identification. The major limitation of minisatelliets VNTRs is that they tend to be clustered at telomeres and are therefore of restricted value in constructing complete human genome maps.[10] Microsatellites were first described with a repeat unit of only two base pairs in the early 1980s.[11] However, the polymorphic nature of microsatellite was not recognized until 1989.[13] The repeat unit in microsatellites may be from 1 to 6 bp and the most common repeat motifs are A, AC, AAAN, AAN, AG, and AT.[12] Such loci are more commonly referred as short tandem repeats (STR). STR is more exploited for genotyping because it has some clear advantages over other methods. It is very abundant in human genome, with a frequency of one in every 6 kb of human genomic sequence.[12] STR is highly polymorphic , so it can provide more genetic information. Because of its short repeat units, the length of STR fragments are usually small, so that it is not sensitive to the DNA degradation or other damages. In addition, the loci are small enough to be analyzed by polymerase chain reaction (PCR).[14]

The most straightforward and reliable way to identify the genotype is to perform DNA sequencing. However, DNA sequencing is usually tedious and expensive. We normally don't need so much information about the full sequence of each sample. Instead, we want the length information of a DNA fragment containing the repeat units so that we can determine the number of the repeat units or compare the sample with a standard to determine the genotype. The technologies for genotyping sample preparation include restricted enzyme digestion[15] and polymerase chain reaction (PCR).[14]

PCR is more robust and efficient. Since its introduction in 1985,[16] it has been used in almost all the research areas of molecular biology. The principle of the reaction is similar to

that of Sanger reaction except that a primer pair is used and no ddNTP is added (Fig. 2). After a number of reaction cycles (typically 30 cycles), the product will be a single-size fragment whose composition is defined by the sequence in between the primer pair. Due to the fact of exponential amplification, the number of the product copies could be 10^6-10^7 times of the number of the template copies. The high efficiency of amplification eventually eliminates the need of radioactive labeling. Laser induced fluorescence detection is sensitive enough and sometimes, even UV absorption method may be applied. In addition, the PCR can be performed with as little as 50 copies of template molecules. It is particularly advantageous when the sample resource is limited, for example, when a trace of blood sample is taken from a criminal scene.

The determination of the size of a PCR product can be based on either separation methods or non-separation methods. In separation mode, typically with electrophoresis, the restriction or PCR fragments are separated into different bands based on their sizes, along with a DNA size marker. The actual sizes of the sample fragments can be calculated from the migration time[17] or by matching them with size marker.[18,19] Mass spectroscopy is a good example of the techniques for non-separation detection of DNA fragments. It enables simultaneous detection of all the fragments, so that the analysis speed is extremely high.[20] But the resolution to large fragments is limited. Now mass spectroscopy is only applied to the analyses of DNA fragments smaller than 200 bp. With development of high-resolution mass technique, its application in DNA analysis will be drawn more attention.

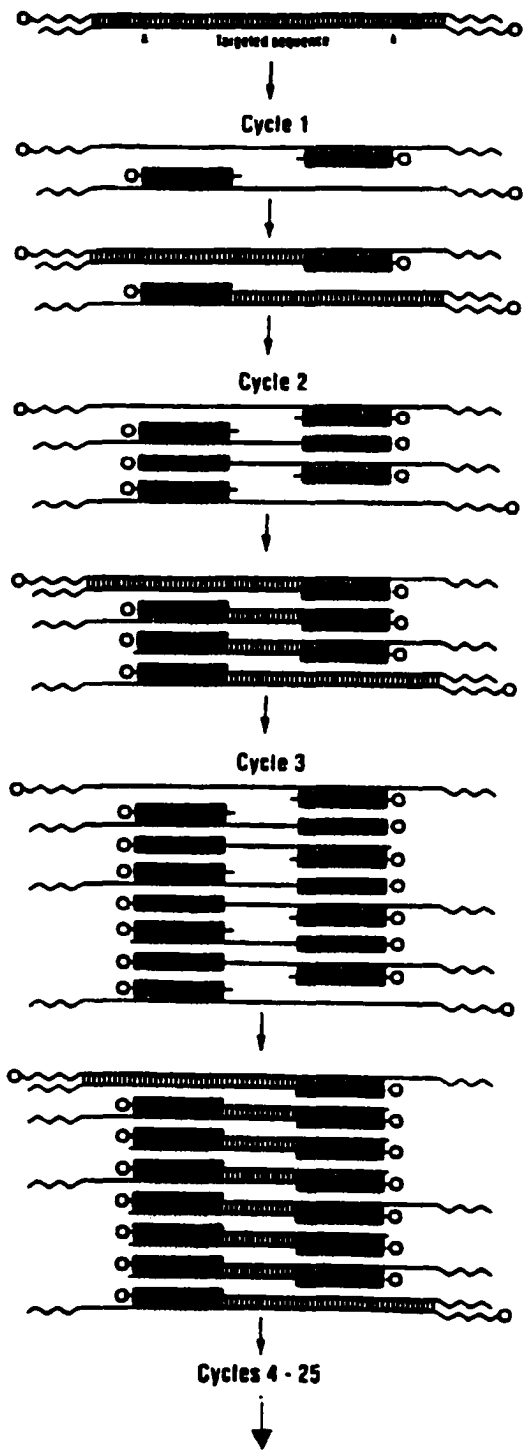


Figure 2. Principle of polymerase chain reaction (PCR)

Mutation Detection

DNA replication usually proceeds with a remarkably high degree of fidelity. Single base substitutions have been estimated to occur at frequencies of between $10^{-9}$and 10^{-11} per incorporated nucleotide.[21] Other mutational events such as insertions, deletions, and duplications occur at widely differing frequencies. Even though, there exists within and between species, and even between organisms, a considerable degree of variation in DNA sequence. Mutations may also occur at a much higher frequency if the organisms are exposed to certain hazardous sources like radioactive radiation and mutagenesis reagents.

Since a large portion of sequence variations in the human genome is caused by single base changes, any method used to detect mutations must be capable of detecting single-base substitutions. Thus single point mutations or single nucleotide polymorphisms (SNP) are the focus of mutational analysis. In some instances, the sequence of a common mutation is known, for example, the three-base deletion in the CFTR gene responsible for about 70% of cystic fibrosis chromosomes. Many methods have been developed to detect such defined mutations, such as the amplification refractory mutation system (ARMS),[22] ligase chain reaction variations,[23] allele-specific oligonucleotide (ASO),[24] and so on. Techniques for the detection of known mutations are a great help in many situations. However, the detection of unknown mutations may be more interesting and have more applications. There is a wide range of methods for this purpose. Separation-based methods include mini-sequencing,[25] single-strand conformational polymorphism (SSCP),[26] heteroduplex analysis,[27] denaturing gradient gel electrophoresis (DGGE),[28,29] chemical cleavage of mismatch,[30] enzyme mismatch cleavage,[31] denaturing HPLC (DHPLC)[32] and so forth. Non-separation methods are mainly based on hybridization, for example, DNA chip technology.[33] Mass spectroscopy

can be applied but has the limitation mentioned before. Although the analysis speed of non-separation methods is fast, the cost and generality still remain hinders for the wide acceptance of these techniques.

Although looking for mutations or markers in a predefined region of a genome is becoming well established, improvements in the throughput of analysis are still needed. This is more obvious when screening of the whole genome is demanded.

Techniques for DNA separation

Separation techniques are widely used in most of the DNA analysis applications. Electrophoresis, especially slab gel electrophoresis, has played a very important role in DNA separations. DNA molecules have same mass-to-charge ratio when the sizes are above 20 bp. Free zone electrophoresis can not differentiate the fragments. When a gel matrix is applied, the separation is solely based on the size difference (conformational difference under certain circumstance) between fragments. Slab gel offers the advantages of high resolution, long read length, and low cost. However, the running time is usually very long, from a few hours (DNA sequencing)[34] to a couple of days (SSCP to detect mutation).[35] The reason for the slow experiment is the slow dissipation of Joule heat generated by the electrophoresis current which limits the application of high voltage. The slab gel is one-time usage. The process of gel preparation is tedious. Another major drawback is that it is not easy for automation. In the past decade, much work has been done to increase the separation speed and to simplify the operation. This includes applying capillary electrophoresis (CE),[36] ultra-thin slab gels,[37] and microchip technology[38] to DNA separations. The major improvement of these techniques is

the better heat dissipation in the system because of the large area-to-volume ratio of the separation channels, so that a much higher electric field can be applied.

Capillary Gel Electrophoresis

In 1988, Cohen and Karger[36] first demonstrated the gel electrophoresis of DNA fragments in a capillary. In that work, a mixture of oligonucleotides (20-mers) were separated in less than 20 minutes. Since then, capillary gel electrophoresis (CGE) has become an attractive alternative to slab gel electrophoresis. It has been applied to different areas of DNA analysis such as sequencing,[39,40] PCR product analysis[41], genotyping[42] and mutation detection.[43] During the development of capillary gel electrophoresis, two major issues, the sieving matrix (as well as capillary wall coating) and the detection system, have been addressed and still remain challenging.

Sieving Matrices

One of the key focuses in developing capillary gel electrophoresis is the development of sieving matrices with high separation performance, high speed and replaceable operation. Cross-linked polyacrylamide (PA), which had been the sieving material in slab gel, was used in CGE in its earliest days. However, due to the instability overtime, irreproducibility in polymerization and rigid nature, cross-linked PA in CE has not been reported to last more than a few runs. The call for replaceable sieving matrices has initiated a tight contest in recent years.

The function of cross-linked PA in DNA separation is to serve as an entangled network. Linear polymers, when their concentrations are above their entangling threshold, can form

similar network and may have similar function as cross-linked PA. Following this point, many linear polymers have been employed for DNA separations, starting from linear polyacrylamide (LPA).[39] These polymers keep the high-resolution property of cross-linked PA, but reduce the running time drastically. In addition, because of the relatively low viscosity of linear polymer solution, the capillary regeneration and gel replacement become feasible. This greatly encourages the development of system automation. Up to date, the polymers used in DNA analysis include LPA and its derivatives,[44,45] cellulose and its derivatives,[46] poly(ethylene oxide) (PEO),[47] polyvinylpyrrolidone (PVP),[48] etc. In terms of separation performance, LPA is still the most popular choice, which achieved base calling up to 1300 bases within 80 min in DNA sequencing.[49] In terms of viscosity, PVP (27 cp at 4.5%)[48] and polydimethylacrylamide (PDMA, 75 cp at 6%)[45] seem to be the least viscous solution with reasonable separation performance.

One important issue related to linear polymers is the capillary coating. In DNA analysis, the buffer pH is usually above 8, which causes the inner surface of the capillary partially negatively charged. Under this condition, DNA separation becomes difficult because a strong electroosmotic flow (EOF) moves against DNA electrophoretic migration. With cross-linked PA, EOF can be suppressed to nearly zero. But with polymer solution, the case should be handled with caution. The low viscosity of polymer solution may allow the EOF to exist in a large value and slow down the migration of DNA. Even worse is, sometimes the EOF could drive the polymer out of capillary along with DNA sample.[50] A solution to this problem is to modify the capillary inner surface to suppress EOF. Most research groups applied covalently-bonded polymers into the capillary. Hjeten's polyacrylamide coating[51] is widely used in DNA analysis. Other coatings like DB ** series

from J&W[47] and polyvinyl alcohol[39] were also applied in DNA analysis. The temperature stability and lifetime of these coatings are less reported, but the validation of the data is important.

Besides covalent coating, dynamic coating with a highly adsorptive polymer is another option and seems more convenient and cost-efficient. Some polymers, such as PEO, PVP and PDMA, have strong affinity toward silica. When the polymer solution flow through the capillary, a thin layer of densely adsorbed polymer will cover the inner surface of the capillary. The interaction between the polymer and the capillary inner surface is so strong that the process of desorption is very slow during the electrophoresis. This dynamic polymer coating has the same function as covalent coating and can suppress EOF to negligible level. Therefore, uncoated capillaries can be applied to DNA separations using this type of polymers as sieving matrices or dynamic coatings.[45,47,49]

While seeking a new sieving matrix for DNA separation, some basic rules should be followed. Separation performance is the ultimate criterion to judge the polymer candidate. There are also some structural requirements as following that can be used as a guidance to select or design sieving matrix candidates:

- Water-soluble
- No Charge
- Chemically Stable
- No Fluorescence in Visible Wavelength Range
- Large Molecular Weight
- No or Low Toxicity
- Low Viscosity

Detection

While absorbance detection is widely used in HPLC, its application in CGE is a far cry and proved to be inadequate. This can be attributed to the short path length provided by the small diameter (10-75 μm) of the capillary tubing. Only when the DNA sample is in large amount, usually after PCR amplification, the absorbance method may be applied.[52,53]

In light of the inadequate sensitivity of absorbance detection, it is not surprising that fluorescence detection in CGE has attracted much attention. A major reason is because a laser beam usually provides enough excitation power to produce so much fluorescence that the signal-to-noise ratio is finally limited by background noise instead of the absolute fluorescence signal. Meanwhile, labeling a fluorophore to a DNA molecule is not a difficult task and will not change the DNA property much. The application of laser in fluorescence detection strengthened the advantage of this method. Extremely high sensitivity (10^{-13} M detection limit) has been reached with laser induced fluorescence (LIF) detection.[54] Up to date, there are four detection modes in capillary electrophoresis. The first is direct focusing laser onto a capillary window.[55] The focal point is at the center of capillary bore. The fluorescence is collected from a 90° angle to the laser beam. The laser scattering light from the capillary wall will limit the detection limit. To avoid the scattering light, a laser beam can be shined through the detection end and axially illuminate the buffer solution.[56] The fluorescence can be collected into a charge-coupled device (CCD) detector. Photo bleaching of the dye in this scheme may limit the laser power used. Another state-of-the-art setup to reduce scattering light is to attach a sheath flow cuvette to the end of the capillary.[57] Because of the excellent optical quality of the cuvette and well–defined sheath flow, high-power laser can be applied without the compromise from increased scattering. The adoption of a confocal

fluorescence microscope as a CE detector offers sensitivity levels comparable to those with sheath flow.[58] In addition, the confocal arrangement makes the alignment of the system easier.

In laser induced fluorescence detection mode, the DNA molecules can be covalently labeled by attaching a fluorophore to the primer if PCR is performed or to the terminator if cycle sequencing is performed. For restriction fragments, intercalating dye like ethidium bromide, picoGreen, TOTO, YOYO, etc,[59,60] will bind to double-stranded DNA at a dye:base pair ratio of about 1:5. With certain laser excitation, the native fluorescence from the dyes is very low while a strong fluorescence can be obtained after their binding to DNA. These dyes can also be used for PCR products.

As discussed before, mass spectroscopy is always a choice for detection. When coupled with CGE, the interface between separation channel and mass inlet becomes important. One possible solution may be sheath flow at the end of separation. This has been accomplished in microchip.[61]

Besides the issues discussed above, other experimental conditions, such as capillary length, electric field, temperature, buffer composition, polymer concentration, and so on, are also critical to good separation of DNA fragments. The selection and optimization of these parameters have been extensively discussed.[62-65]

High-throughput DNA Analysis

Although capillary gel electrophoresis is much faster than traditional slab gel, the throughput of analysis of large batch of samples has no improvement, even it is worse. The reason lies on the fact that the slab gel has multi-channels (48 lanes from ABI 377

sequencer). A number of samples can be analyzed simultaneously in slab gel and the slow speed can be compensated by the parallel analysis. The current DNA analysis requires, especially under the demand of Human Genome Project and high throughout screenings, that the capillary method should be able to process a large number of samples at one time. Thus the multi-channel devices are conceived and generated.

Multiple Capillary Array

In the past few years, our group and several other groups have developed different setups and detection systems to accommodate parallel arrays for DNA sequencing in capillary electrophoresis. The major difference among these designs is the detection part. Generally, the detection modes can be summarized into two types: scanning mode and imaging mode. In scanning systems, either the capillaries or the detection system must be translated along the scan axis. In an imaging detection system, all capillaries must be illuminated simultaneously.

Mathies' group first developed the confocal scanning multiple-capillary array. This system provides high sensitivity by using a high-numerical-aperture microscope objective, while the background fluorescence and scattering from the capillary are minimized through confocal sectioning. 25 capillaries are packed side by side. A translation stage is used to sweep the capillary array across the focal point of the laser back and forth at a rate of 1 Hz.[66] The small, defined excitation region enables the use of a modest laser power and elimination of cross talk between adjacent capillaries. High repetition rate scan systems with sensitive closed feedback loops are required and therefore limit the number of capillaries that may be

scanned. A recent development of this system used a rotating objective lens instead of sweeping capillaries. It can lead up to 1000 capillaries.[67]

The imaging detection systems have used multi-sheath flow, on-column illumination and fiber-optic array illumination.

Kambara's group developed sheath flow imaging capillary array.[68] In a detection cell filled with buffer, there is a 1-mm tip-to-tip gap between a bundle of gel filled capillaries and same number open capillaries. Sheath flow is formed by putting the buffer reservoir above the sheath flow cell and sample eluted from the gel-filled capillary will be brought into the open capillary. Laser beam is aligned with the long axis and excites the sheath flow region. A CCD camera is used to collect and image the fluorescence. Because the DNA sample is illuminated in the gel-free region, very low background fluorescence and scattering light is detected. Therefore, better signal-to-noise ratio can be achieved in this system. Two-dimensional packing of the capillary array has been acclaimed. However, it is quite difficult to flush the sheath flow chamber between runs.

Our group first reported the on-column imaging design.[69] 100 capillaries are densely packed. A laser beam is expanded by a cylindrical lens and then line-focused onto the capillary array at a 45° angle. The illuminated region of the capillary array is imaged onto the CCD camera. Because the laser beam is dispersed, a large laser power is need to excite the capillary array. However, this design has no moving parts and can provide 100% duty cycle. Using this scheme as detection mode, we are capable of performing 96-capillary DNA sequencing separation over 400 bases within 2 hours[70] and simultaneous 96-sample genotyping within 1 hour.[19]

An alternative design to save the laser power is side-entry excitation.[71,72] Instead of using expanded laser beam, A focused laser beam passes through all the capillaries along the axis of the capillary array. The transmitted laser beam through one capillary can be reused by next one. One problem with this design is the uniform excitation due to the gradual decrease of laser intensity from capillary to capillary by multiple reflection and scattering. The difficult alignment of the system also limits the number of the capillaries in the array.

An approach using fiber-optic illumination was developed by Quesada and Zhang.[73] Eight optic fibers were coupled with eight capillaries perpendicularly. Laser excitation and fluorescence collection were performed through the optic fibers. Low-numerical aperture excitation and high-numerical aperture collection were used to achieve high sensitivity. The requirement of high laser power and difficult alignment of the fiber-capillary junction will limit the application of this scheme to a large capillary-array.

Microchip

Microchips are miniaturized multiple-channel devices. The computer industry has brought a steady performance improvement due to increased system complexity and integration while costs were reduced by automated batch fabrication. The similar idea was developed in several laboratories in early '90s. Due to the lack of suitable micropumps and microvalves, chip-based HPLC was not quite successful. However, driven by electric forces, CE showed a great potential to be transferred to chip-based devices with the promise of a further reduction in analysis time and a higher degree of integration.

Woolly and Mathies first demonstrated DNA sequencing of 200 bases in 10 min on CE chips.[74] Since then, the development of chip technology in DNA analysis field has made

great progress. Research has been conducted to improve the resolution on chip. These developments include channel design, channel coating, better sieving matrix composition, decreasing in the length of injection zone.[75,76] Recent achievements include separation of more than 400 bases in 15 min with one-color DNA sequencing[75] and 600 bases in 20 min with four-color sequencing.[76] These performances are very impressive, but still no better than capillary-based results at present time, because the loss of resolution due to decreased channel length can not be compensated by reduced time. Compared to DNA sequencing, analysis of genotyping, restriction fragments, PCR products and short oligonucleotides are more suitable for microchip, where less resolution is required than sequencing.[77-79] High-speed genotyping of 12 samples in 160 s was accomplished in microchip.[75]

In addition to the capability of faster separation speed, microchip channel can be fabricated with essentially zero dead volume intersections. This enables the integration of reaction, sample pretreatment, separation and fraction collection. An intrinsic advantage of microfabricated devices is the potential to produce arrays of separation channels for high-throughput applications with negligible additional costs.

Up to now, most publications on microchip focus on the use of quartz or glass devices. Besides the advantageous of electrical, thermal, and optical properties of these materials, surface modification methods can be easily transferred from CE. The problem with microchip separation at this stage is the time-and labor-intensive fabrication procedure, sample introduction and chip regeneration. These steps will slow down the cycle operation and detriment the throughput. A disposable microchip based on polymer material may be the solution to chip regeneration.[80] But it is still under development. Problems with uniformity of

fabrication and fluorescence detection need to be solved. For now, the capillary-array is still the mature device for high throughput DNA Analysis.

Our Goal

Our goal is to develop suitable sieving matrices and capillary coatings for DNA analysis and apply our multiple capillaries array system to perform high-throughput DNA genotyping and mutation detection.

References

1. Sanger, F., Nicklen, S., Coulson, A. R. *Proc. Natl. Acad. Sci. USA* **1977**, 74, 5463-5467
2. Smith, L. M., Sanders, J. Z., Kaiser, R. J. et al. *Nature* **1986**, 321, 674
3. Ansorge, W., Sproat, B. S., Stegemann, J., et al. *J. Biochem. Biophys. Meth.* **1986**, 13, 315
4. Prober, J. M., Trainor, G. L. Dam, R. J. *Science* **1987**, 238, 336
5. Knowlton, R. G., Brown, V. A., Braman, J. C., et al. *Blood* **1986**, 68, 378
6. Nakamura, Y., Leppert, M., O'connell, P., et al. *Science* **1987**, 235, 1616
7. Nobel, D. *Anal. Chem.* **1995**, 67, 613-615 A
8. Jones, K.W. J. Med. Genet. **1973**, 10 (3): 273-281
9. Jeffreys, A. J., Wilson, V., Thein, S.L. *Nature* **1985**, 316, 76.
10. Armour, J. A. L., Jefferies, A. J. *Curr. Opin. Genet. Dev.* **1992**, 2, 850
11. Miesgeld, R., Krystal, M., Arnheeim, N. *Nucleic Acids Res.* **1981**, 9, 5931
12. Beckmann, J. S., Weber, J. L. *Genomics* **1992**, 12, 627
13. Weber, J. L., May, P. E. *Am. J. Hum. Genet.* **1989**, 44, 388

14. Gill, P., Ivanov, P. L., Kimpton, C., et al. *Nature Genet.* **1994**, 6, 130

15. Nakamura, Y., Carlson, M., Krapcho, K., et al. *Am. J. Hum. Genet.* **1988**, 43, 854

16. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., et al. *Science* **1985**, 230 (4732): 1350-1354

17. Butler, J. M., McCord, B. R., Jung J. M., et al. *Electrophoresis* **1995**, 16, 974-980

18. Wang, Y., Ju, J., Carpenter, B. A., et al. *Anal. Chem.* **1995**, 67, 1197-1203

19. Gao, Q., Pang, H. M., Yeung, E. S. *Electrophoresis*, **1999**, in press

20. Butler, J. M., et al. *Int. J. Med.* **1998**, 112 (1), 45

21. Lundstrom, R., Tavare, S., Ward, R. H. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 5961

22. Newton, C. R., Graham, A., Heptinstall, L. E., et al. *Nucleic Acids Res.* **1989**, 17, 2503

23. Karthigesu, V. D., Mendy, M., Fortuin, M. *FEMS Microbiol. Lett.* **1995**, 131 (2): 127-32

24. Chehab, F. F., Wall, J. *Hum. Genet.* **1992**, 89, 163

25. Syvänen, A. C. *Clin. Chim. Acta.* **1994**, 226, 225

26. Ellison, J., Dean, M., Goldman, D. *Biotechniques* **1993**, 15, 684

27. Ganguly, A., Rock, M. J., Prockop, D. J. *Proc. Natl. Acad. Sci., USA* **1993**, 90, 10325

28. Myers, R. M., maniatis, T., Lerman, L. *Methods Enzymol.* **1987**, 155, 501-527

29. Riesner, D., Henco, K., Steger, G. *Adv. Electrophoresis* **1991**, 4, 169-250

30. Ramus, S. J., Cotton, R. G. *Biotechniques* **1996**, 21, 216

31. Youil, R., Kemper, B. W., Cotton, R. G. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 87

32. Ofner, P. J., Underhill, P. A. *Am. J. Hum. Genet.* **1995**, 52, 266

33. Masanori, G., Masahisa, H., Tomoko, O., et al. *Genet. Anal.: Biomol. Eng.* **1997**, 14(2), 47-50

34. Luckey, J. A., Drossman, H., Kostichka, A. J., et al. *Nucleic Acids Res.* **1990**, 18, 4417-4421

35. Orita, M., Iwahana, H., Kanazawa, H., et al. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 2766-2770

36. Cohen, A. S., Najarian, D. R., Paulus, A., et al. *Proc. Natl. Acad. Sci. USA* **1988**, 85, 9660-9663

37. Chen, D., Peterson, M. D., Brumley, R. L., et al. *Anal. Chem.* **1995**, 67, 3405-3411

38. Woolley, A. T., Mathies, R. A. *Proc. Natl. Acad. Sci. USA* **1994**, 91, 11348-11352

39. Ruiz-Martinez, M. C., Berka, J., Belenkii, A., et al. *Anal. Chem.* **1993**, 65, 2851-2858

40. Tan, H., Yeung, E. S. *Anal. Chem.* **1998**, 70, 4044-4053

41. Zhang, N., Yeung, E. S. *J. Chromatogr.* **1998**, 714, 3-11

42. Zhang, N., Yeung, E. S. *Anal. Chem.* **1996**, 68, 2927-2931

43. Gelfi, C., Orsi, A., Righetti, P. G., et al. *J. Chromatogr.* **1994**, 657, 201-205

44. Carrilho, E., et al. *Anal. Chem.* **1996**, 68, 3305-3313

45. Madabhushi, R. S. *Electrophoresis* **1998**, 19, 224-230

46. Bashkin, J., Marsh, M., Barker, D., Johnston, R. *Appl. And Theor. Electrophor.* **1996**, 6, 23-28

47. Fung, E. Yeung, Yeung, E. S. *Anal. Chem.* **1995**, 67, 1913-1919

48. Gao, Q., Yeung, E. S. *Anal. Chem.* **1998**, 70, 1382-1388

49. Salas-solano, O., Carrilho, E., Miller, A. W. *Anal. Chem.* **1998**, 70, 3996-4003

50. Yin, H. F., Lux, J. A., Schomberg, G. *J. High Resolut. Chromatogr.* **1990**, 13, 624-627

51. Hjerten, S. *J. Chromatogr.* **1985**, 347, 191-198

52. Gelfi, C., Righetti, P. G., Magnani, C., et al. *Clin. Chim. Acta* **1994**, 229, 181-189

53. Oto, M., Suehiro, T., Akiyama, Y., Yuasa, Y. *Clin. Chem.* **1995**, 41, 482-483

54. Soper, S. A., Shera, E. B., Martin, J. C.,et al. *Anal. Chem.* **1991**, 63, 432

55. Yeung, E. S., Wamg, P., Li, W., Giese, R. W. *J. Chromatogr.* **1991**, 559, 183

56. Taylor, J. A., Yeung, E. S. *Anal. Chem.* **1992**, 64, 1741-1744

57. Zhao, J., Chen, D., Dovichi, N. J. *J. Chromatogr.* **1992**, 608, 117

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

59. Heller, C., Pakleza, C., Viovy, J. *Electrophoresis* **1995**, 16, 1423-1428

60. Mitnik, L., Heller, C., Viovy, J. *Science* **1995**, 267, 219-222

61. Zhang, B., Karger, B. L. *Anal. Chem.* submitted

62. Figeys, D., Dovichi, N. J. *J. Chromatogr.* **1993**, 645, 311-317

63. Nishikawa, T., Kambara, H. *Electrophoresis* **1994**, 15, 215-220

64. Luckey, J. A., Smith, L. M. *Anal. Chem.* **1993**, 65, 2841-2850

65. Issaq, H. I., Chan, K. C., Muschik, G. M. *Electrophoresis* **1997**, 18, 1153-1158

66. Mathies, R. A., Huang, X. C. *Nature* **1992**, 359, 167-169

67. Kheterpal, I., Mathies, R. A. *Anal. Chem.* **1999**, 71, 31-37A.

68. Takahashi, S., Murakami, K, Anazawa, T., Kambara, H. *Anal. Chem.* **1994**, 66, 1021-1026

69. Ueno, K., Yeung, E. S. *Anal. Chem.* **1994**, 66, 1424-143

70. Pang, H. M., Pavarski, V., Yeung, E. S. *J. Biophysics Biochem.* **1999**, submitted

71. Lu, X., Yeung, E. S. *Applied Spectroscopy* **1995**, 49, 605-609

72. Tan, H., Yeung, E. S. *Anal. Chem.* **1998**, 70, 4044-4053

73. Quesada, M. A., Zhang, S. *Electrophoresis* **1996**, 17, 1841-1851

74. Woolley, A. T., Mathies, R. A. *Anal. Chem.* **1995**, 67, 3676-3680

75. Woolley, A. T., Sensabaugh, G. F., Mathies, R. A. *Anal. Chem.* **1997**, 69, 2181-2186

76. Liu, S., Shi, Y., Ja, W. W., Mathies, R. A. *Anal. Chem.* **1999**, 71, 566-573

77. Manz, A, Harrison, D. J., Verpoorte, E. M., et al. *J. Chromatogr.* **1992**, 593, 253-258

78. Burns, M. A., Mastranelo, C. H., Sammarco, T. S., et al. *Proc. Natl. Acad. Sci. USA* **1996**, 93, 5556-5561

79. Volkmuth, W. D., Austin, R. H. *Nature* **1992**, 358, 600-602

80. Effenhauser, C. S., Bruin, G. J. M., Paulus, A., Ehrat, M. *Anal. Chem.* **1997**, 69, 3451-3457

CHAPTER 2. EVALUATION OF CAPILLARY COATINGS IN DNA SEQUENCING IN POLY(ETHYLENE OXIDE) SOLUTION

A paper presented as a poster at 11th HPCE Conference

Qiufeng Gao and Edward S. Yeung

Abstract

Polyethylene oxide (PEO) has been proved to be a successful gel matrix for DNA separation. Since most other matrices require coated columns, PEO's capability of being performed in bare silica capillaries is more attractive. To optimize the separation system, different columns with variety of coatings were compared for DNA sequencing applications. These capillaries include non-coated, DB-wax coated, newly developed FC coated and polyvinylpyrrolidone (PVP) dynamically coated columns. For a new column, the PVP-coated capillary showed better performance than the bare capillary. Both the bare capillary and the PVP coated capillary showed better performance than DB-wax and FC-coated capillaries. It was also found that covalently-coated capillaries will degrade in a rate similar to that in bare capillaries and PVP coated capillaries have longer lifetime. Reasons for the degradation of coatings are discussed.

Introduction

Capillary gel electrophoresis is now a popular method for DNA separation.[1] Linear polymer solutions as replaceable matrices are more attractive polymer matrices than cross-linked ones because linear polymers are less viscous and capillaries are regeneratable.[2,3]

Compared to capillary electrophoresis with cross-linked gel, capillary inner wall plays an important role regarding to column performance with linear polymer matrices. The key point here is the electroosmotic flow (EOF). In cross-linked capillary electrophoresis, the polymers are covalently bonded to the capillary inner wall. EOF is eventually eliminated under this condition. However, with linear polymer solution, the residual silanol groups are active and EOF exists. EOF migrates against DNA electrophoretic flow and has detriment to the separation of DNA fragments in gel. Since the net motion of DNA fragments is dictated by (μ_{DNA}-μ_{EOF}), the large DNA fragments are affected more severely than the small fragments. Therefore, one quickly loses the ability to separate large bases in the run. To reduce the effect of EOF and get reproducible results, most of the experiments were performed in coated capillaries.[3-5] A common method to coat the capillary with polyacrylamide (PA) was developed by Hjeten.[6] Unfortunately, the capillary coating procedure is undoubtedly laborious and the coating degrades with usage. This is not surprising since PA also breaks down with time on interaction with the typical buffers used for DNA sequencing.[7] There is a need for better surface treatment procedure for the capillary columns to retain their performance over many runs.

Polymer solution of mixed polyethylene oxide (PEO) was developed as a new matrix in our group.[8] It has been successfully applied to DNA sequencing[9-11] and genotyping.[12] Fung and Yeung first reported the application of PEO to DNA sequencing in bare capillaries.[9] They compared PEO with LPA in terms of separation resolution and migration time. The performance of PEO in bare capillaries is similar to that of LPA in coated capillaries. Kim and Yeung later accomplished DNA sequencing up to 1000 bases with PEO in bare capillaries.[10] In both publications, the authors used a capillary regeneration procedure based

on HCl treatment followed by methanol washing. This procedure takes more than one hour to bring the capillaries back to the operational condition. The parameters that affect the separation performance, such as polymer composition, electric field, capillary length, temperature, etc, were extensively studied.[9-12] However, less study was performed to the effect of capillary coatings in a long term. To further optimize the separation system, we compared the DNA sequencing performance of this matrix in several different columns: bare silica columns, DB-wax coated columns, FC coated columns and polyvinylpyrrolidone (PVP) coated columns. In this article, we report our evaluation on long-term performance of these coatings used with PEO gel matrix.

Experimental Section

Laser-Excited Fluorescence Detection

The experiment setup is similar to that reported before[13] except that it is in a compact box. Figure 1 is an illustration of this setup. A 5-mW He-Ne laser with 543.5 nm output from Melles Griot (Irvine, CA) was used for excitation. A PMT was used to collect fluorescence signal. Two 610 nm long pass filters were used in front of the PMT window to block scattered laser light. The PMT output was transferred through a 10kΩ resistor to a 24-bit A/D interface (Lawson Labs, Kalispell, MT) at 5 Hz and stored and analyzed in a computer.

Capillaries and Reagents

Bare capillaries with 75μm I.D. (or 50 μm I.D.) and 375μm O.D. were obtained from Polymicro Technologies (Phoenix, AZ). Coated capillaries (DB-Wax with 50μm I.D. and 375μm O.D., FC-coated capillaries with 75 μm I.D. and 190 μm O.D.) were obtained from

J&W (Folsom, CA). All chemicals for making buffer were purchased from ICN Biochemicals (Aurora, OH). Courmarin 334 was from Eastman Kodak (Rochester, NY). The concentration used for determination of EOF was 10^{-7} M. PEO polymers were obtained from Aldrich Chemical (Milwaukee, WI). PVP powder (Mw = 1,000,000) was obtained from PolyScience (Warrington, PA). Fuming hydrochloric acid was obtained from Fisher (Pittsburgh, PA). PGEM/U DNA sequencing samples were obtained from Nucleic Acid Facilities (Iowa State University, Ames, IA).

Column Regeneration

After each run, the bare capillaries were treated with 0.1 N HCl for 0.5 h followed by 0.5 h methanol washing. The coated capillaries were treated just by fast flushing with methanol. The PVP dynamic coating was made by flushing a 2% PVP solution (in 1xTBE) through the capillary. After these treatments, PEO gel was filled again through a micro-syringe.

Buffer and PEO Gel

Buffer solution is 1xTBE (89 mM Tris, 89 mM Boric acid, 2 mM EDTA) with 3.5 M Urea in deionized water. The sieving matrix was prepared by mixing 1.5% (w/v) 8,000,000 Mw and 1.4% (w/v) 600,000 Mw PEO in above-mentioned buffer. Stirring was maintained at room temperature for 24 hrs. The matrix preparation procedure was similar for 2.0% 8,000,000 Mw and 1.4% 600,000 Mw PEO gel. 2% PVP solution was made by shaking the mixture of buffer and PVP powder for 5 min and then leaving it for 10 min.

DNA Sample Treatment and Injection

The DNA sample was denatured by heating it in a denaturing solution (5:1 formamide-50mM aqueous EDTA solution) at 95 °C for 3min. Then bottom of the sample vial was dipped into a bottle filled with -20°C ice-ethylene glycol until the injection was finished. Injection was performed under 100v/cm for 40s without pre-run in all experiments.

Results and Discussion

Characteristics of the Capillary Coatings

People have used surfactants and GC columns in CE to manipulate electroosmotic flow.[14] The DB-wax coating was originally developed for gas chromatography.[15] The coating material is bonded and cross-linked polyethylene glycol (PEG). So it has a high polarity. It is stable at pH range of 2-8. Though it was not designed for the harsh, aqueous environment of CE, it has been proven effective for the separation of DNA and protein and for the prevention of protein absoption to the capillary wall in CE.

The FC coating was specially designed for CE. It is a highly hydrophobic, cross-linked fluorocarbon polymer layer. FC coating has a wider pH adaption range from 2 to 10. It was claimed that FC coating has better performance and longer lifetime than GC columns like DB-wax in CE applications.[15]

The PVP coating is a dynamic coating which is held by the absorption interaction between the polymer molecules and the silanol groups on the inner capillary wall. The interaction is so strong that the brokerage of the coating is very slow and will not be noticed during the electrophoresis. It can be flushed off by water or refreshed by new PVP solution. So it is possible to regenerate the dynamic coating between each run.

EOF in Capillaries with Different Coatings

The primary function of the capillary coatings is to suppress EOF. To evaluate the performance of these capillaries, we first compared the EOF suppressing effect in these capillaries. Among neutral dyes, Courmarin 334 is small and the adsorption on the capillary wall is negligible.[15] So it was used as marker for the determination of EOF. Different coated capillaries were filled with 1xTBE buffer. Dynamic injection of marker for 5s at the height difference of 10 cm was performed. The electrophoresis was conducted in 1xTBE buffer with 3.5M urea (pH = 8.2), with a capillary effective length of 20 cm. The electric field was 200 V/cm. The results are shown in table 1.

TABLE 1. EOF values in different coated capillaries with 1x TBE buffer

Column type	Bare silica	DB-wax	FC-coated	PVP-coated
μ_{eof} (cm^2/vs) x10^5	44	1.0	2.1	< 0.1

In bare capillaries, EOF exists at a pretty high value comparable to DNA electrophoretic mobility. All the coatings are efficient for reducing EOF by more than 20 folds to a negligible level.

Performance of New Columns

The ultimate criterion to judge the capillary coatings is the separation performance. Figure 2 shows the difference of separation performance in a new bare capillary and in a new DB-wax capillary. Since the two types of capillaries have the same inner and outer diameters,

same total and effective lengths and they were filled with the same batch of gel, the performance difference should be due to the effect of coating. For small base pairs, there is no significant difference in term of resolution. But for large fragments (>300 bases), the bare capillary is better than DB-wax capillary.

Similar result was obtained while comparing a bare capillary with an FC-coated capillary (Figure 3). We can still see a little bit difference through the whole read window though performance of the FC column is quite close to that of a bare capillary. So bare capillaries seem to have better performance than coated ones in the first runs. One possible reason for that is PEO has a strong interaction with silica wall. Our group has demonstrated that PEO can form a layer of dynamic coating on bare silica surface and then reduce EOF.[16] Hydrogen bonding between PEO and hydroxyl group on silica surface was also discussed by other authors.[17,18] The interaction may help PEO molecules form stable, stuck-to-wall network structure, which is preferred for better separation. On the other hand, because coated capillary walls are lack of this hydrogen bonding with PEO, the residual EOF will still affect the separation.

While the bare capillaries have better performance than covalently-coated capillaries, the PVP-coated capillary showed same performance to that in bare capillary for small fragments but better performance for large fragments (Figure 4). It should be also noted that the migration time in the PVP coated capillary is shorter than that in the bare capillary. This is more obvious for the large fragments. A good explanation that may contribute to this performance is better EOF-suppressing effect of PVP coating than PEO self-coating.

Lifetime of Columns

It has been acknowledged that bare capillaries will degrade after a few times of usage. Actually, In our system, a new bare capillary can be used up to 25 runs if the regeneration procedure described above is followed.[9] A coated capillary is expected to be used much longer because of its inert characteristics to buffer and to PEO.

However, our result showed that covalently-coated capillaries do not necessarily have longer lifetime than bare capillary. Figure 5 showed the degraded performance of DB-wax coated columns after 23 runs. Figure 6 is the electropherogram of DNA sequencing in a FC-coated capillary which had been used for 23 runs. The separation was much worse for the large fragments in the used coated capillaries. Figure 7 shows the electromicroscopic graph of the DB-wax coatings before and after degradation. The new coating is smooth but the degraded one has many "holes" on it. The holes may contribute to the worse performance of the coated capillary after degradation. The fact is gel matrix is much more viscous than buffer solution. When column is regenerated, the friction between gel and coating will result in physical damage to region of weak coating. In addition, during running, the coating was in a basic environment and there is chemical degradation occurring to coating. DB-wax has stability to a narrow range of pH (2-8). FC coating has stability to a larger pH scale (2-10) which enables it to have longer lifetime in TBE buffer. A less viscous PEO gel may make the capillary regeneration easier and prolong the capillary's lifetime.

With PVP as dynamic coating, the column regeneration becomes much easier. The lifetime of PVP-coated capillary appeared to be much longer than both bare capillaries and covalently-coated capillaries. This was better reflected in the lifetime change of capillary array. We have reported 96-capillary DNA sequencing with PEO gel. The problem with this

system was the difficulty of array regeneration.[18] The life time of bare capillaries in the array appeared to be much shorter than that of the single capillary. A capillary array lasted only 3 runs even with long-time HCl flushing and methanol conditioning. Using 2 % PVP solution as dynamic coating, the lifetime of the capillary array was greatly prolonged. Over 30 runs of DNA sequencing have been performed in the same array and no sign of array degradation was observed.[19, 20] The detail was discussed in reference 19 and 20.

Conclusion

Through multiple runs of DNA sequencing in columns with different coating, we evaluated the performance of the coatings in terms of performance and longevity. We conclude that bare capillaries are the choice for PEO system. FC-coated capillaries have similar performance to that of bare columns. Thus, they can be used to keep resolution when bare capillaries are not applicable. Covalently-coated capillaries have certain lifetime with PEO as sieving matrix, generally between 20-30 runs. Degradation of the coating is due to chemical erosion and physical friction. Coatings designed for CE like the FC coating is more stable than GC coatings when used in CGE. A probable way to prolong the lifetime of the capillary is to reduce the viscosity of PEO gel. PVP dynamic coating is a successful solution to handling the problem of capillary regeneration. This is especially beneficial to multiple capillaries array electrophoresis.

Acknowledgments

The Ames Laboratory is operated for the US Department of Energy by Iowa State University under contract No.W7405-Eng-82. This work was supported by the Director of

Energy Research, Office of Basic Energy Science, Division of Chemical Sciences, and the Office of Health and Environmental Research.

References

1. Swerdlow, H., Gesteland, R. *Nucleic Acids Res.* **1990**, 18, 1415-1419.
2. Ruiz-Martinez, M. C., Berka, J., Belenkii, A., et al. *Anal. Chem.* **1993**, 65, 2851-2858.
3. Carrihlo, E., Ruiz-Martinez, M. C., Berka, J., et al. *Anal. Chem.* **1996**, 19, 3305-3313.
4. Boček, P., Chrambach, A. *Electrophoresis* **1991**, 12, 1059-1061.
5. Bashkin, J., Marsh, M., Barker, D., Johnson, R. *Appl. and Theor. Electrophor.* **1996**, 6, 23-28.
6. Hjerten, S. *J. Chromatogr.* **1985**, 347, 191-198.
7. Starita-Geribaldi, M., Houri, A., Sudaka, P. *Electrophoresis* **1993**, 14, 773-781.
8. Chang, H. T., Yeung, E. S. *J. Chromatogr.* **1995**, 669, 113-123.
9. Fung, E. N., Yeung, E. S. *Anal. Chem.* **1995**, 67, 1913-1919.
10. Kim, Y., Yeung, E. S. *J. Chromatogr.* **1997**, 769, 321-326.
11. Li, Q., Yeung, E. S. *Appl. Spectroscopy* **1995**, 49, 1528-1533.
12. Zhang, N., Yeung, E. S., *Anal. Chem.* **1996**, 68, 2927-1931.
13. Lee, T. T., Yeung, E. S. *Anal. Chem.* **1991**, 63, 2842-2848.
14. J&W Scientific Products Mannual, 1997.
15. Preisler, J., Yeung, E. S. *Anal.Chem.* **1996**, 68, 2885-2889.
16. Madabhushi, R. S., Vainer, M., Dolnik, V., et al. *Electrophoresis* **1997**, 18, 104-111.
17. Iki, N., Yeung, E. S. *J. Chromatogr.* **1996**, 731, 273-282.
18. Li, Q., Yeung, E. S. unpublished result.

19. Pavarski, V., Gao, Q., Yeung, E. S. *Abstract of 216th ACS National Meeting*, **1998**.

20. Pang, H., Pavarski, V., Yeung, E. S. *J. Biophys. and Biochem.* submitted.

Figure Captions

Figure 1: Setup of single capillary electrophoresis.

Figure 2: Comparison of DNA separation in bare capillary and in DB-wax coated capillary. (A) and (C) are for bare capillary (B) and (D) are for DB-wax. Compare (A) with (B) and (C) with (D). Condition: New capillaries with 50 μm I.D., 375 μm O.D. for both. 55 cm total length, 41 cm effective length, 200v/cm electric field; 1.5%+1.4% PEO gel.

Figure 3: Comparison of DNA separation in bare capillary and in FC-coated capillary. (A) and (C) are from a bare capillary (B) and (D) are from a FC-coated capillary. Compare (A) with (13) and (C) with (D). Condition: New capillaries with 75 μm I.D., 375 μm O.D. for bare capillary and 190 μm O.D. for FC-coated column. 65 cm total length, 50 cm effective length, 200v/cm electric field; 2.0%+ 1.4% PEO gel.

Figure 4: Comparison of DNA separation in bare capillary and in PVP-coated capillary. (A) and (C) are from a bare capillary (B) and (D) are from a PVP-coated capillary. Compare (A) with (13) and (C) with (D). Condition: 65 cm total length, 50 cm effective length, 200v/cm electric field; 1.5%+ 1.4% PEO gel.

Figure 5: Electropherograms of the 1^{st} run and 25^{th} run in a DB-wax coated capillary. Condition: New capillaries with 50 μm I.D., 375 μm O.D. for both. 55 cm total length, 41 cm effective length, 200v/cm electric field; 1.5%+1.4% PEO gel.

Figure 6: Electropherograms of the 1^{st} run and 25^{th} run in a FC- coated capillary. Condition: 65 cm total length, 50 cm effective length, 200v/cm electric field; 2.0%+ 1.4% PEO gel.

Figure 7: Electromicroscopy of a new and a degraded DB-wax capillary inner coating. "Good" was from a new capillary. "Bad" was from a degraded capillary.

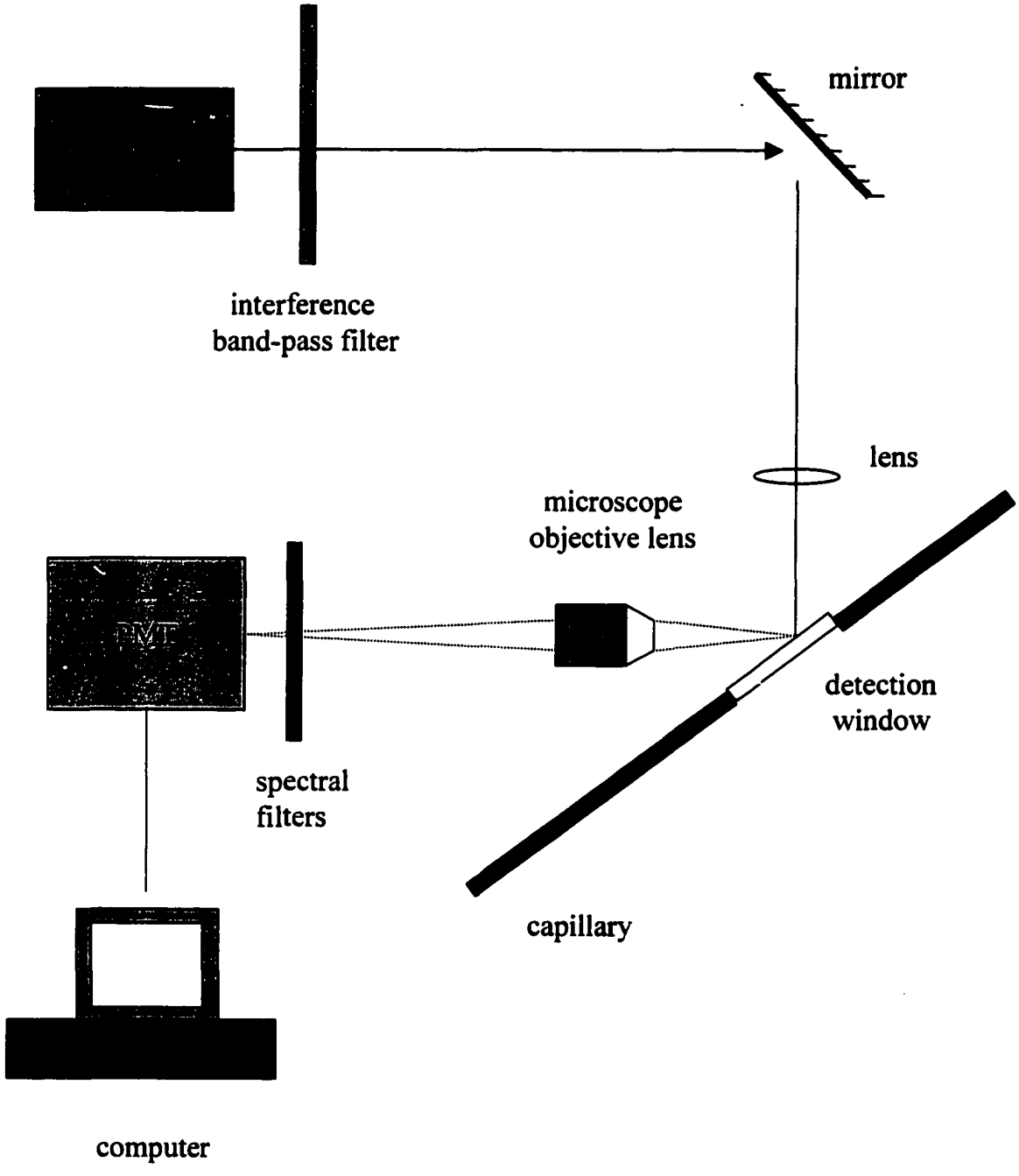

mirror
interference
band-pass filter
lens
microscope
objective lens
PMT
detection
window
spectral
filters
capillary
computer

Figure 1

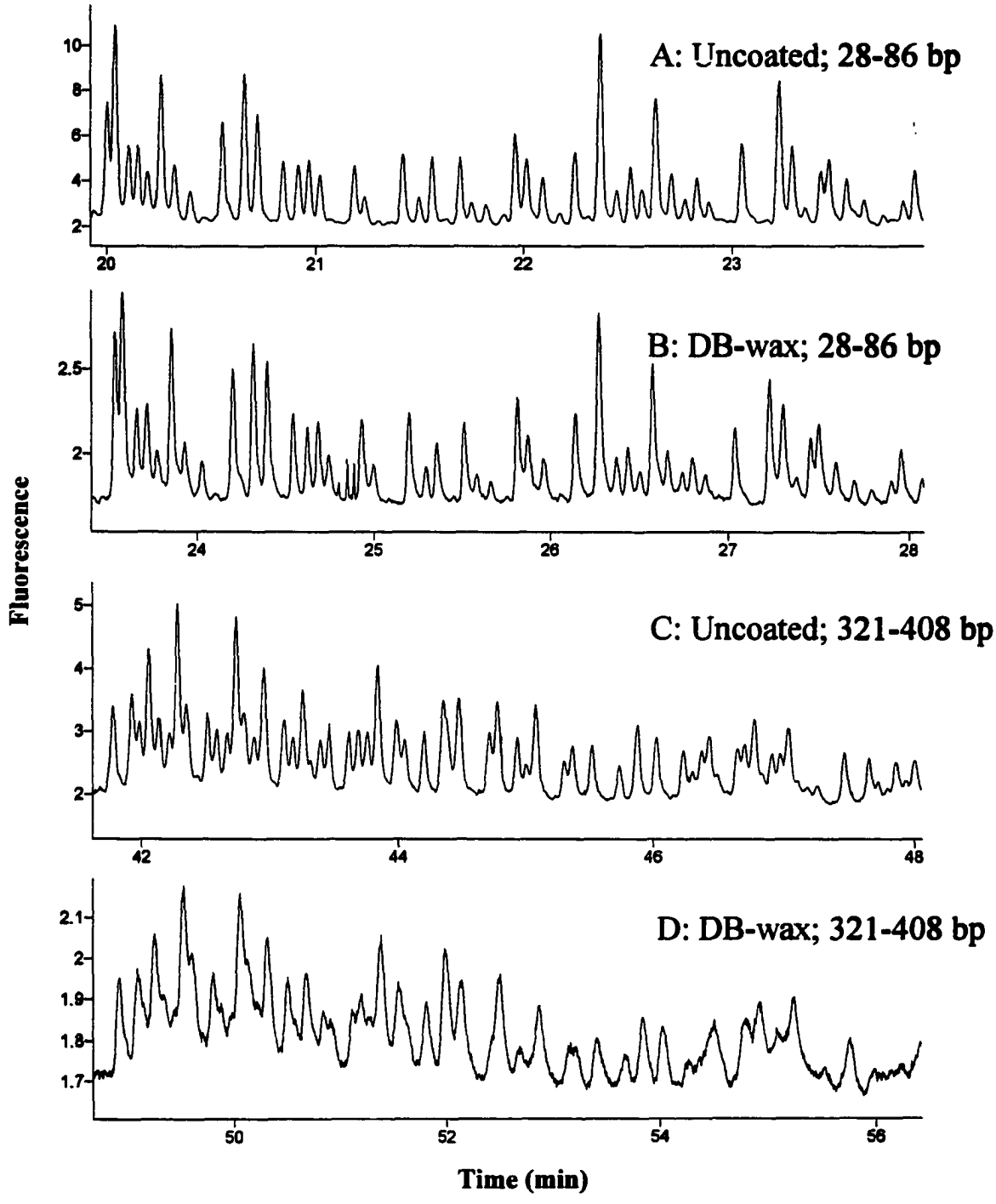

A: Uncoated; 28-86 bp
B: DB-wax; 28-86 bp
C: Uncoated; 321-408 bp
D: DB-wax; 321-408 bp
Fluorescence
Time (min)

Figure 2

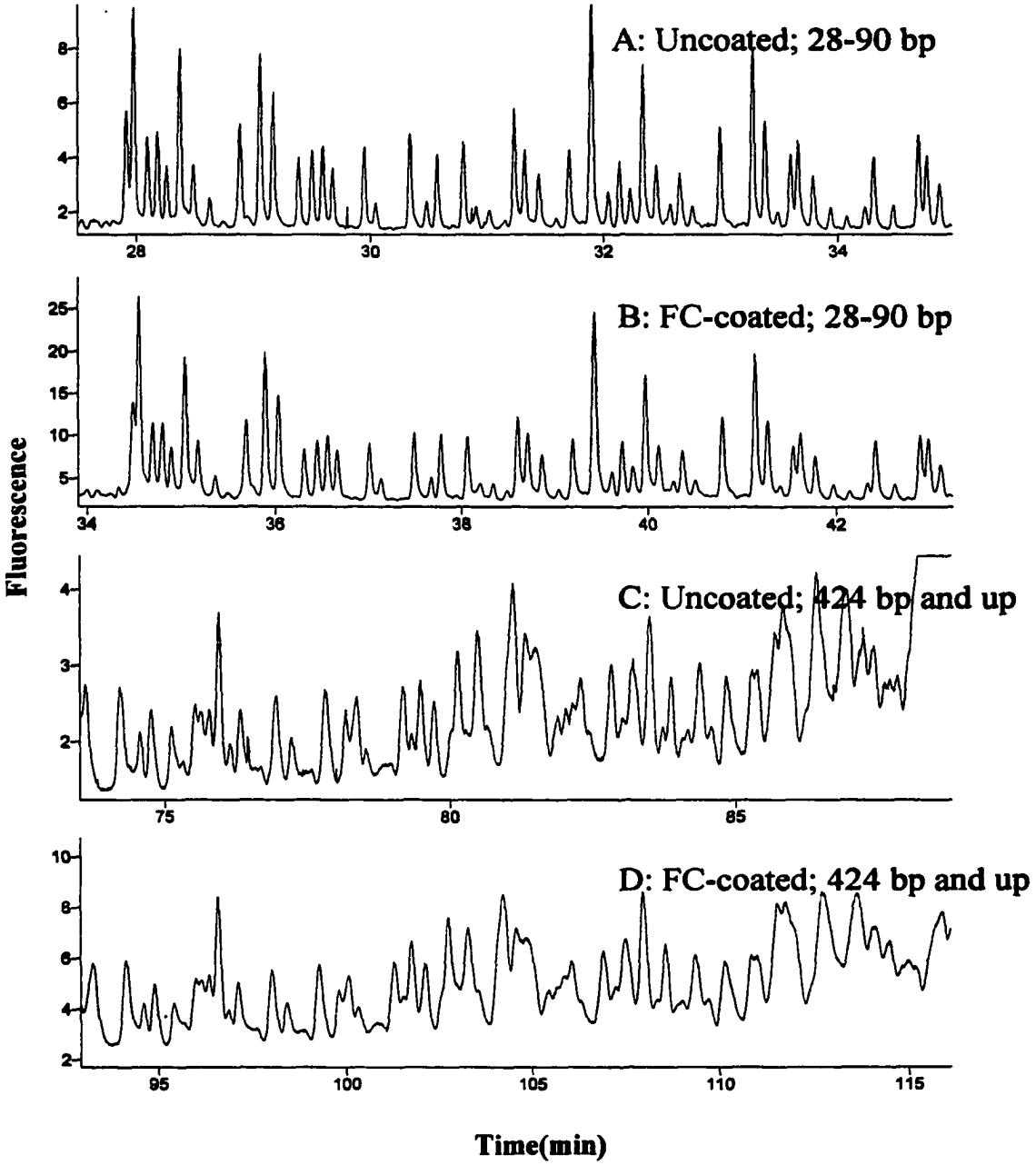

A: Uncoated; 28-90 bp
B: FC-coated; 28-90 bp
C: Uncoated; 424 bp and up
D: FC-coated; 424 bp and up
Fluorescence
Time(min)

Figure 3

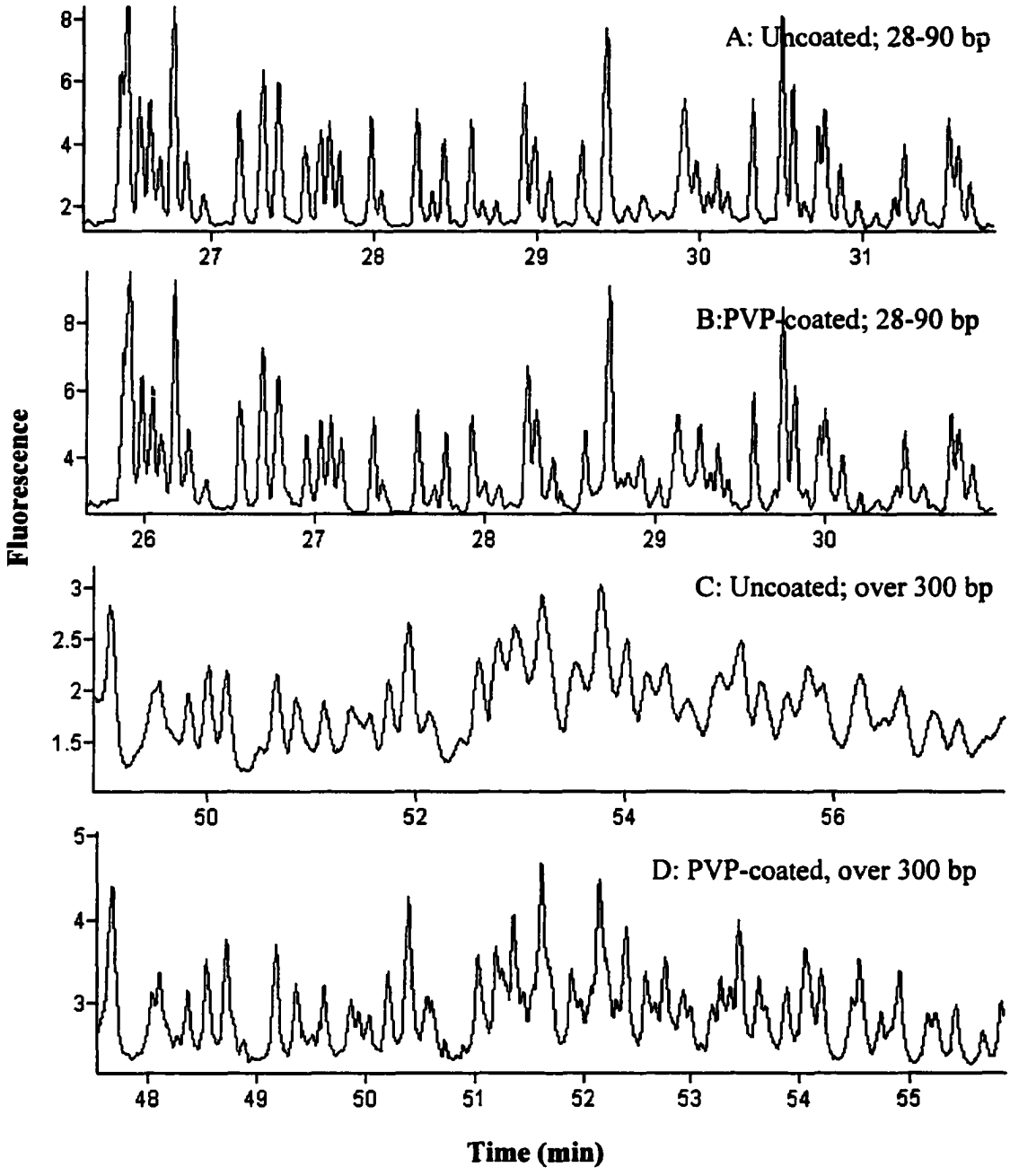

A: Uncoated; 28-90 bp
B:PVP-coated; 28-90 bp
C: Uncoated; over 300 bp
D: PVP-coated, over 300 bp
Fluorescence
Time (min)

Figure 4

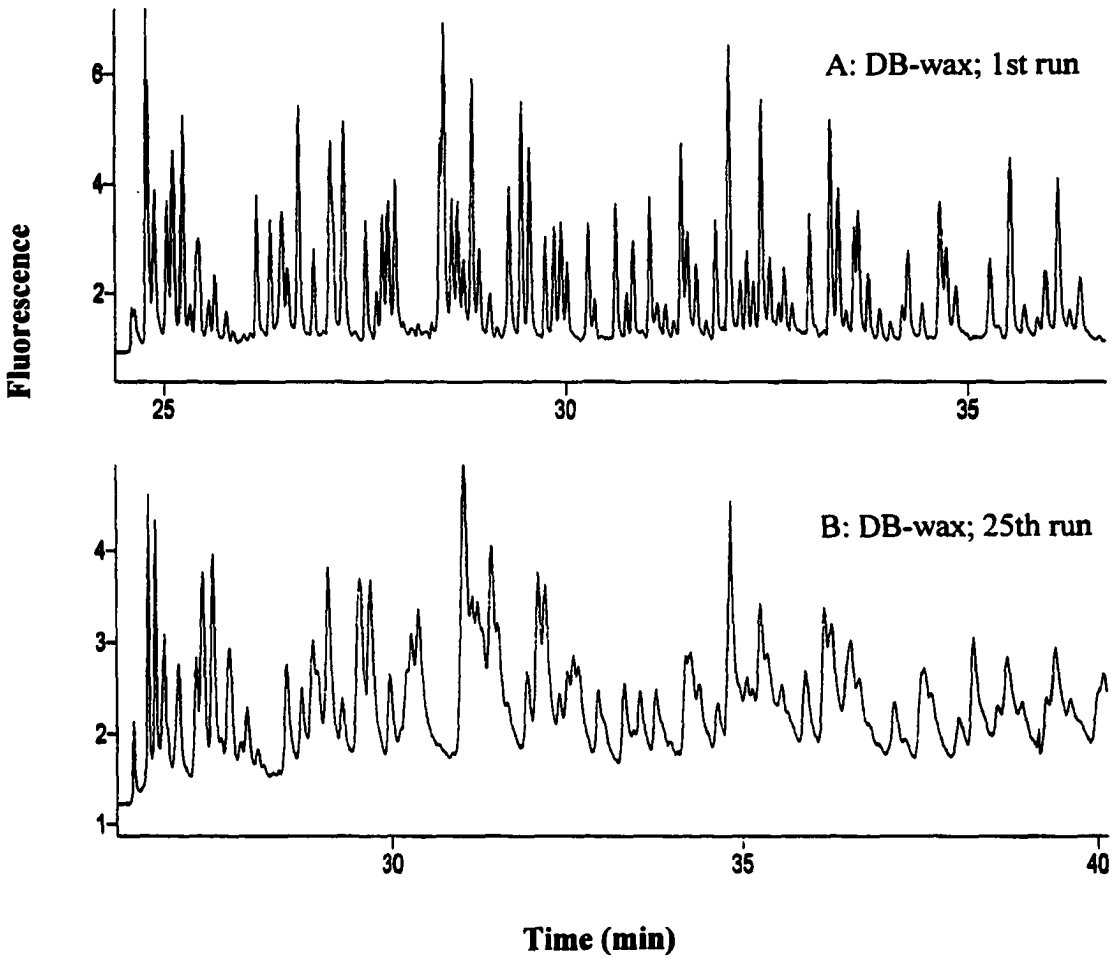

A: DB-wax; 1st run
B: DB-wax; 25th run
Fluorescence
Time (min)
6
4
2
25
30
35
4
3
2
1
30
35
40

Figure 5

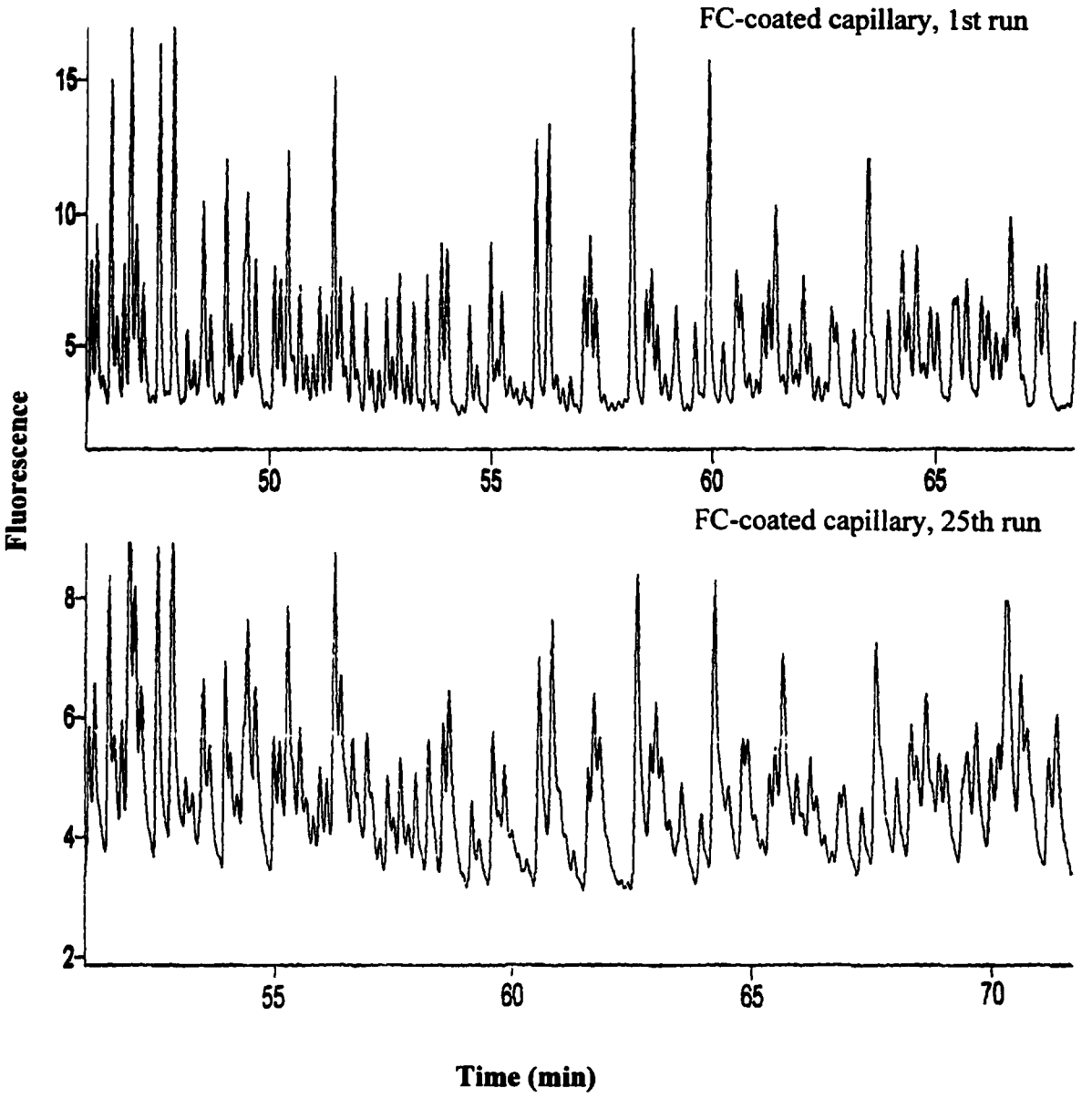

FC-coated capillary, 1st run
15
10
5
50
55
60
65
FC-coated capillary, 25th run
8
6
4
2
55
60
65
70
Fluorescence
Time (min)

Figure 6

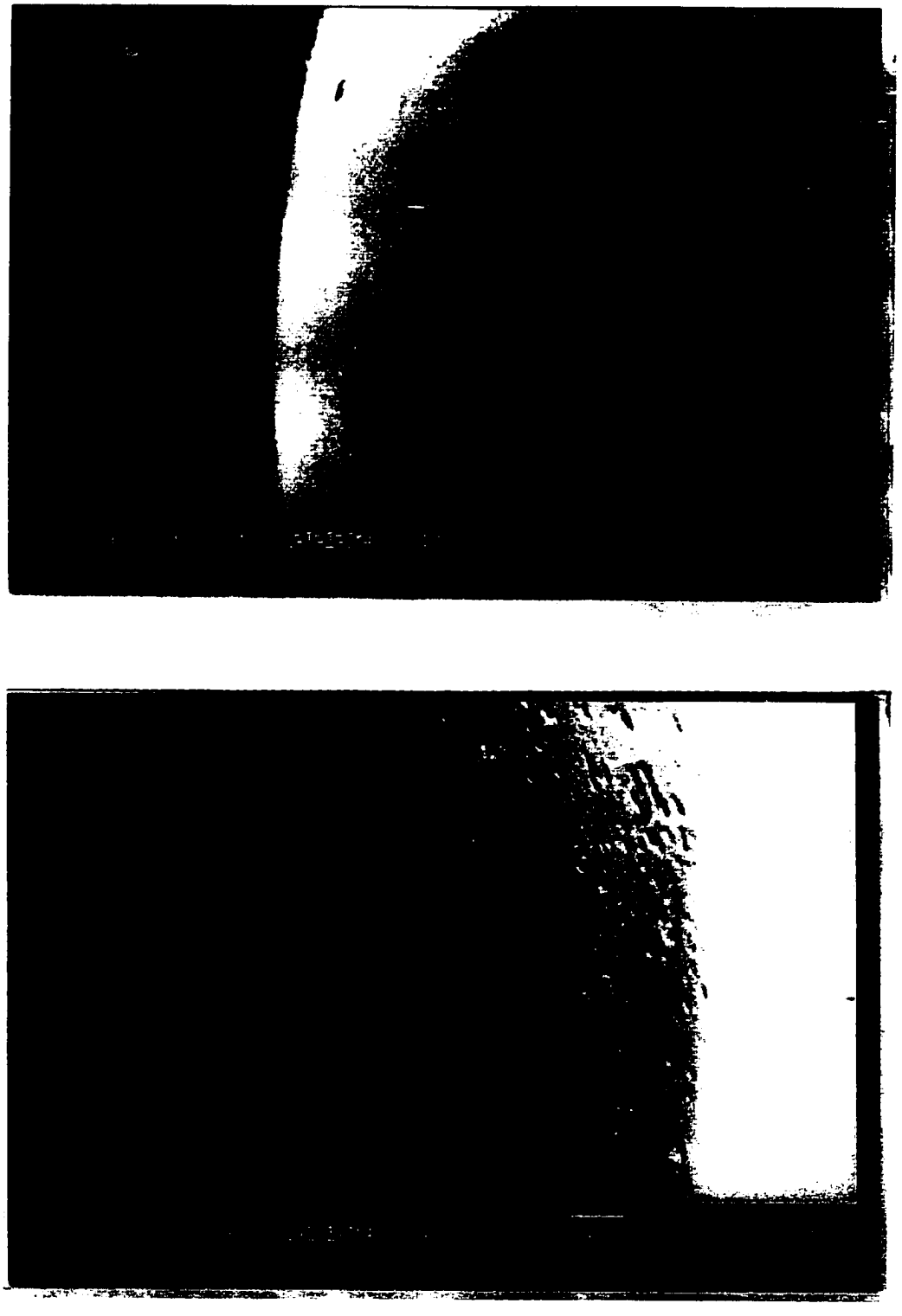


Figure 7

CHAPTER 3. A NEW MATRIX FOR DNA SEPARATION: GENOTYPING AND SEQUENCING USING POLY(VINYLPYRROLIDONE) SOLUTION IN UNCOATED CAPILLARIES

A paper published in Analytical Chemistry [1]

Qiufeng Gao and Edward S. Yeung

Abstract

We report a new sieving matrix for DNA separation based on commercially available polyvinylpyrrolidone (PVP). The new sieving matrix has a very low viscosity at moderate concentrations, e.g. 27 cp at 4.5%. Its excellent self-coating property can reduce electroosmotic flow to a negligible level. Column regeneration between runs is very simple and effective. Successful separations were achieved in uncoated capillaries. For genotyping, we show that D1S80, 100-bp ladder and Amelogenin sex determination system can be baseline separated as double-stranded DNA, and vWF, TH01, TPOX and CSF1PO short tandem repeats (STR) can be separated with single-base resolution as single-stranded DNA in this new matrix. Sequencing of M13mp18 showed good resolution up to 500 bases in a solution of high molecular weight fraction extracted from commercially available PVP. The feasibility of adaptation to a multi-capillary array system is discussed.

---

[1] Reprinted with permission from Analytical Chemistry 1998, 70, 1382-1388

Introduction

Capillary electrophoresis (CE) has become an attractive alternative to the slab gel in a wide range of DNA separations including DNA sequencing[1-6] and genotyping.[7,8] The primary advantages of CE include fast speed, increased separation efficiency, and the possibility to automate sample loading. Because cross-linked polyacrylamide (PA) is unstable under the high electric field typically applied in capillary electrophoresis, linear polymer solutions became the medium of choice for DNA separation. Recently, Barron[9] counted at least 12 water soluble polymers which have been used for DNA separation, including PA, cellulose and its derivatives, poly(ethylene glycol), polyvinyl alcohol, galacto-mannan, and agarose. Another new matrix, poly(N-dimethylacrylamide) (DMA) was also reported.[10] The effect of gel composition,[11-14] capillary,[12-14] and electric fields[13,14] on the separation has been studied systematically. In published reports, except for poly(ethylene oxide) (PEO) and DMA, all other polymers have to be used in coated columns to provide the high resolution needed for DNA sequencing. With few exceptions, the polymer solutions are also generally viscous.

Besides separation performance of the polymer solution, a major issue in capillary electrophoresis is column regeneration. Although there is no solid evidence, column degradation seems related to polymer solution viscosity.[6] The more viscous the polymer solution is, the more difficult it is to be replaced. The accumulation of residual polymer on the column wall will create inhomogeneous regions in the column which will degrade the separation efficiency. So a less viscous matrix is always preferred if the performance can be maintained.

Studies of satellite DNA and variable number of tandem repeats (VNTR) have shown that DNA is frequently polymorphic.[15] In practice, the data obtained from two VNTR probes can yield powers of exclusion which are vastly greater than can be obtained when HLA and isozyme typing are employed together. VNTR loci are usually analyzed under non-denaturing conditions. The human D1S80 locus is often used as a model for VNTR analysis because it has a large number of possible alleles.

The analysis of short tandem repeats (STR) is another typing method. STR loci consist of short, repetitive sequence elements of 3-7 base pairs in length. Compared to the VNTR system, the STR system is more tolerant to DNA damage. Separation of STR loci is usually performed under denaturing conditions because of the existence of heteroduplex and secondary structures due to the small structural difference (3-7 base pairs) between.[16] Even under denaturing conditions, hybridization can easily occur. Careful handling of the sample is required.[17] The combined regions vWF, TH01, TPOX and CSF1PO can be used as a STR model because these loci cover the size range from 179 base pairs to 327 base pairs without overlapping. The amelogenin human sex system can be used for evaluation of the resolving power of the new separation matrices. For DNA sequencing, M13mp18 control DNA has been used as a standard. Its sequence is well known and the peak assignment is straightforward.

In this paper, we report a new matrix based on commercially available polyvinylpyrrolidone (PVP), which has a very low viscosity. We demonstrate that PVP possesses favorable characteristics and superior separation performance in CE for both genotyping and sequencing in uncoated capillaries.

Experimental Section

Capillary and Reagents

Capillaries with 75 μm i.d. and 365 μm o.d. were purchased from Polymicro Technologies (Phoenix, AZ). All chemicals for making 1xTBEbuffer were obtained from Sigma (St. Louis, MO); ethidium bromide (EthB) was from Molecular Probes (Eugene, OR); coumarin 334 was from Eastman Kodak (Rochester, NY); polyvinylpyrrolidone (PVP) with M_r = 1,000,000 was from Polyscience (Warrington, PA); and urea was from ICN Biomedicals (Aurora, OH).

DNA Samples

ds-DNA. AmpliFLP D1S80 VNTR ladder was obtained with PCR amplification kit from Perkin-Elmer Applied Biosystems Division (PE-ABI, Foster City, CA) and Amelogenin sex system was from Promega (Madison, WI).

ss-DNA. CTTv STR system was purchased with PCR amplification kit from Promega one week before each experiment and stored in a –20 °C freezer. Unlabeled primer pair and 5'-fluorescein-labeled primer pair for TH01 locus (5'-GTGGGCTGAAAAGCTCCCGATTA-T-3' and 5'-ATTCAAAGGGTATCTGGGCTCTGG-3') were custom synthesized by the Nucleic Acids Facility of Iowa State University (Ames, IA) and Integrated DNA Technology (Coralville, IA), respectively. PCR for the alleles of TH01 locus samples was obtained using custom synthesized DNA as the primer pair and diluted TH01 standard ladder as the template. The mixture of commercial standard TH01 ladder and TH01 9.3 (4:1, v/v) was diluted 1000 times with 1xTE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The modification to the protocol suggested by Promega is as follows: 2 cycles of denaturing at 95

°C for 1 min, annealing at 65 °C for 30 s and extension at 70 °C for 30 s; 40 cycles of denaturing at 95 °C for 1 s, annealing at 65 °C for 30 s, and extension at 70 °C for 30 s. The thermal cycler was an Idaho Technology Thermal Cycler.

DNA Sequencing. Dye-labeled primer sequencing kit was purchased from PE-ABI. M13mp18 control DNA was obtained from Life Science. The parameters of cycle sequencing with T terminator only are as follows: hold at 95 °C for 3 min, 15 cycles of denaturing at 95 °C for 2 s, annealing at 50 °C for 1 s, and extension at 65 °C for 40 s; 15 cycles of denaturing at 95 °C for 2 s, annealing at 65 °C for 0 s, extension at 65 °C for 40 s. The product was purified by ethanol precipitation method. The sample was then dried by vacuum for 2 min. The dried sample was kept in a –20 °C freezer until used.

Polymer Solution and Buffer Preparation

ds-DNA. 1xTBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) was used as the buffer. Commercially available PVP powder (M_r = 1,000,000) was added to the buffer solution to yield 0.05% ~ 4.5% (w/w) polymer solutions. After adding the powder, the mixture was shaken for 5 min followed by vacuum degassing for another 5 min, or the mixture was stirred with a magnetic stirrer for 1 h without degassing. This solution was used immediately for viscosity measurement, electroosmotic flow (EOF) test and double-stranded DNA separation.

ss-DNA. For the STR system, denaturing 1xTBE buffers with 3.5 M urea and with 7 M urea were used. Similar to the above procedure, PVP (M_r = 1,000,000) was added to the buffer solutions to form a 5.5% (w/w) polymer solution.

DNA Sequencing. For DNA sequencing, a fraction of high molecular weight PVP extracted from commercial PVP was used. First, PVP was dissolved in deionized water to form a 5% solution. Then, 3 volumes of acetone:isopropanol (1:1, v/v) was added into the polymer solution. The mixture was shaken vigorously and the turbid solution was allowed to stand overnight. The upper layer was discarded. The viscous bottom layer was collected on a Teflon plate and dried naturally in room air. The recovery yield is about 50%. Sequencing matrix of 5% (w/w) polymer solution was made by stirring the mixture of high molecular weight PVP solid and 1xTBE with 3.5 M urea buffer solution for 1 h.

Viscosity Measurement

Viscosity measurement was performed with a modified-Ubbelohde type capillary viscometer set in a thermostated water bath at 20 ± 0.1 °C. Density differences between polymer solutions were corrected in the calculations. Duplicate experiments were performed to verify the data. The averages of the duplicate results were used to construct the curves of viscosity vs. concentration.

EOF Measurement

UV-visible absorption detection was used to measure electroosmotic flow (EOF) in 1xTBE buffer, 0.1% and 1% PVP (M_r = 1,000,000) solutions in 1xTBE buffer. The incident light wavelength was set at 452 nm. A 50-cm long capillary with 20 cm effective length from the anode to the detection window was filled with polymer solution without any pretreatment. The polymer solution served as buffer in the reservoirs as well. 50 μM coumarin 334 in 1xTBEbuffer solution was used as a neutral marker. The injection was performed with

gravity at 10 cm above the cathode for 30 s. The applied electric field was 250 V/cm. Duplicate experiments were performed in each case.

CE Setup and Procedure

ds-DNA. A He-Ne green laser at 543.5 nm (Melles Griot, Irvine, CA) was used for excitation. The experimental setup is similar to that described before.[18] A RG610 cutoff filter was used to block the stray light. The fluorescence signal from the photomultiplier tube (PMT) was transferred directly through a 10-kΩ resistor to a 24-bit A/D converter at 5 Hz (Lawson Labs, Kalispell, MT) and stored in a 486/33 computer (Gateway, North Sioux City, SD).

The capillaries used for ds-DNA separation had a total length of 35 cm and an effective length of 30 cm. The electric field varied from 120 V/cm to 180 V/cm. 1xTBE was the running buffer, with 1 μg/ml EthB incorporated as the intercalating dye. The same concentration of EthB was introduced into the polymer solutions acting as the sieving matrix. The column was equilibrated at the running voltage for 1 min before injection. The D1S80 ladder and Amelogenin ladder were injected directly for 5 s at the running voltage. The capillary was rinsed with deionized water three times and methanol once between runs. After rinsing and refilling with matrix, the capillary was ready to be used for the next run.

ss-DNA. An air-cooled Ar+ laser with 488 nm emission (Uniphase, San Jose, CA) was used. A 488-nm interference bandpass filter was used between the laser and the capillary. The whole setup was similar to that above except that a combination of a 488-nm notch filter and a 520-nm interference bandpass filter was used in front of the PMT. Data acquisition was accomplished with a 10-kΩ resistor, a 24-bit A/D interface at 4 Hz (Justice Innovation,

Palo Alto, CA; Model DT2802) and a 486/33 computer. The total length of the capillary was 55 cm, with an effective length of 50 cm. The running voltage was 11 kV. The capillary was filled with PVP solution by a 3 ml syringe and the capillary was pre-run for 2 min at the running voltage before injection. The standard CTTv sample was made and treated exactly the same way as described before.[17] The PCR product from TH01 was mixed with formamide (1:1) and heated at 95 °C for 3 min. Injection was performed at 11 kV for 20 s while the sample was still being heated at 95 °C. The high temperature runs were performed by circulating hot water around the capillary with a pump. After separation, the capillary was cleared and regenerated by fast flushing with deionized water and methanol.

DNA Sequencing. The same setup for ds-DNA was used. The total length of the capillary was 55 cm, with an effective length of 50 cm. The running voltage was 8.25 kV. After being filled with polymer solution, the capillary was equilibrated for 2 min. The dry M13mp18 sample was cooked at 95 °C for 3 min with 2 μl formamide:EDTA (5:1). Injection was performed at 150 V/cm for 15 s. The capillary regeneration procedure was similar to that described for ds-DNA.

Results and Discussion

Viscosity of PVP Solutions

Among all the water-soluble neutral polymers, PVP is an unusually soluble one. Its water solution also has a much lower viscosity compared to other polymer solutions with the same concentration and molecular weight. Figure 1 shows the viscosity of PVP (M_r = 1,000,000) as a function of concentration in 1xTBE buffer. At the moderate concentrations used for DNA separations, the viscosity is very low, e.g. 27 cp for the 4.5% solution and only

3 cp for the 1% solution. The advantage is obvious. With such low viscosity, the sieving matrix can be easily filled and replaced, even with the pressure generated by a common 3-ml syringe. This is also good for column regeneration, which requires one to remove all materials near the column surface,[6,17] because Pouiselle flow implies that higher pressure or longer time is needed to clean out the material near the capillary wall. This is particularly suitable for a capillary array system, which is more problematic in terms of gel filling and regeneration. Homogeneity is another concern when the gel is viscous. With PVP, homogeneous solutions can be made within a short time (see experimental section). This indicates that fast matrix setup and sample separation can be performed in a robust instrument with multiple capillaries.

Figure 2 is the plot of reduced viscosity vs. concentration. Reduced viscosity is defined as (relative viscosity – 1)/concentration. The critical aggregation concentration (C*) for a polymer solution can be estimated from such a plot.[19] A significant increase in the slope is evidence that the C* is reached. We estimate C* is about 0.6% for PVP (M_r = 1,000,000).

Electroosmotic Flow

EOF, which is used as the driving force in CZE and in electrochromatography, is considered as a disadvantage in DNA separations because the counterflow means longer separation times and causes the sample zone to broaden due to diffusion. In the worst case, it can expel the sieving matrix from the capillary.[20] So, in most DNA separation experiments, bonded coatings were used on columns to minimize EOF. An alternative way to reduce EOF is to use dynamic coatings of linear polymers in bare fused-silica columns. In the previous studies, our group demonstrated that EOF could be substantially reduced in PEO matrix due

to the dynamic coating formed by PEO on the capillary wall as well as the high viscosity of the matrix.[6] We also developed a method to monitor the change of EOF during electrophoresis in PEO.[21] The results showed that PEO coating is stable at pH 7 but unstable at higher pH, such as pH = 8.2, which is the typical pH value for DNA separations. At pH 8.2, 0.2% PEO solution still shows an EOF mobility of 2.5 x 10^{-4} cm^2 V^{-1} s^{-1}, which is 50% of the EOF mobility in plain buffer. Since the PVP solution is much less viscous than PEO gel, we are more concerned with a changing EOF over a long period of time. Our way to measure the EOF mobility is to calculate the EOF velocity from the migration time of an uncharged compound and the capillary length.[22]

Table 1. Electroosmotic flow mobilities for various 1xTBE buffer solutions in fused-silica capillary tubes

	0% PVP un-coated	0.5% PVP un-coated	1% PVP un-coated	0% PVP DB-wax coated
migration time for neutral marker	3 min	49 min	3 h 40 min	2 h 12 min
μEOF (cm^2 V^{-1} s^{-1})	4.5 x 10^{-4}	2.7 x 10^{-5}	6.0 x 10^{-6}	1.0 x 10^{-5}

For PVP added to 1xTBEbuffer solutions, the results showed there was no significant difference between the migration times in consecutive runs with the same solution. But the

migration times in 0.1% and 1% PVP solutions were much longer than that in the buffer itself. By comparing the EOF values in three solutions in a bare capillary and in 1xTBEin a DB wax coated capillary, it can be seen that the EOF was reduced by over one order of magnitude in PVP solution (Table 1). In 1% PVP, the EOF suppression effect is even better than that in a coated capillary. This fact might be explained by surface adsorption. PVP has stronger hydrogen bonding with the residual hydroxyl groups on the capillary wall than PEO due to the existence of hydrophilic carbonyl groups.[23] Thus, PVP molecules can form a denser coating on the wall even at higher pH. In addition, since the coating is dynamic and the solution is not viscous, the coating may be more uniform over the whole capillary. Therefore, PVP may be an excellent candidate as a dynamic coating material even in capillary zone electrophoresis. For PEO, we found that the quality of separation depends on the procedure employed for filling the polymer solution. The more viscous solution sometimes results in coating defects that affect the electrophoretic resolution.

Separation Performance

ds-DNA. Figure 3 shows the separation of D1S80 alleles in two different PVP solutions. The pooled D1S80 ladder has total of 27 alleles ranging from 369 bp to 801 bp with 16 bp between adjacent alleles. Usually it is hard to separate all 27 alleles, especially for the last several fragments. Here, we found that in a PVP solution of moderate concentration (4.5%), all 27 alleles can be separated with good resolution (Figure 3a). As the concentration decreased and the voltage applied to the capillary decreased as well, the separation became more efficient and less time consuming (Figure 3c). The same trend can be found over the PVP concentration range from 4.5% down to 1% (w/w). Further decreases

in the PVP concentration to 0.5% will result in peak broadening and bad separation. The improvement in separation can be explained by a combination of two factors. First, a lower concentration can increase the mesh size of the polymer solution, which allows more efficient passage of large DNA fragments.[24,25] Even though there is greater diffusion for the smaller fragments, sufficient resolving power exists to preserve the performance. Second, a lower electric field strength can reduce the reptation effect, thus giving better separation for large DNA fragments.[26] When the polymer concentration drops below the critical concentration, interaction between polymer molecules becomes less efficient and even local entanglement may not exist. The interaction between diluted polymer and DNA fragments alone was not sufficient to recognize the 16 base pair difference between alleles at this size range.

Among the non-crosslinked polymers that were used in high resolution DNA separation in published reports, only PEO and DMA have been used in bare capillaries. But, multiple steps and long conditioning times are needed to regenerate the column surface.[6] By contrast, with simple flushing in 3 min between runs, the column can be reused at least 30 times if PVP is used as the sieving matrix. The performances of the 1st run and the 31st run of D1S80 in 4.5% PVP are compared in Figure 3a and Figure 3b. The separation efficiency and separation time were both reproducible with alleles (14,41) appearing at (21.4 min, 25.3 min) and (21.5 min, 25.5 min) in the 1st and the 31st runs respectively.

There is a reasonable explanation for the reliable regeneration of capillaries. On the one hand, PVP is readily soluble in water and in methanol. PVP in water solution has a fairly low viscosity (27 cp). Any polymer adsorbed on the walls can be easily washed out when rapidly flushed with water and methanol. On the other hand, when new matrix is filled into the capillary, the low viscosity, homogenous polymer solution facilitates PVP molecules to

be adsorbed onto the wall to form a complete and uniform coating. As discussed above, EOF will be dramatically suppressed by the PVP coating and therefore, high separation efficiency will be maintained.

Figure 4 shows the separation of the amelogenin human sex determination system. It contains a 212 bp fragment generated from a specific segment of the human X chromosome and a 218 bp fragment generated from the corresponding human Y chromosomal segment. Because there is only a 6 bp difference in the two fragments, this sample is usually treated as a short tandem repeat (STR) system and separated under denaturing conditions. It can also be analyzed as ds-DNA under non-denaturing conditions.[27] Separating and detecting DNA in double-stranded form has certain advantages over single-stranded form. It does not require chemical bonding of dye to the primer pair before PCR and needs less sample treatment. We treated the amelogenin system as double-stranded DNA in our experiment. Since DNA fragments of 212 bp and 218 bp are not very large, a higher concentration (4.5% PVP) was used for the separation.[6,9] To improve resolution, a longer capillary (50 cm effective length) was used. It can be seen that the two fragments were completely resolved despite the presence of some impurities and heteroduplex component in the sample.[16]

ss-DNA. The human CTTv STR system contains vWF, TH01, TPOX and CSF1PO loci ranging from 139 bp to 327 bp with 4 bp difference between alleles in each locus. A denaturing condition with 3.5 M urea was utilized previously in PEO gel and was shown to be sufficient for keeping the CTTv ladder denatured.[17] The separation of CTTv ladder in 5.5% PVP with 3.5 M urea is shown in Figure 5. Despite the extra peaks and shoulders caused by secondary components, baseline resolution was still obtained in the CSF1PO locus, which is the most difficult group of the four loci to be separated. Careful examination of the

electropherogram pattern indicates a one to one correspondence between the primary peaks (marked with asterisk) and the secondary peaks over the entire electropherogram. The intensity ratio of the primary peaks to the secondary peaks are the same for all the alleles in each locus. Since this could be due to insufficient denaturing power of this PVP solution, two sets of extreme denaturing condition were then investigated.

First the separation of the same CTTv sample was performed with 3.5 M urea in a capillary heated to 83 °C while the injection end was immersed in 95 °C 1xTBEbuffer without urea (Figure 5b). At such a high temperature, the urea inside the capillary will be decomposed and the decomposition products can accumulate into bubbles. Our solution to this problem is to apply a higher electric field of 300 V/cm across the capillary. Since the major decomposition products of urea in hot water are ammonia and cyanic acid,[28] these will be present in the form of salt if their concentrations are very low. Salts will be driven out of the capillary by the high voltage applied immediately after the polymer solution is filled into the capillary. When the migration rate of the decomposition products is equal to the formation rate, a steady state is established such that the concentration of the decomposition products is below that required for the bubble formation. It can be seen from Fig. 5b that the migration time was substantially reduced and the separation performance was still excellent under this condition. It can also be seen that the secondary peaks moved closer to the primary peaks than that in Fig. 5a but the intensity ratios remained the same.

Under an alternative condition, the locus of vWA was tested in 5.5% PVP with 7 M urea at 65 °C (Figure 5c). This condition was reported to be sufficient to eliminate secondary structures in DNA sequencing.[29] The result is similar to that in Figure 5b—the secondary

peaks moved closer to the primary ones than that in Fig. 5a but the peak intensity ratio remained the same.

Figures 5b and 5c show that it is unlikely that rehybridization causes the secondary peaks. We then investigated the purity of the primers supplied with the genotyping kit. Surprisingly, each primer pair showed two peaks in the electropherogram when there should only be one. The electropherogram of the commercial primer pair for TH01 is shown in Figure 6a. Figure 6b is the electropherogram of our custom synthesized primer pair. Figure 6c is the electropherogram of the commercial primer pair mixed with our own primer pair. Our custom synthesized primer pair contained only one fluorescein-labeled strand according to ref. 30. It is clear that our dye-labeled primer matches one of the two structures (marked with asterisk) of the commercial dye-labeled primer. The other is therefore an altered structure which cannot be distinguished from the original one either in slab gel or in PEO matrix. The secondary peaks in Fig. 5 are due to PCR amplification associated with the altered structure of the dye-labeled primer.

Figure 7 shows the separation of the PCR products based on our custom synthesized primer pair and commercial TH01 alleles as the template. These alleles range from 179 bp to 203 bp with 4 bp repeat difference. We selected the TH01 locus for investigation because this locus has a common "9.3" allele which is a single-base deletion of the allele 10. 9.3 stands for 9 repeats plus 3 bases. When the "9.3" allele is added into the TH01 locus, it can provide a critical evaluation of the separation performance of the new matrix. As expected, the PCR products showed a clear pattern of single peaks without any secondary structure. This serves to identify the peaks marked with asterisks in Figures 5a and 5b. Although the separation of allele "9.3" with allele 10 was not complete, it was sufficient for distinguishing

between them. The above data indicate that with properly controlled PCR and the method described in ref. 17, the 5.5% PVP and 3.5 M urea matrix can be used for the assignment of unknown genotype peaks.

DNA Sequencing. We have shown above the feasibility of using PVP as a sieving matrix for genotyping. Single-base separation was observed under denaturing conditions (Fig. 7). We then tested a 7% PVP solution, in 1xTBEwith 3.5 M urea, for separating DNA sequencing fragments. The separation of T-terminated M13mp18 fragments showed reasonable separation up to 350 bases.

In previous work,[6,31] it has been demonstrated that solutions of higher molecular weight polymer could extend the read length in DNA sequencing. We adopted a method[32] to extract the higher molecular weight fraction from commercially available PVP (M_r = 1,000,000) and made a solution of this extract for sequencing. PVP is readily soluble in water and alcohol but not in acetone. In a mixture of water and acetone, the solubility depends greatly on the molecular weight and composition of the solvent. The higher molecular weight fraction is less soluble than the lower molecular weight fraction. Higher acetone concentrations will precipitate more PVP. Since commercial PVP has a wide polydispersity,[33] it is possible to precipitate the higher molecular weight fraction by selecting a certain composition of water-acetone solvent. In our experiments, 10% isopropanol was added into acetone first to make the dependence of precipitation on composition less steep to make it easier to control the process.

Figure 8 shows the separation of T-terminated M13mp18 sequencing ladder in 5% extracted PVP in 1xTBEbuffer with 3.5 M urea. The sample was also tested with PEO matrix for comparison. Single-base separation was extended from 350 bases in 7%

commercial PVP to 530 bases in 5% extracted PVP. Partial separations were obtained at 439/440, 491/492, and 501/502 bases. The blot at 426 bases was due to cycle sequencing. The extracted PVP also gave improved resolution for genotyping when compared to Figure 7. On-going studies in our laboratory will be focused on mobility shifts and spectral properties of the sequencing dyes in PVP solutions, so that actual sequencing applications can be initiated.

Our measurement indicates that the viscosity of the 5% extracted PVP solution was about 60 cp. For comparison, matrix based on 1.5% PEO 600,000 M_r, 1.4% PEO 8,000,000 M_r has a viscosity of 1200 cp, and non-crosslinked PA (6% T) has a viscosity of 4900 cp at room temperature.[6] The low viscosity of PVP solutions is one of the most distinctive features that makes it a unique matrix for DNA separations.

Conclusions

As a new matrix, 1,000,000 M_r PVP and the high molecular weight fraction extracted from 1,000,000 M_r PVP can be applied for DNA separation for genotyping and sequencing with good resolution. By varying the concentration and electric field, genotyping can be accomplished with commercial 1,000,000 M_r PVP. Single base pair resolution was reached for ss-STR systems. With extracted PVP, single-base separation can be achieved beyond 500 bases under sequencing conditions. The favorable characteristics of PVP solutions of a very low viscosity and efficient EOF suppression make capillary regeneration much easier and therefore, makes PVP most suitable for high-throughput capillary array electrophoresis.

Acknowledgments

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 National Institutes of Health under grant No. HG-01385.

References

1. Cohen, A. S., Najarian, D. R., Paulus, A., Guttman, A., Smith, J. A., Karger, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *85*, 9660-9663.
2. Smith, L. M. *Nature* 1991, *349*, 812-813.
3. Drossman, H., Luckey, J. A., Kostichka, A. J., D'Cunha, J., Smith, L. M. *Anal. Chem.* **1990**, *62*, 900-903.
4. Swerdlow, H., Zhang, J. Z., Chen, D. Y., Harke, H. R., Grey, R., Wu, S., Dovichi, N. J., Fuller, C. *Anal. Chem.* **1991**, *63*, 2835-2841.
5. Hung, X. C., Quesada, M. A., Mathies, R. A. *Anal. Chem.* **1992**, *64*, 2149-2154.
6. Fung, E. N., Yeung, E. S. *Anal. Chem.* **1995**, *67*, 1913-1919.
7. Zhang, N. Y., Yeung, E. S. *Anal. Chem.* **1996**, *68*, 2927-2931.
8. Wang, Y., Ju, J., Carpenter, B. A., Atherton, J. M., Sensabaugh, G. F., Mathies, R. A. *Anal. Chem.* **1995**, *67*, 1197-1203.
9. Barron, A. E., Sunada, W. M., Blanch, H. W. *Electrophoresis* **1996**, *17*, 744-757.
10. Menchen, S., Johnson, B. F., Madabhushi, R., Winnik, M. *Proc. SPIE* **1996**, *2680*, 294-303.
11. Figeys, D., Dovichi, N. J. *J. Chromatogr.* **1993**, *645*, 311-317.
12. Nishikawa, T., Kambara, H. *Electrophoresis* **1994**, *15*, 215-220.

13. Luckey, J. A., Smith, L. M. *Anal. Chem.* **1993**, *65*, 2841-2850.

14. Issaq, H. I., Chan, K. C., Muschik, G. M. *Electrophoresis* **1997**, *18*, 1153-1158.

15. Nakamura, Y., Leppert, M., O'Connell, P. et al. *Science* **1987**, *235*, 1616-1622.

16. Wang, Y., Wallin, J. M., Ju, J., Sensabaugh, G. F., Mathies, R. A. *Electrophoresis* **1996**, *17*, 1485-1490.

17. Zhang, N. Y., Yeung, E. S. *J. Chromatogr.* **1997**, *768*, 135-141.

18. Chang, H. T., Yeung, E. S. *J. Chromatogr.* **1995**, *669*, 113-123.

19. Goodwin, J. W., Hughes, R. W., Lam, C. K. et al. *Advances in Chemistry Series* **1989**, *223*, 365-378.

20. Yin, H. F., Lux, J. A., Schomberg, G. *J. High Resolut. Chromatogr.* **1990**, *13*, 624-627.

21. Preisler, J., Yeung, E. S. *Anal. Chem.* **1996**, *68*, 2885-2889.

22. Walbroehl, Y., Jorgenson, J. W. *Anal. Chem.* **1986**, *58*, 479-481.

23. Parnas, R. S., Chaimberg, M., Taepaisitphongse, V., Cohen, Y. *J. Colloid and Interface Sci.* **1989**, *129*, 441-450.

24. Grossman, P. D., Soane, D. S. *Biopolymers* **1991**, *31*, 1221-1228.

25. Baba, Y., Ishimaru, N., Samata, K., Tsuhako M. *J. Chromatogr.* **1993**, *653*, 329-335.

26. Viovy, J. L., Duke, T. *Electrophoresis* **1993**, *14*, 322-329.

27. Isenberg, A. R., McCord, B. R., Koons, B. W., Budowle, B., Allen, R. O. *Electrophoresis* **1996**, *17*, 1505-1511.

28. Blakeley, R. L., Teston, A., Andrews, R. K., Zerner, B. *J. Am. Chem. Soc.* **1982**, *104*, 612-614.

29. Kleparnik, K., Foret, F., Berka, J., Goetzinger, W., Miller, A. W., Karger, B. L. *Electrophoresis* **1996**, *17*, 1860-1866.

30. Edwards, A., Civitello, A., Hammond, H. A., Caskey, C. T. *Am. J. Hum. Genet.* **1991**, *49*, 746-756.

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

32. Levy, G. B., Caldas, I., Fergus, D. *Anal. Chem.* **1952**, *24*, 1799-1803.

33. Campbell, H., Kane, P. O., Ottewill, I. G. *J. Polymer Sci.* **1954**, *12*, 611-620.

Figure Captions

Figure 1. Viscosity of PVP as a function of concentration (M_r = 1,000,000) in 1xTBEbuffer solution. Temperature: 20 ± 0.1 °C.

Figure 2. Reduced viscosity of PVP as a function of concentration (M_r = 1,000,000) in 1xTBEbuffer solution. Temperature: 20 ± 0.1 °C.

Figure 3. Electropherogram of D1S80 allelic ladder under different conditions: (a) 4.5% PVP, 180 V/cm, 1st run; (b) 4.5% PVP, 180 V/cm, 31st run; and (c) 1.0% PVP, 120 V/cm, 1st run. 1xTBE buffer, 35 cm total length, 30 cm effective length. Sample injection: 2 s, 5 s, and 5 s at the running voltage, respectively.

Figure 4. Electropherogram of human amelogenin system in 4.5% PVP. 200 V/cm, 55 cm total length, 50 cm effective length. Injection: 5 s at 200 V/cm.

Figure 5. Separation of CTTv ladder in 5.5% PVP under different conditions: (a) 1xTBE buffer with 3.5 M urea, room temperature, 200 V/cm; (b) 1xTBE buffer with 3.5 M urea, column at 83 °C, tip in 95 °C 1xTBE buffer, 300 V/cm; and (c) 1xTBE buffer with 7 M urea, column at 65 °C, tip in 95 °C 1xTBE buffer, 200 V/cm. Column: 55 cm total length, 50 cm effective length. Hot injection at 200 V/cm for 20 s after denaturing at 95 °C for 3 min. For the TH01 region, the primary peaks are marked with asterisks.

Figure 6. Electropherogram of the primer pair for TH01 in 5.5% PVP: (a) commercial primer pair; (b) custom synthesized primer pair; and (c) commercial primer pair mixed with custom synthesized primer pair. The primer pairs from the two sources have the same structures but only one strand was labeled with dye in each case. Conditions: 1xTBE buffer with 3.5 M urea, room temperature, hot injection for 2 s at 200

V/cm. DNA concentration: 0.1 μM. Denaturation: 95 °C for 3 min. The correct primer sequence gives rise to the peaks labeled with asterisks.

Figure 7. Electropherogram of PCR products of TH01 ladder and TH01 9.3 allele in 5.5% PVP. Conditions: Same as in Figure 6a.

Figure 8. Continuous electropherogram of M13mp18 Sanger fragments in 5% extracted PVP. Conditions: 1xTBE buffer with 3.5 M urea, room temperature, 150 V/cm. Column: 55 cm total length, 50 cm effective length. Hot injection for 15 s at 150 V/cm after denaturation at 95 °C for 3 min.

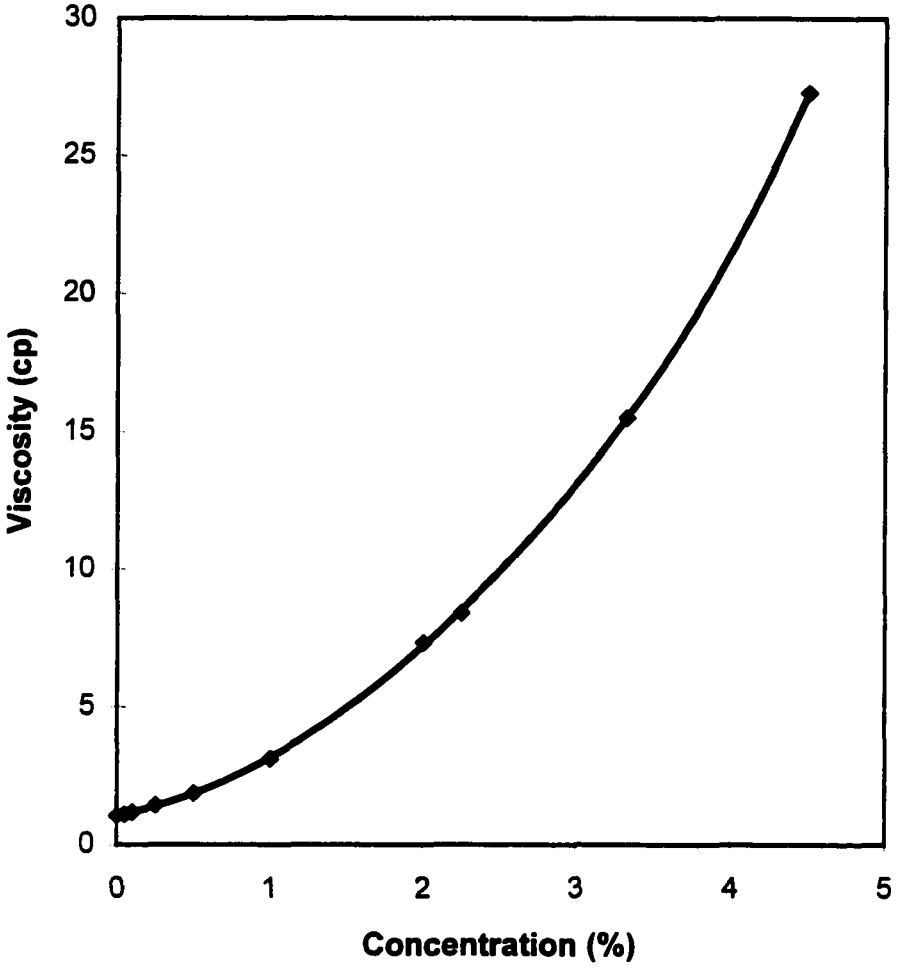


Figure 1

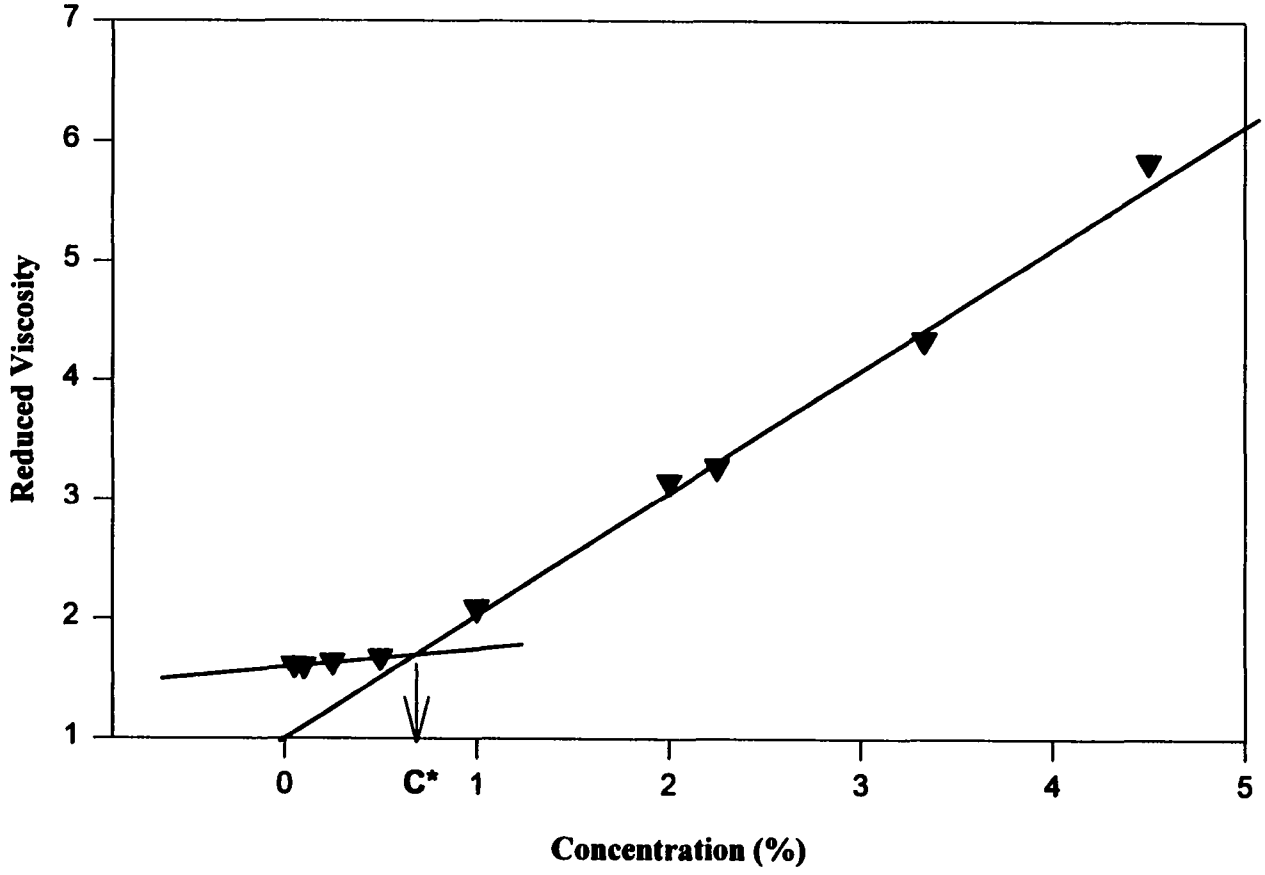


Figure 2

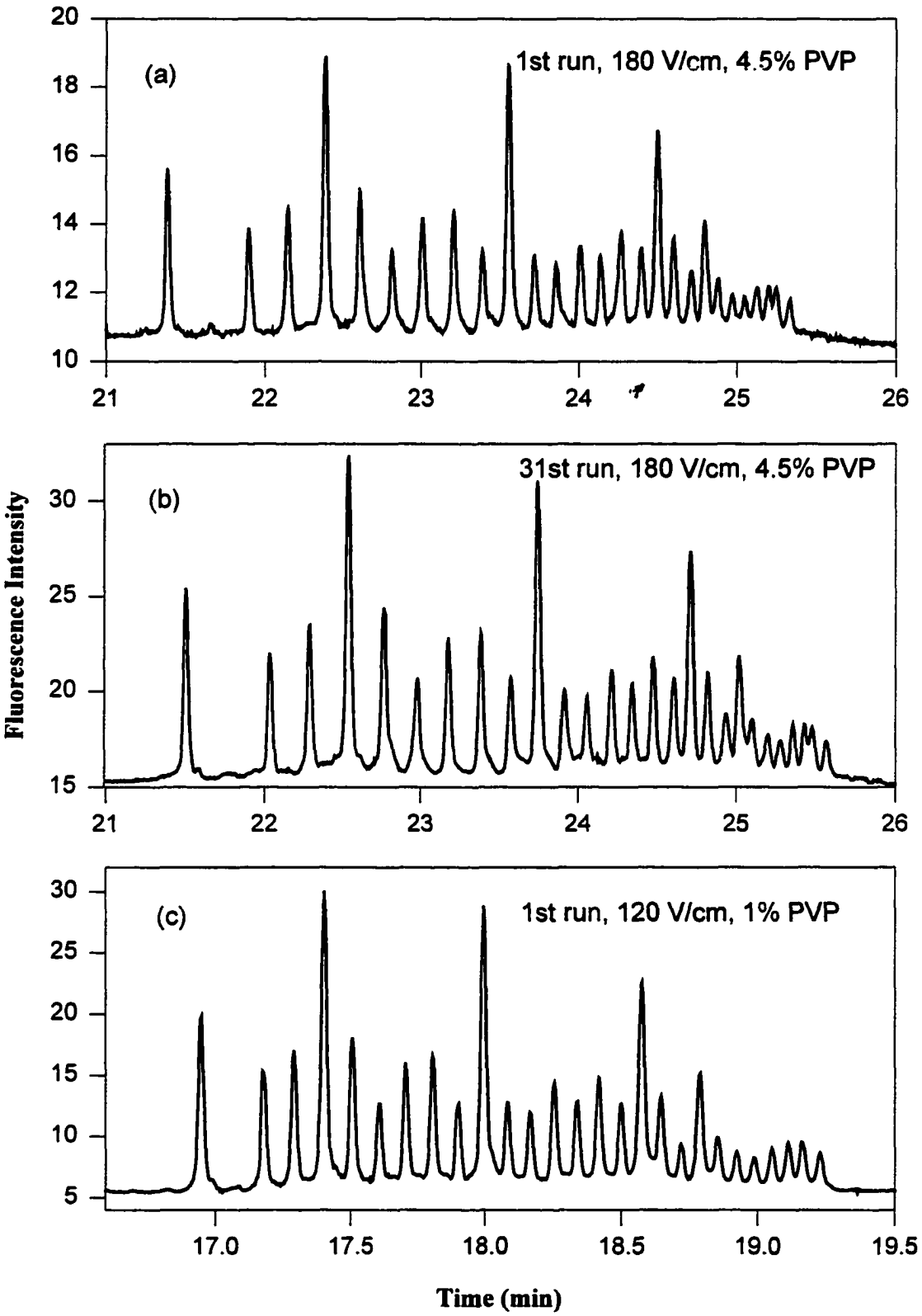

(a)
1st run, 180 V/cm, 4.5% PVP
(b)
31st run, 180 V/cm, 4.5% PVP
(c)
1st run, 120 V/cm, 1% PVP
Fluorescence Intensity
Time (min)

Figure 3

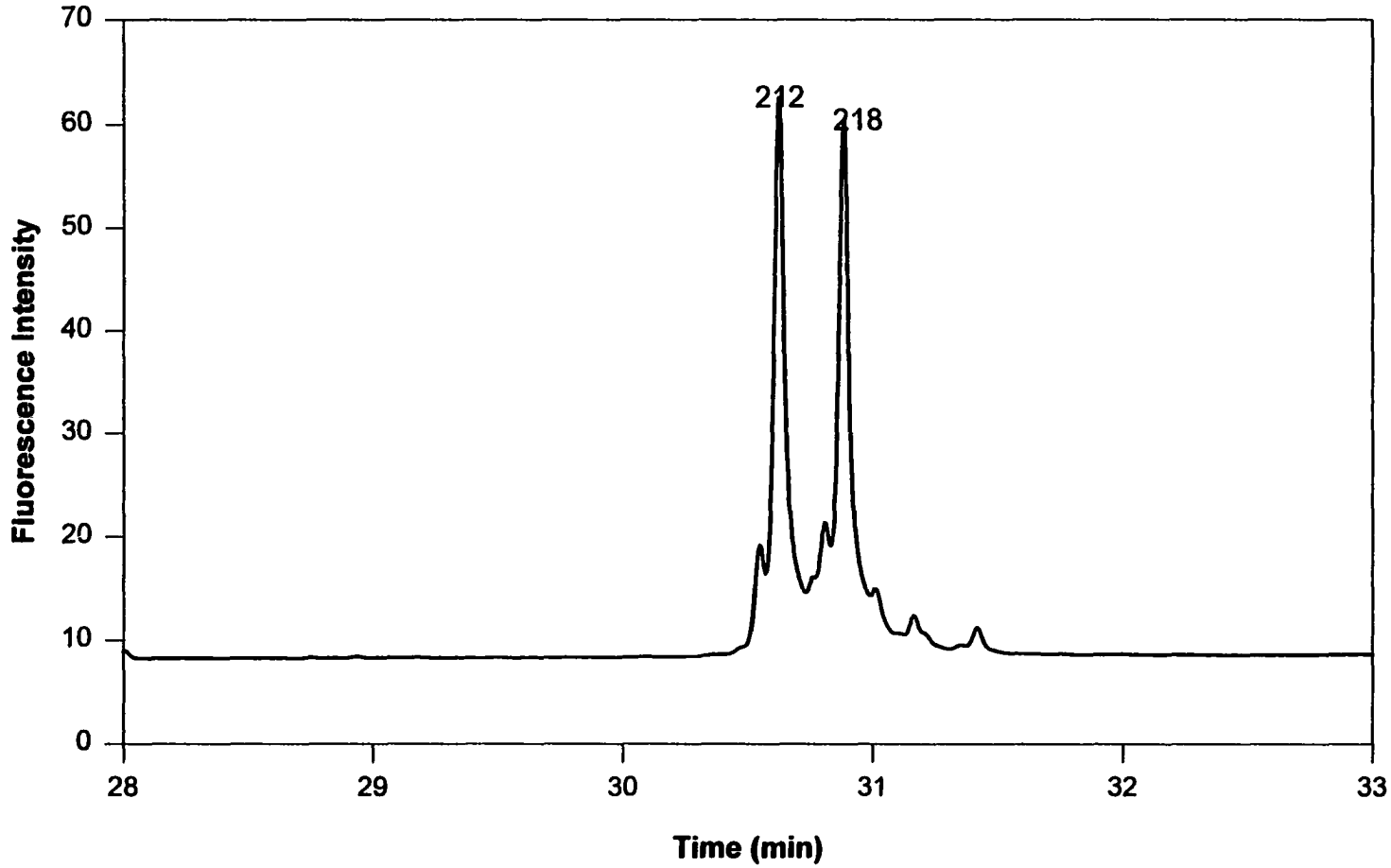

70
60
50
40
30
20
10
0
Fluorescence Intensity
212
218
28
29
30
31
32
33
Time (min)

Figure 4

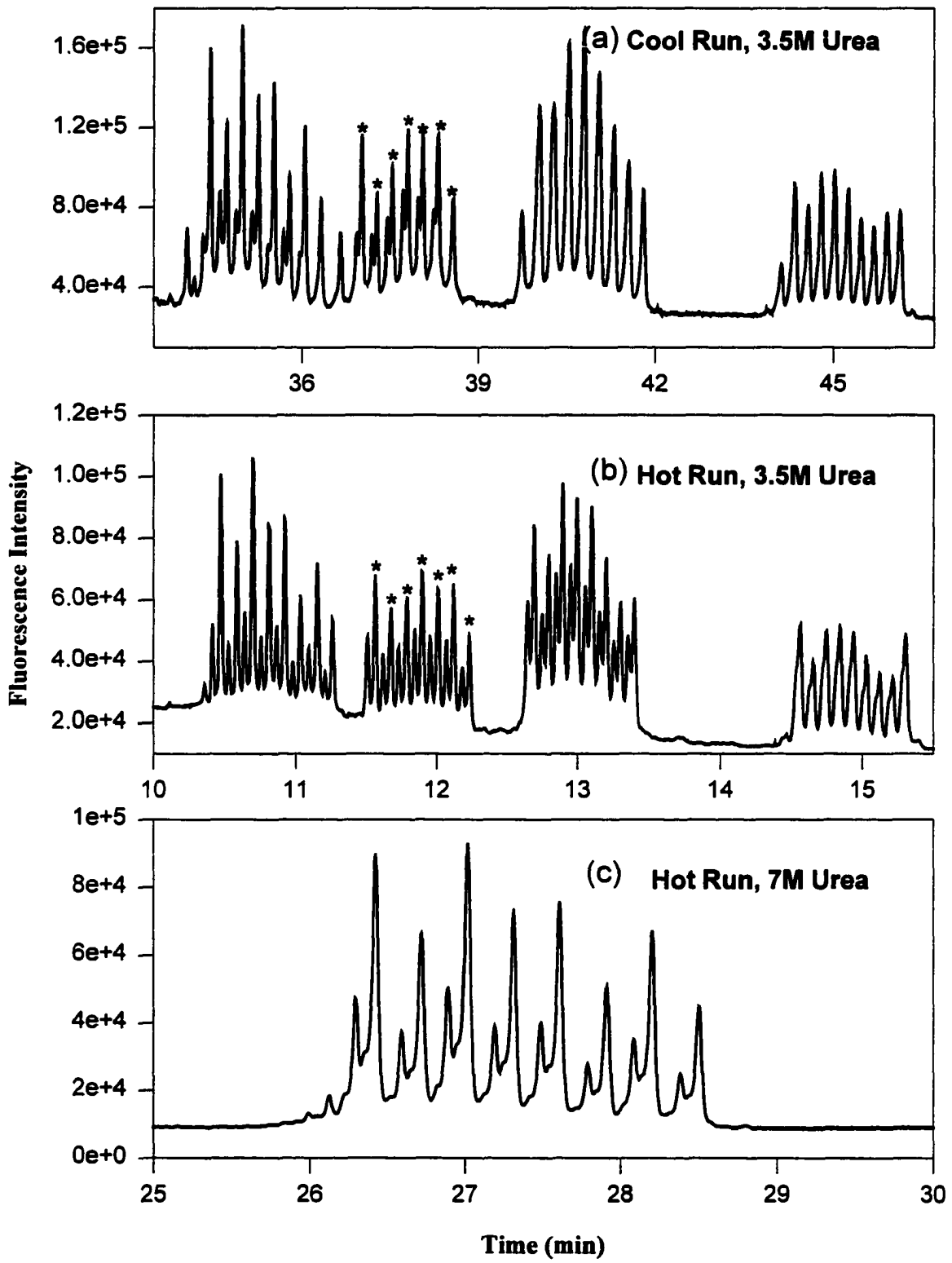

(a) Cool Run, 3.5M Urea
(b) Hot Run, 3.5M Urea
(c) Hot Run, 7M Urea
Fluorescence Intensity
Time (min)
1.6e+5
1.2e+5
8.0e+4
4.0e+4
36
39
42
45
1.2e+5
1.0e+5
8.0e+4
6.0e+4
4.0e+4
2.0e+4
10
11
12
13
14
15
1e+5
8e+4
6e+4
4e+4
2e+4
0e+0
25
26
27
28
29
30

Figure 5

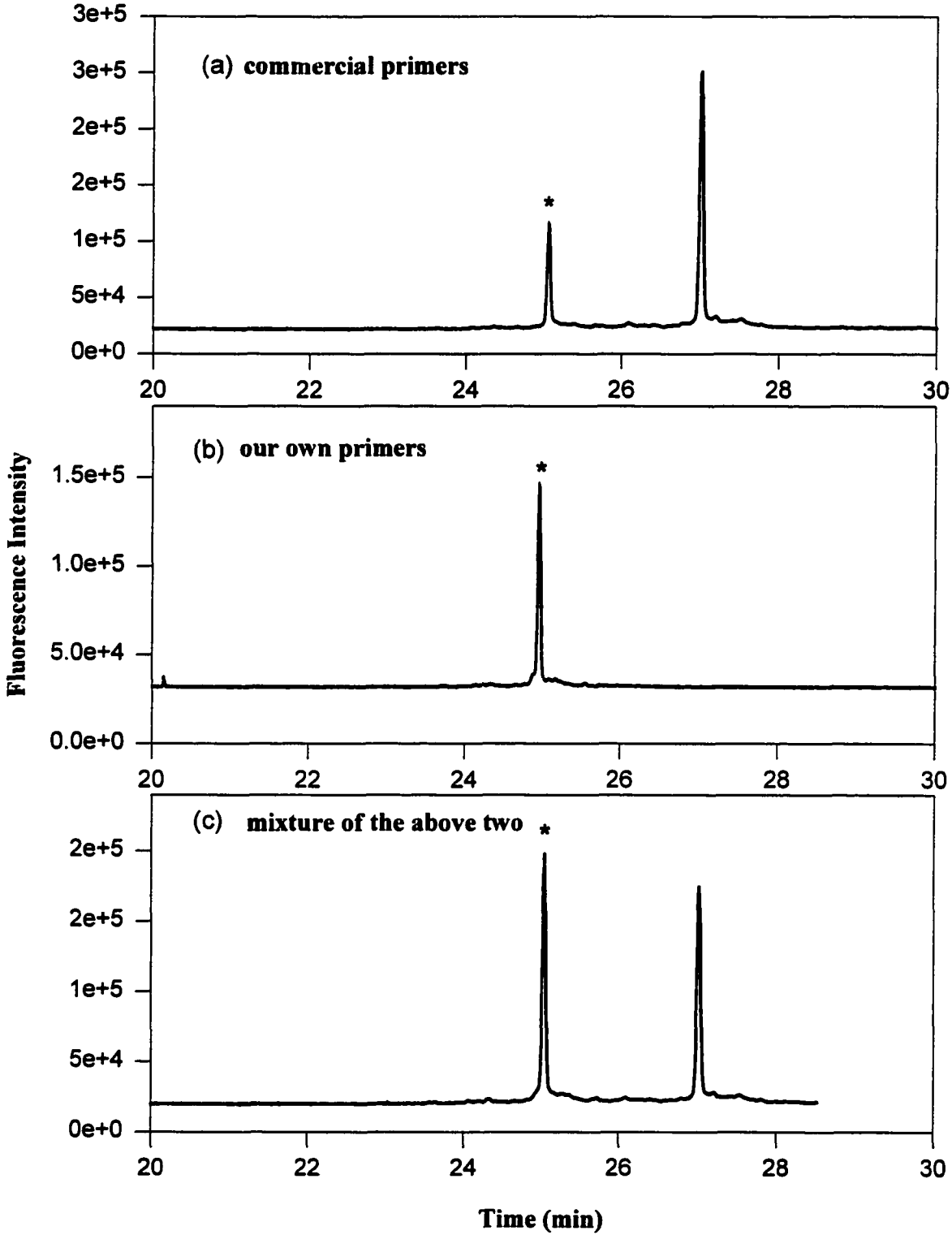

(a) commercial primers
(b) our own primers
(c) mixture of the above two
*
Fluorescence Intensity
Time (min)
3e+5
3e+5
2e+5
2e+5
1e+5
5e+4
0e+0
1.5e+5
1.0e+5
5.0e+4
0.0e+0
2e+5
2e+5
1e+5
5e+4
0e+0
20
22
24
26
28
30

Figure 6

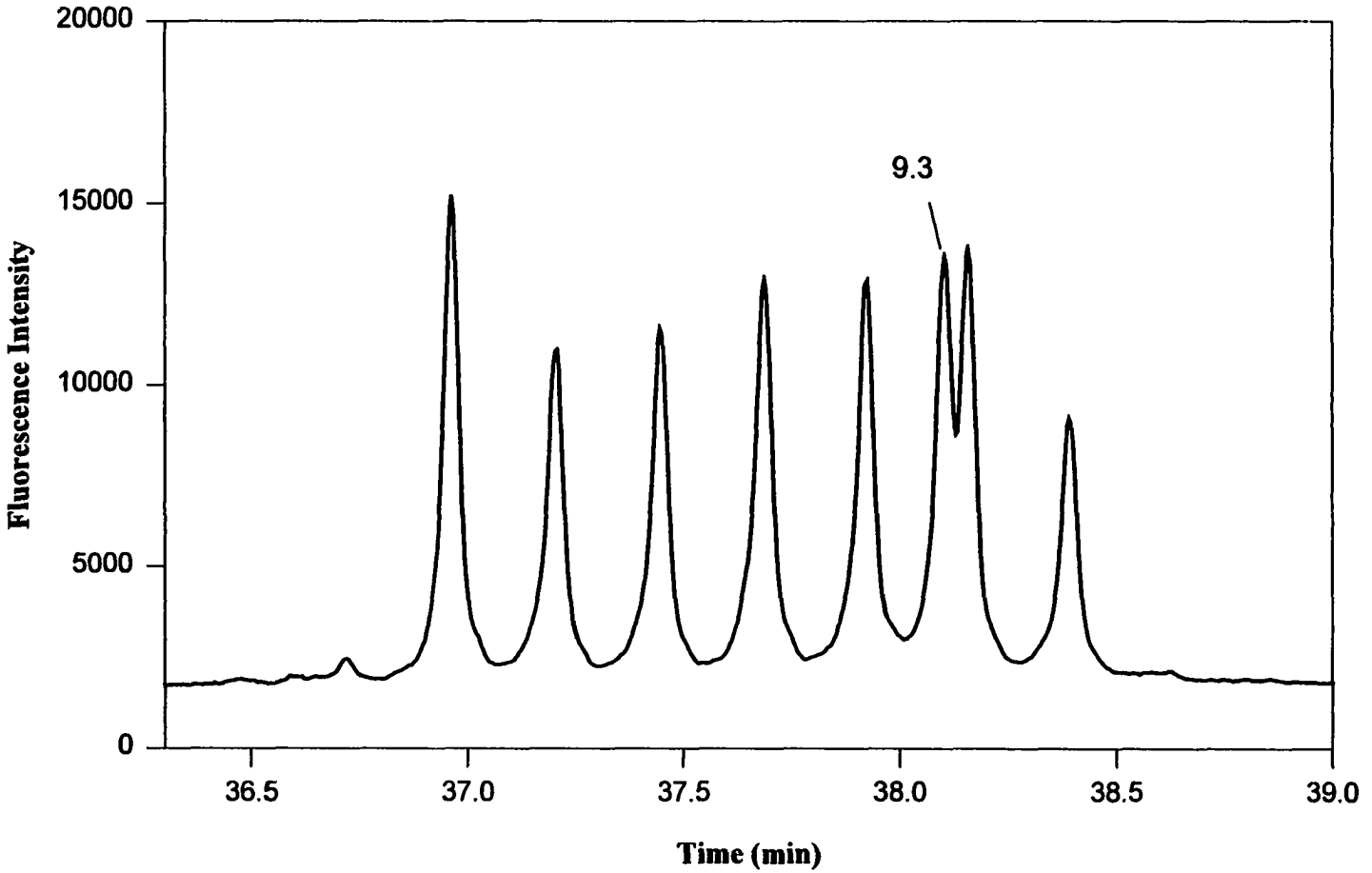

20000
15000
10000
5000
0
Fluorescence Intensity
9.3
36.5
37.0
37.5
38.0
38.5
39.0
Time (min)

Figure 7

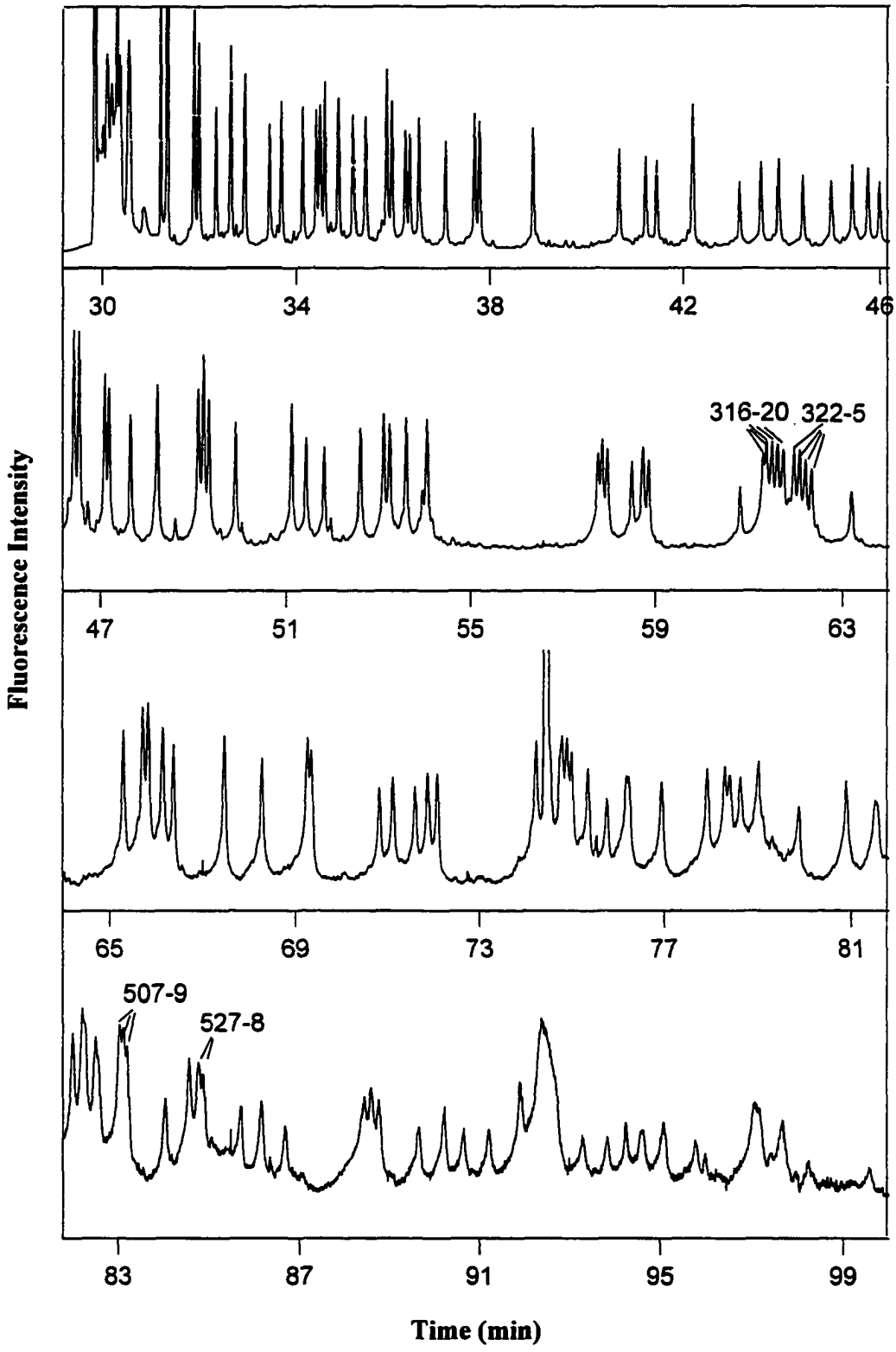

Fluorescence Intensity
30
34
38
42
46
316-20
322-5
47
51
55
59
63
65
69
73
77
81
507-9
527-8
83
87
91
95
99
Time (min)

Figure 8

CHAPTER 4. HIGH-THROUGHPUT MUTATION DETECTION BY TEMPERATURE GRADIENT CAPILLARY ELECTROPHORESIS WITH POLYVINYLPYRROLIDONE SOLUTION

A paper prepared for submission to Analytical Chemistry

Qiufeng Gao and Edward S. Yeung

Abstract

Single point mutation detection has called upon much attention recently. Though many methods have been reported, low-cost, high-throughput and high-detection-rate methods are still in demand. We present a fast and reliable mutation detection scheme based on temperature gradient capillary electrophoresis. Temperature gradient was controlled externally. A large temperature gradient (10 °C) was applied. A precision of 0.02 °C was set while a temperature ramp of 0.7 °C/min was performed. Multiple unlabeled samples from PCR reaction were injected and analyzed. Ethidium bromide was used as intercalating dye for laser induced fluorescence detection. The mutations were identified by comparing the electropherogram pattern of a herteroduplex with that of a homeduplex reference in continuous runs. All five mutations were successfully detected with high confidence. This scheme should be suitable for multiple capillary array electrophoresis and for screening single point mutations among large batch of samples.

Introduction

Mutation detection has become increasingly important in the fields of genetics, molecular diagnostics and cancer research.[1] Typically single nucleotide polymorphisms (SNP) are drawing more attention because they are the most common form of genetic variations. This type of single-base substitution in the genome occurs at a frequency of >1% in the human population. A recent estimate is that there is about one SNP per 1000 bp.[2] Other types of SNP involve insertion and deletion, and are found to occur at about one per 12 kb. SNP can be used to study genetic linkages[3] and for the diagnosis of diseases, especially cancer.[4]

The ultimate way to fully characterize a mutation is to perform DNA sequencing to the corresponding sample. However, the current DNA sequencing techniques are still laborious and expensive. Large scale DNA sequencing to detect mutation is not practical at this stage because a large portion of sequencing will give negative results considering the mutation is the minority in most cases. To save time and cost, screening methods have been developed to identify both known point mutations and unknown point mutations before any further step is taken.

The detection of mutation can be accomplished by using oligonucleotide arrays[5] or DNA chips.[6,7] Even though the number of spots that can be packed into a small area is very large,[8] one has to use multiple spots to span a mismatch. The match/mismatch discrimination is not entirely definitive, since different sequences have different melting temperatures. Ideally, one would have slightly different temperatures at each probe spot. The other issue is time. In a representative mode of operation,[2] the DNA is applied to the chip and hybridization is carried out at 44 °C for 15 hours at 40 rpm. The chips are then washed and

stained before imaging. A third issue is that the DNA chips are presently quite costly if one wants to span all possible mutations and probe scores of clinical samples at a time. Clearly further development is needed to speed up the process and to make it more cost effective.

Mass spectroscopy has been suggested to sequence DNA[9] or at least to detect alterations in DNA.[10] Deletions and insertions are readily detected by mass changes. Substitutions are not so obvious because of the limited mass resolution of instruments that are reasonably accessible at present. Positional switches will not be detected at all. Recent success in detecting ions from large DNA fragments is encouraging.[11] Given time and more research, mass spectroscopy may well become a suitable tool for detecting polymorphism in the genome.

A popular electrophoresis method to detect polymorphism is to rely on slight changes in conformations in single-stranded DNA (SSCP).[12] This is performed under mildly denaturing conditions. SSCP has been applied to detect *p53* mutations[13] and has been adapted for capillary electrophoresis.[14] There are three important considerations for SSCP analysis. First, the "mildly" denaturing condition is not well defined and may have to be optimized for each DNA region. This is because of conformation, and therefore changes in conformation, is sequence specific. Second, visualization after separation is not routine. The introduction of a radionucleotide probe or a fluorescence label into the DNA strand requires knowledge of specific sequences of DNA regions around the point of mutation. Third, at present the assay is not reliable with fragments greater than around 200bp and the sensitivity is only 60-95%.[15,16]

For the analysis of double-stranded DNA, conformation sensitive gel electrophoresis (CSGE) has been demonstrated.[17] This is based on slight differences in conformations

between the homoduplex and the heteroduplex DNA fragment. Just like SSCP above, the gel and buffer conditions are not universal. When applied together with SSCP, the mutation detection rate may reach near 100%. [18]

A different approach is to use denaturing gradient gel electrophoresis (DGGE).[19] Separation is performed at a constant temperature but with a gel constructed to provide various degree of denaturation along its length. The gradient gel in principle can resolve all mismatches, but irreproducibility in creating identical gels makes implementation difficult. It is desirable to attach an artificial GC-rich sequence to the respective ends of the two strands to provide optimum separation. By using this single method, the mutation detection rate can reach near 100%.

Compared to SSCP/CSGE, DGGE can tolerate longer DNA fragments and is less time-consuming. An analogue of DGGE is temperature gradient gel electrophoresis (TGGE).[20] In TGGE, instead of denaturant gradient along the gel, a spatial or temporal temperature gradient is used to perform the same function. A simpler scheme is to apply constant denaturing capillary electrophoresis (CDCE).[21] But this is limited to defined mutations.

A totally different approach that has been found to be quite successful is denaturing high performance liquid chromatography (DHPLC).[21] The specific interaction involves is ion-pair reverse-phase retention. DHPLC has been applied to study human evolution via Y-chromosome mismatches,[22] to quantify RT-PCR by introducing mutant RNA as an internal standard,[23] and to detect germline and somatic mutations.[24] The present limitations of DHPLC are the need to establish specific temperatures for distinct types of mismatches and inability to run multiple samples at the same time.

Capillary electrophoresis comprises fast analysis, small amount of sample requirement, and high sensitivity. It has been successfully used in many DNA analysis fields like sequencing and genotyping. Recently developed multiple capillary array technique provided further application perspective for high-throughput analysis. Therefore, though DGGE was developed and used in slab gel format, its transplant to capillary electrophoresis has drawn much interest. Karger et al reported p53 mutation detection by CDCE with laser induced fluorescence.[25] Gelfi and Righitti first reported DGCE.[26] In addition to the denaturant gradient, they also applied a secondary polymer concentration gradient to refocus the sample band to recover sensitivity. While their result is quite impressive, the construction of the gradients is tedious and hard to be reproducible, especially for a capillary array. The same authors also developed TGCE,[27] where the temperature sweeping was generated internally via ohmic heat produced by voltage ramp over time. A temperature profile vs. voltage must to be established prior to the experiment. The main challenges for this technique are the narrow temperature sweeping range and dependent temperature control on the electric field.

In this paper, we present a different construction of mutation detection by temperature gradient capillary electrophoresis. The temperature was controlled externally by a heating plate. Temperature gradient was programmed through a computer with a precision of 0.02 °C per step. An intercalating dye was used for laser induced fluorescence detection. Polyvinylpyrrolidone, which was successfully used in sequencing and genotyping, was used as sieving matrix. Nine Single-point mutation samples were tested. Fast and accurate identification was achieved with this system

Experimental Section

Chemical Reagents

1x TBE buffer was prepared by dissolving 8.5 g premixed TBE buffer powder (Amerosco, Solon, OH) into 500 ml dionized water. Ethidium bromide was incorporated into the TBE buffer with a final concentration of 0.5 μg/ml. Polyvinylpyrrolidone (PVP) was obtained from Sigma (St. Louis, MO). The sieving matrix was made by dissolving 3% (w/v) of 360,000 M_w PVP into 1x TBE buffer with dye. The mixture was shaken for 2 min and left standing for 1 h to remove bubbles.

Table 1. Characteristics of the DNA samples

Sample Name	Length (bp)	Mutation Type	Mutation Position	Primer Sequence
M_2	209	A to G	169	For =AGGCACTGGTCAGAATGAAG Rev = AATGGAAAATACAGCTCCCC
M_{60}	388	T insertion	243	For = GCACTGGCGTTCATCATCT Rev = ATGTTCATTATGGTTCAGGAGG
M_{69}	256	T to C	221	For=GGTTATCATAGCCCACTATACTTTG Rev= ATCTTTATTCCCTTTGTCTTGCT
M_{122} (a)	393	T to C	73	For= TGGTAAACTCTACTTAGTTGCCTTT Rev= CAGCGAATTAGATTTTCTTGC
M_{122} (b)	393	T to C	289	For= TGGTAAACTCTACTTAGTTGCCTTT Rev= CAGCGAATTAGATTTTCTTGC

DNA Sample

The point mutation samples and their references were kindly provided by Dr. Peter Oefner and Dr. Peidong Shen (Stanford University). The symbols and characteristics of the samples are listed in table 1. The heteroduplex was formed by mixing, denatuing and then reannealing the reference sample and mutant at an amount ratio of 1:1. The meltmap of the samples were constructed with the facilitation of a computer program *Melt94*.[28]

Experimental Setup

A laboratory-assembled CE-system with laser-induced fluorescence (LIF) detection and a heating plate was used in this work. The heating plate was an aluminum plate with dimension of 48cm x 10cm x 0.8 cm. A 300 w heating element and a platinum sensor were closely attached underneath. A 84-cm capillary with 375 μm o.d. and 75 μm i.d. (PolymicroTechnologies, Phoenix, AZ) was attached to the heating plate along the 48-cm direction. There was a 18-cm part in front of and a 8-cm part after the heating plate. The capillary was immersed into a layer of high thermal conductivity paste on the heating plate for better temperature control (Fig. 1) The temperature was regulated and monitored by an Eppendorf TC-50 temperature controller (Brinkmann Instruments, Westbury, NY). The temperature controller was modified so that it was able to be programmed by computer. The temperature control program was written in LabView language. A temperature setting from 58 °C to 78 °C with a precision of 0.02 °C per step was allowed. In deed, since the capillary has a small volume and an excellent contact with the heating plate due to the existence of thermal conductivity paste, the temperature in side the capillary will follow the temperature reading of the plate very well at the ramp of experimental condition.

The design of the CE instrument was similar to that reported in previous papers[29]; briefly, a 514 nm Ar^+ laser was used for excitation. The applied laser power was 5 mw. A PMT with 590nm long-pass filter was used for collecting the fluorescence. The sampling rate was 5 Hz.

Experimental Procedure

The capillary was filled with 3% PVP solution without any pretreatment. 10 s at 220 V/cm field strength was used for sample injection. The temperature was preset to 61 °C before the injection. The temperature program was as following: 4 minutes gradient delay after injection, 0.7 °C/min ramp at a precision of 0.02 °C per step. The gradient stopped at 71 °C. Then the heating plate was set back to 61 °C automatically. After each run, fresh PVP matrix was used to simply flush out the old one to maintain the separation efficiency in the subsequent run. During idle periods, the capillary array was stored in deionized water.

Results and Discussion

Temperature Programming

The application of temperature gradient is to partially denature the DNA. Because the heteroduplex has a mismatch, it will start to melt at a lower temperature than the homoduplex. A large change in electrophoretic mobility on melting can be observed in a sieving medium, such as a gel or a long chain linear polymer solution. The heteroduplex will exhibit a retarded migration behavior near melting temperature compared to the homoduplex under the same condition. The mutation can be identified by the electrophoretic pattern difference between homoduplex and heteroduplex at appropriate temperature. For accurate

comparison of the patterns, a reproducible temperature profile is required in the continuous runs. By using the heating plate, accurate and reproducible temperature control can be obtained. The temperature programming is independent on the electric field so that arbitrary temperature gradient can be selected. Excellent migration time and electropherogram pattern reproducibility was observed from run-to-run with a test heteroduplex sample. The relative standard deviation of migration time was within 2%.

Since the mobility retardation occurs only when the DNA fragments begin to melt, the part of the capillary that is not on the plate will not affect the mobility of the fragments. We can simply take out the time that the fragments spend in the unheated parts of the capillary and start the temperature gradient once the DNA fragments migrate into the heated part. The mobility of DNA fragments in the unheated part can be calculated from running the same sample under room temperature. Electric field correction should be carried out because the voltage drop along the capillary is not even any more when a part of capillary is heated. A way to do the electric field correction is through the current calculation. Since the property of the unheated part is independent of heating of the other part, the current through it should be proportional to the electric field added to it. After adjusting the current to the value under heated condition, the DNA mobility obtained at room temperature should be the mobility in the unheated part when the heating plate is working. Fig 2 is the electropherogram of the mixture of M_2 and M_{122}. These are the smallest and largest DNA fragments among the samples. The experimental conditions are listed in the figure caption. It can be calculated that it takes 4 min 30s and 4 min 50s for the M_2 and M_{122} respectively to migrate into the heated region after injection. The migration time for the rest samples to get into the heated region will fall in between. Considering the temperature ramp is 0.7 °C/ min, the temperature

gradient will start from 61.2 °C for the M_{122} sample. But compared to the 10 °C gradient range, the increase of starting temperature is negligible.

Detection Scheme

The traditional way to indicate the DNA bands in gel electrophoresis is to apply silver stain or ethidium bromide stain after electrophoresis. Obviously, these schemes can not be transferred to capillary electrophoresis directly. In capillary-based mutation detection, UV absorbance[27] and laser induced fluorescence detection[25] are the most popular detection methods. In laser induced fluorescence detection mode, the dye is usually covalently bonded to the primer. UV absorbance detection is universal and all the fragments will be detected. One drawback is that the short light path of capillary will limit the sensitivity and a high concentration of DNA is needed. Laser induced fluorescence is sensitive, but dye-labeled primer is costly. In addition, the single strands from melting and PCR reaction will interfere with the identification.[25]

Ethidium bromide has been used for detection of ds-DNA in capillary electrophoresis at room temperature. By incorporating ethidium bromide into buffer and sieving matrix, raw PCR products can be detected. The binding of the dye and DNA is fast though maybe not complete. Fig. 3 is the comparison of signal level with and without heating. Under the heated condition, the signal dropped to 50-70% of that at room temperature. Partial melting of the double-stranded DNA and possibly weaker binding between DNA and dye at high temperature are the reason for the signal drop. Though the unheated part of the capillary after the heating plate may help the recovery of the signal in some degree, the renaturing of DNA and dye binding process is not efficient enough to resume to resume the signal completely.

Even though, this detection method is still more sensitive than the UV absorbance method because no detectable signal was observed in UV absorbance with the same sample under the same electrophoresis condition. An additional advantage is that ethidium bromide has much stronger binding to ds-DNA than ss-DNA. Only the double stranded form can be detected. So the electropherogram pattern is very clean.

Meltmaps of the Samples

The selection of the temperature range was arbitrary and was to cover as many mutations as possible. For most mutations, the sequence of the target DNA fragment is already known. The position and type of the mutation can be either known or unknown. A meltmap of the target DNA fragment can facilitate the determination of the temperature gradient range. Fig. 4 shows the meltmaps of the five samples. The meltmap shows the equilibrium (50%) temperature, T_m, along the sequence at which each base pair in the molecule has equal probability for the helical or melted states as long as the strands remain associated. The low melt domain of all the samples fall into the temperature range of 60 °C ~ 70 °C and the mutation positions are all in the low melt domain. So A temperature range of 10 °C between 61 °C and 71 °C should be able to cover all the samples.

For an unknown mutation sample, if the mutation occurs at the high melt domain, it will be difficult to detect because the temperature will fully denature the double strands. In that case, a GC clamp added to the primer may add an artificial high melt domain to the meltmap and help to strengthen the binding between the two strands. A broader temperature range can be applied or two different temperature gradient runs can be performed.

Resolution Characteristics

Figure 5 is the electropherograms of the homoduplex samples and the heteroduplex samples. Fast separations were achieved. These samples were tested under the same temperature gradient. Though the samples had different fragment lengths and different mutations, they all showed recognizable pattern differences between homoduplex and heteroduplex. This temperature gradient is not the best for all the samples. Better resolution for the heteroduplex were achieved in some samples by using a different and narrower temperature range and slower ramp (Fig 6 and 7). However, the purpose of the experiment is to identify the mutations, not characterize them. Even a slight pattern change may lead to the recognition of a mutation since this is a highly reproducible system. Perfect separation of the fragments in heteroduplex is not necessary.

The pattern change can be determined by the existence of additional peaks (at most 4 total), peak shoulders, or even broader peak width. The first change is quite obvious and can lead to the identification with high enough confidence. The latter two are less confident and may cause false positive. But a false positive is less detrimental than a false negative in clinical diagnosis. Moreover, the confidence level can be determined from the 2% relative STD and the level of the pattern change.

PVP solution has been proved to be able to separate DNA sequencing fragments and genotyping fragments. The 3% PVP used in this experiment was very dilute yet gave good separation performance. The viscosity of this solution is less than 10 cp. The excellent EOF suppressing effect[29] may contribute to the reproducible results obtained in this experiment.

Conclusion

We have demonstrated a reliable temperature gradient capillary electrophoresis system for high-throughput mutation detection. 1-hour total turnaround time was required for a pair of homoduplex and heteroduplex samples. Large temperature range enables the coverage of many mutations so that a universal experimental condition can be applied to different samples. The use of polyvinylpyrrolidone makes the capillary regeneration very easy and result reproducible because of its low viscosity and excellent EOF suppressing effect. Sensitive detection was accomplished by incorporation of ethidium bromide into the buffer. This scheme should be suitable for capillary array mutation detection.

Acknowledgements

The Ames Laboratory is operated for the US Department of Energy by Iowa State University under contract No.W7405-Eng-82. This work was supported by the Director of Energy Research, Office of Basic Energy Science, Division of Chemical Sciences, and the Office of Health and Environmental Research.

References

1. Sidransky, D. *Science* **1997**, 278, 1054-1058.
2. Wang, D. G., et al. *Science* **1998**, 280, 1077-1082.
3. Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J., Cavalli-Sforza, L. L., *Proc.Natl. Acad. Sci (USA)* **1996**, 93, 196-200.
4. Sidransky, D., et al. *Science* **1991**, 252, 706-709.
5. Southern, E. M., *Electrophoresis* **1995**, 16, 1539-1542.

6. Chee, M., et al. *Science* **1996**, 274, 610-614.

7. Winzeler, E. A., et al. *Science* **1998**, 281, 1194-1197.

8. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., Solas, D. *Science* **1991**, 251, 767-773.

9. Taranenko, N. I., Allman, S. L., Golovlev, V. V., Taranenko, N. V., Isola, N.R., Chen, C. H. *Nucl. Acids Res.* **1998**, 26, 2488-2490.

10. Alper, J. *Science* **1998**, 279, 2044-2045.

11. Berkenkamp, S., Kirpekar, F., Hillenkamp, F. *Science* **1998**, 281, 260-262.

12. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T. *Proc. Natl. Acad. Sci. (USA)* **1989**, 86, 2766-2770.

13. Schneider-Stock, R., Haeckel, C., Radig, K., Pein, C.-D., Roessner, A. *Electrophoresis* **1997**, 18, 2849-2851.

14. Henbenbrock, K., Williams, P. M., Karger, B. L. *Electrophoresis* **1995**, 16, 1429-1436.

15. Eng, C., Vijg, *J. Nat. Biotechnol.* **1997**, 15, 422-426.

16. Gromp, M. *Nat. Genet.* **1993**, 5, 111-117.

17. Ganguly, A., Rock, M. J., Prockop, D. J. *Proc. Natl. Acad. Sci. (USA)* **1993**, 90, 10325-10329.

18. Cotton, R. G. H. *Mutation Detection*, P180, Oxford University Press, 1997.

19. Fischer, S. G., Lerman, L.S. *Proc. Natl. Acad. Sci. (USA)* **1983**, 80, 1579-1583.

20. Riesner, D., Steger, G., Zimmat, R., Owens, R. A., Wagonhofer, M., Hillen, W., Vollbach, S., Henco, K. *Elelctrophoresis* **1989**, 10,377-389.

21. Oefner, P. J., Underhill, P. A. *Am. J. Hum. Genet.* **1995**, 57, A266.

22. Underhill, P. A., Jin, L., Lin, A. A., Meddhi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., Oefner, P. *J. Genome Research* **1997**, 996-1005.

23. Hayward-lester, A., Oefner, P. J., Doris, P. A., *Biotechniques* **1996**, 20, 250-257.

24. Liu, W., Smith, D. I., Rechtzigel, K. J., Thibodeau, S. N., James, C. D. *Nucl. Acids Res.* **1998**, 26, 1396-1400.

25. Khrapko, K., hanekamp, J. S., Thilly, W. G., Belenkii, A., Foret, F., Karger, B. L. *Nucl. Acids Res.* **1994**, 22, 364-369.

26. Gelfi, C., Righetti, S. C., Zunino, F., Della Torre, G., Pierotti, M. A., Righetti, P. G. *Electrophoresis* **1997**, 18, 2921-2927.

27. Gelfi, C. Righetti, P. G., Cremonesi, L., Ferrari, M. *Electrophoresis* **1994**, 15, 1506-1511.

28. Lerman, L. S., Beldjord, C. *Mutation Detection: A Practical Approach*, P45-50, Oxford University press, 1998.

29. Gao, Q., Yeung, E. S. *Anal. Chem.* **1998**, 70, 1382-1390.

Figure Captions

Figure 1. Heating device structure.

Figure 2. Electropherogram of M_2 and M_{122} at room temperature. Conditions: 70 cm capillary effective length; 300 V/cm electric field; inject 10s; 3% PVP with 0.5 μg/ml ethidium bromide in buffer; current was the same as under heated condition.

Figure 3. Comparison of intensities from M_{60} under room temperature and under 61°C. Conditions: 84 cm total length; 48 cm on the heating plate; inject 10 s.

Figure 4. Meltmaps of the five samples.

Figure 5. Electropherograms of the homoduplex and heteroduplex samples. A pair of homoduplex and heteroduplex is potted together for comparison. (a), (b), (c), (d), and (e) refer to M_2, M_{60}, M_{69}, M_{122} (I) and M_{122} (II) respectively. Conditions: temperature gradient from 61 °C to 71 °C with a ramp of 0.7 °C/min; 200 V/cm electric field; inject 10 s.

Figure 6. Electropherograms of the homoduplex and heteroduplex of M_{60} under different conditions. (a) was obtained with a temperature gradient from 68-74 °C and a ramp of 0.5 °C /min. (b) was obtained with a temperature gradient from 61-68 °C and a ramp of 0.5 °C/min. The other conditions were the same as in figure 5.

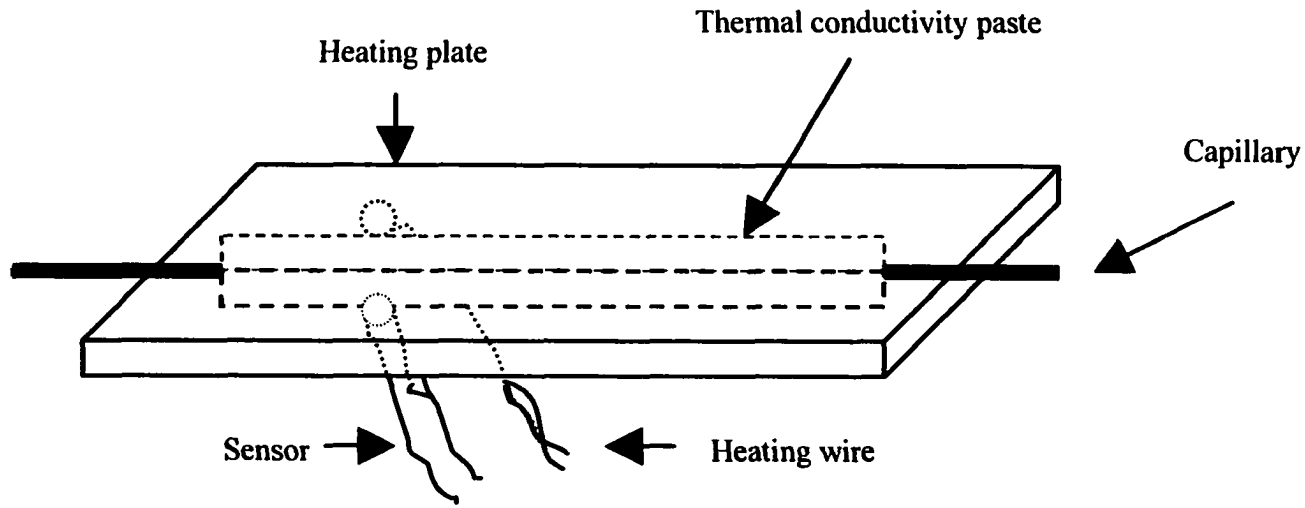

Heating plate
Thermal conductivity paste
Capillary
Sensor
Heating wire

Figure 1

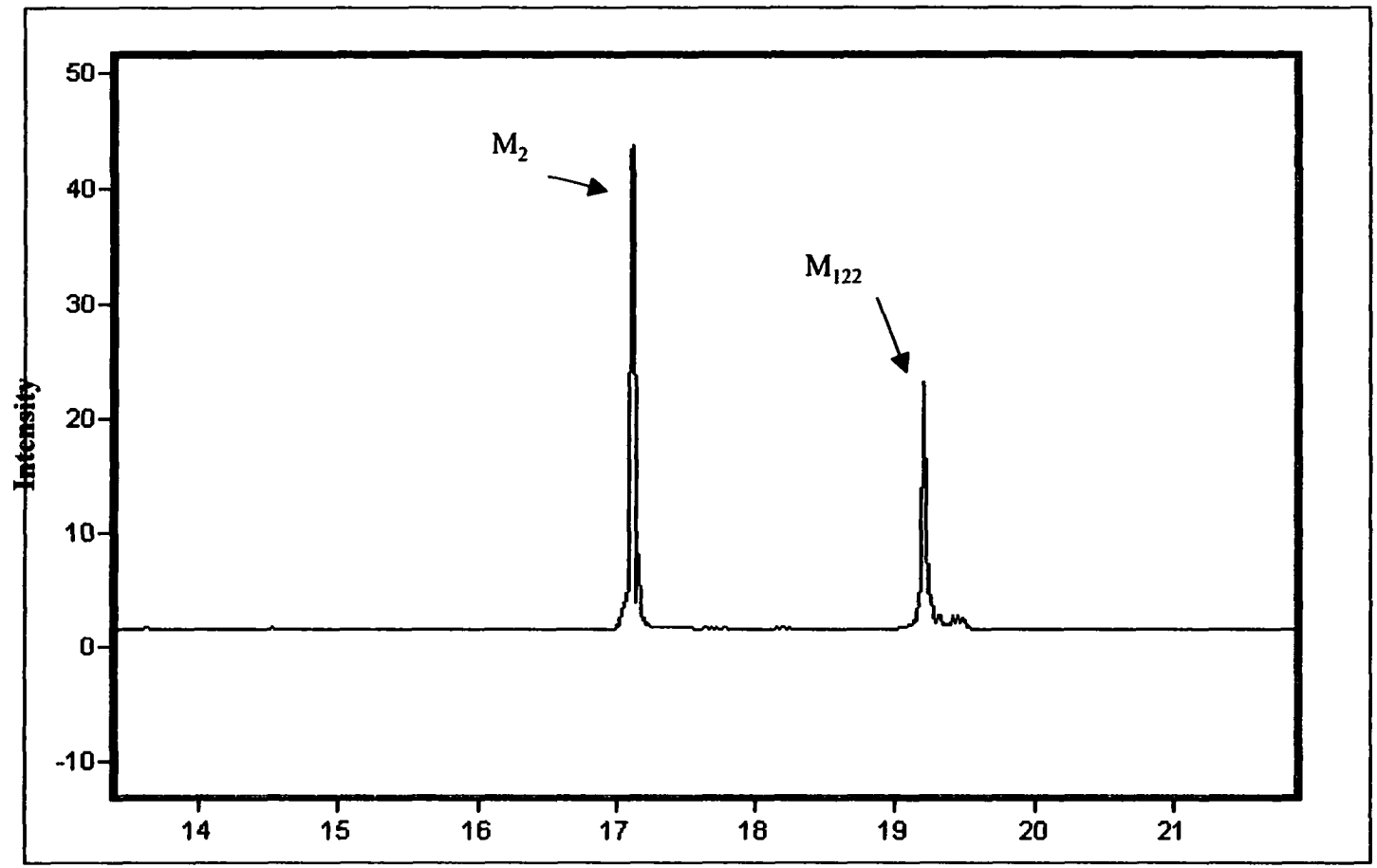

50
40
30
20
10
0
-10
Intensity
M2
M122
14
15
16
17
18
19
20
21
Time min)

Figure 2

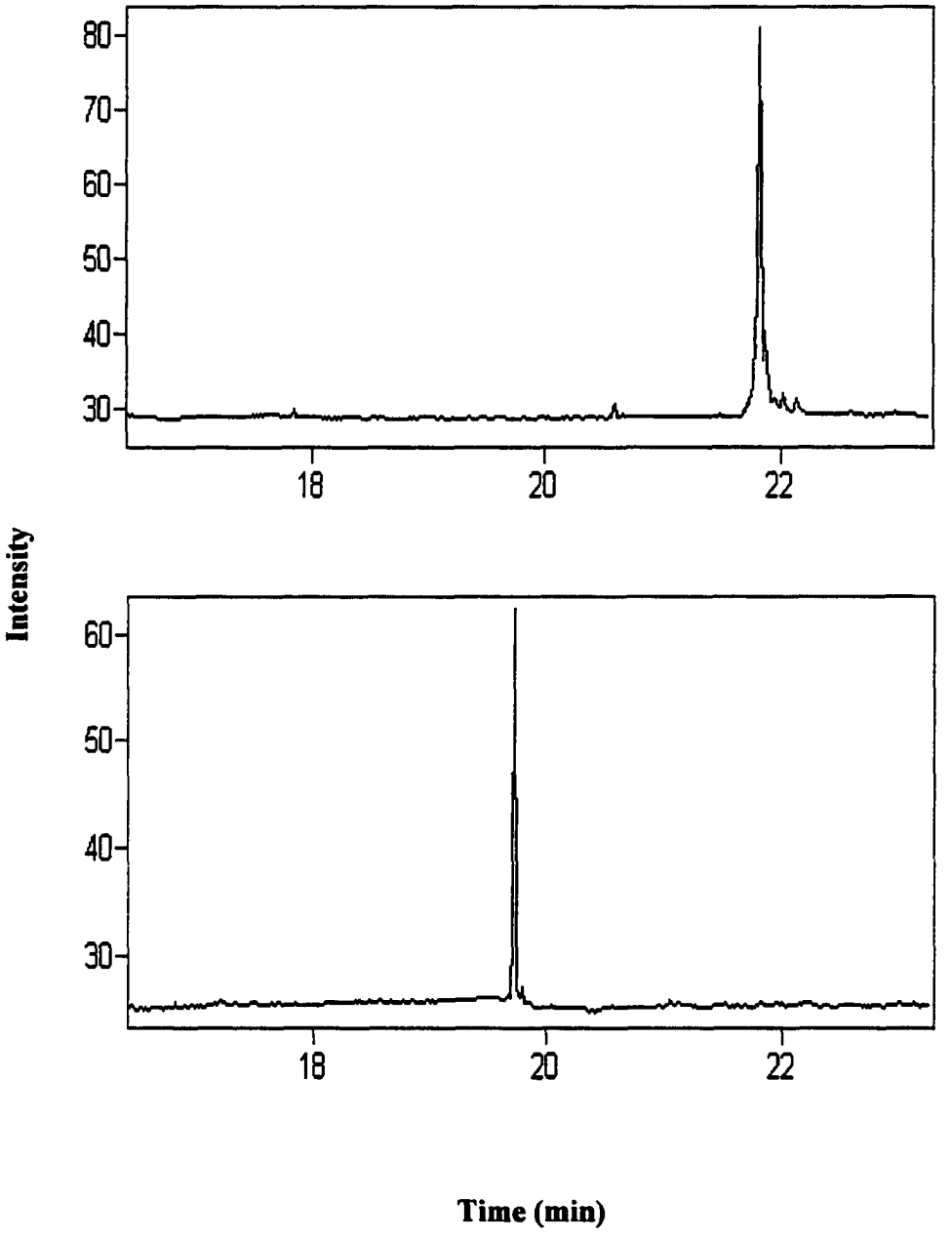


Figure 3

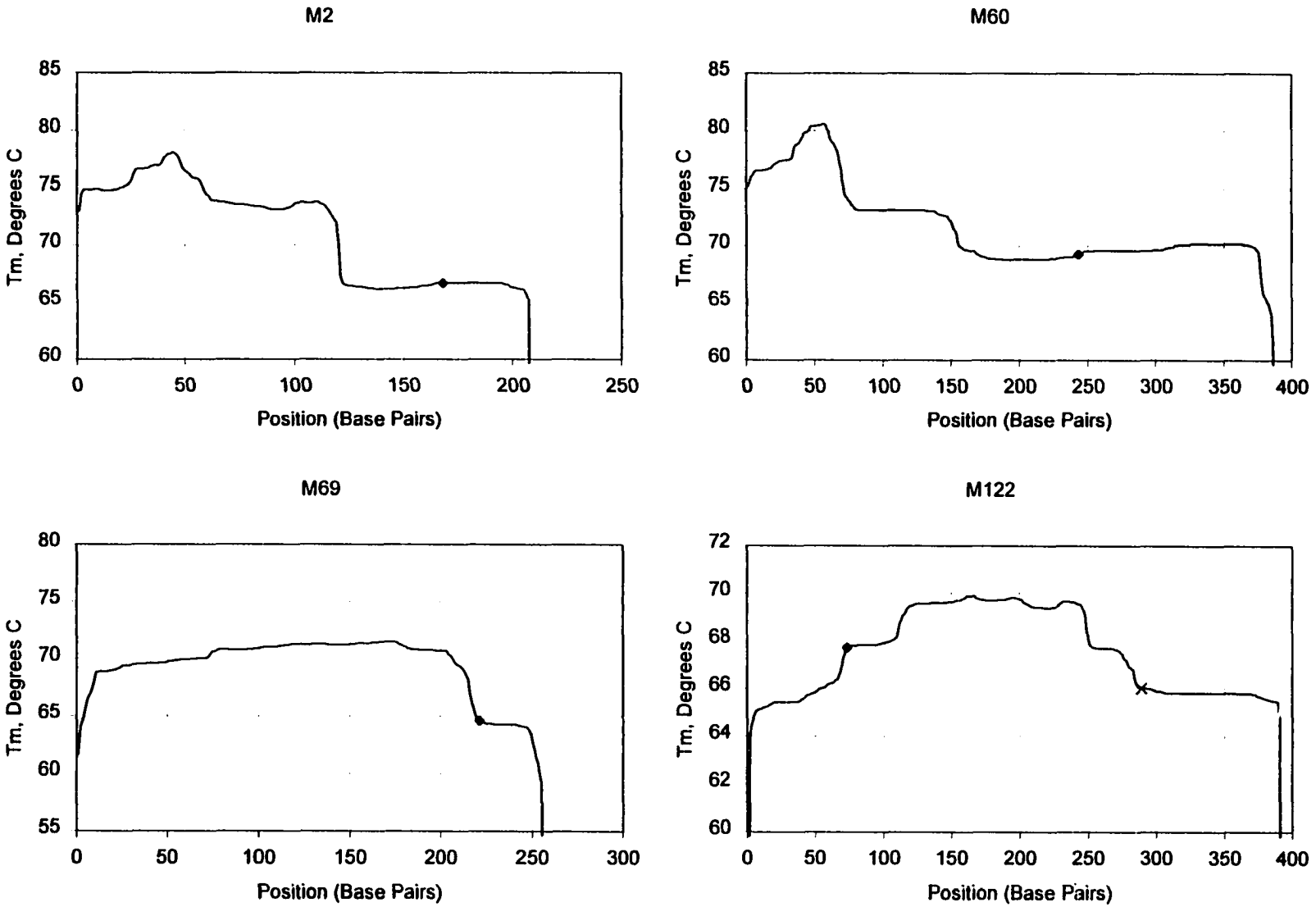

M2
Tm, Degrees C
85
80
75
70
65
60
0
50
100
150
200
250
Position (Base Pairs)
M60
Tm, Degrees C
85
80
75
70
65
60
0
50
100
150
200
250
300
350
400
Position (Base Pairs)
M69
Tm, Degrees C
80
75
70
65
60
55
0
50
100
150
200
250
300
Position (Base Pairs)
M122
Tm, Degrees C
72
70
68
66
64
62
60
0
50
100
150
200
250
300
350
400
Position (Base Pairs)

Figure 4

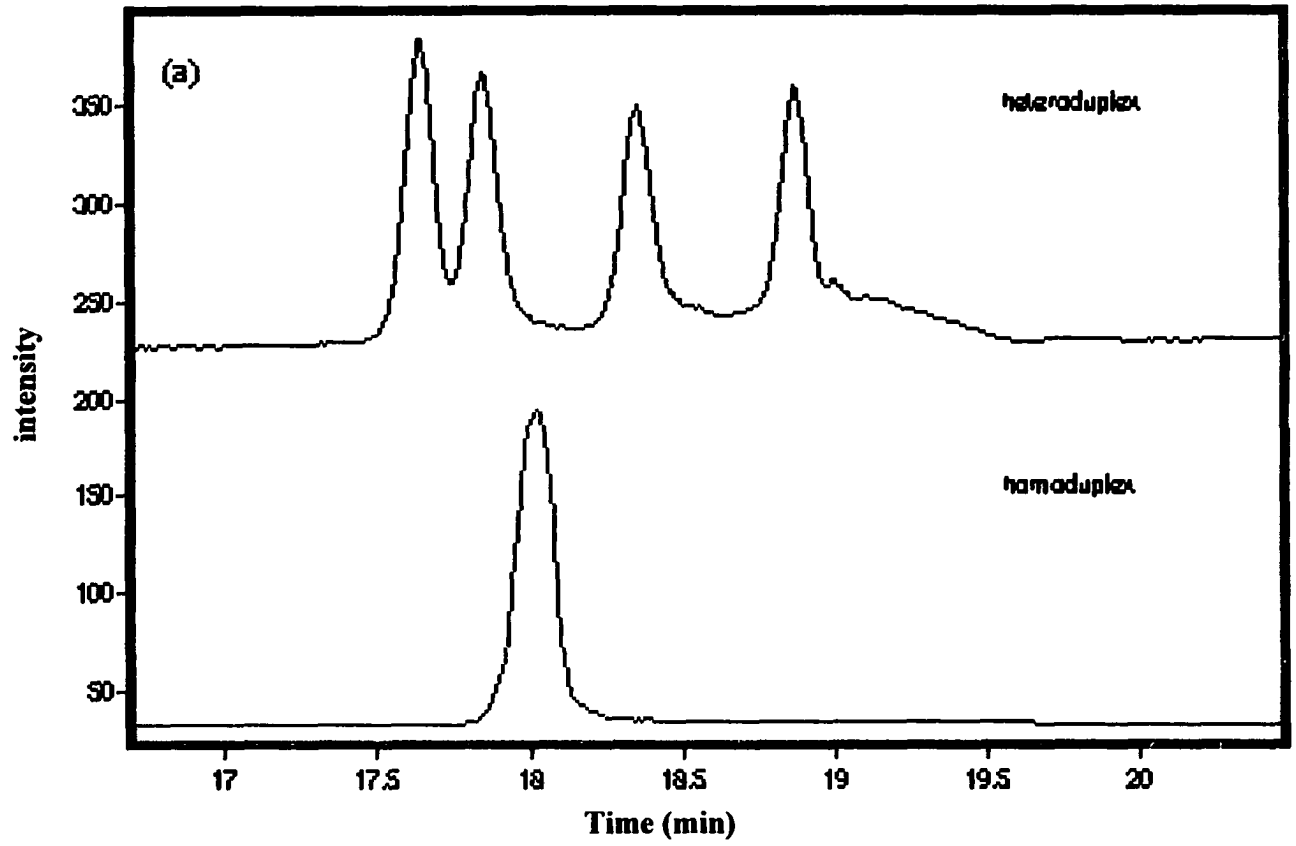

(a)
heteroduplex
homoduplex
intensity
350
300
250
200
150
100
50
17
17.5
18
18.5
19
19.5
20
Time (min)

Figure 5

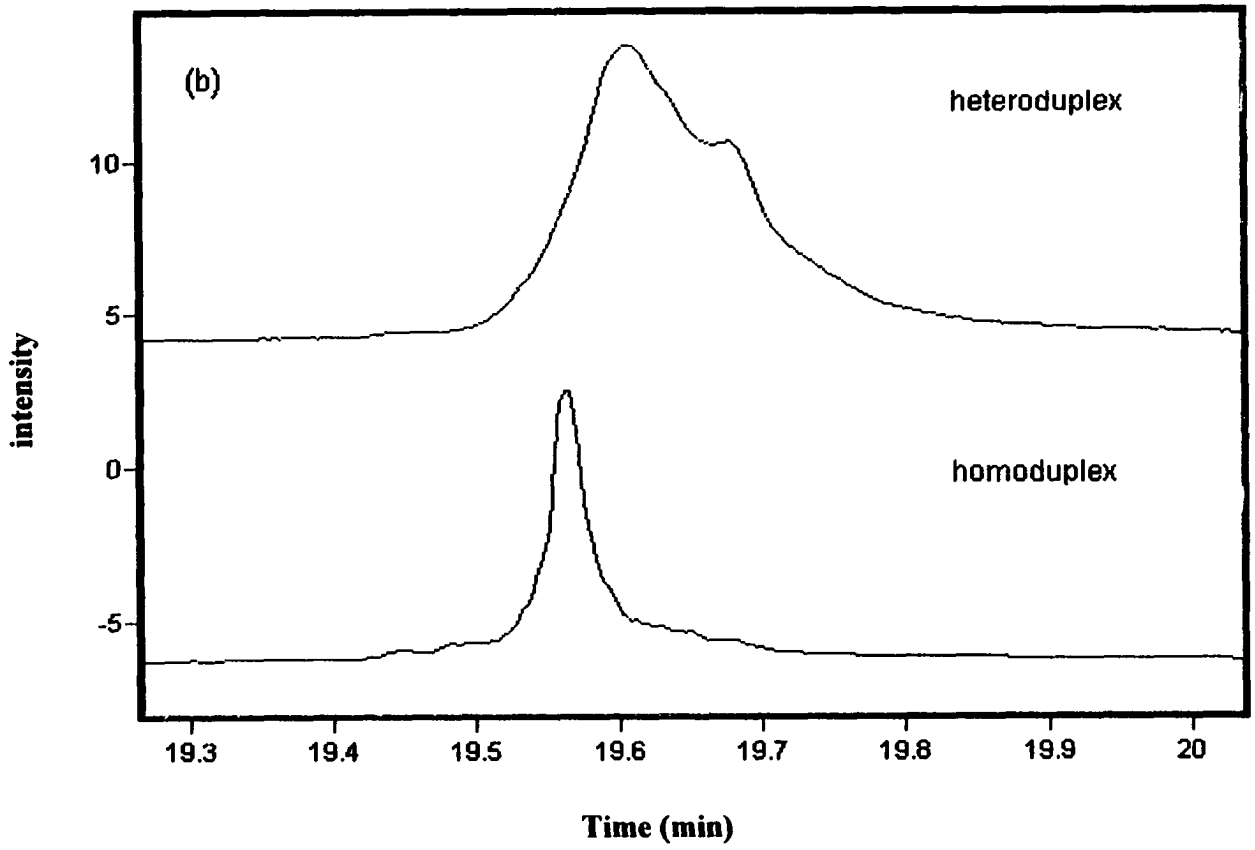


Figure 5 (continued)

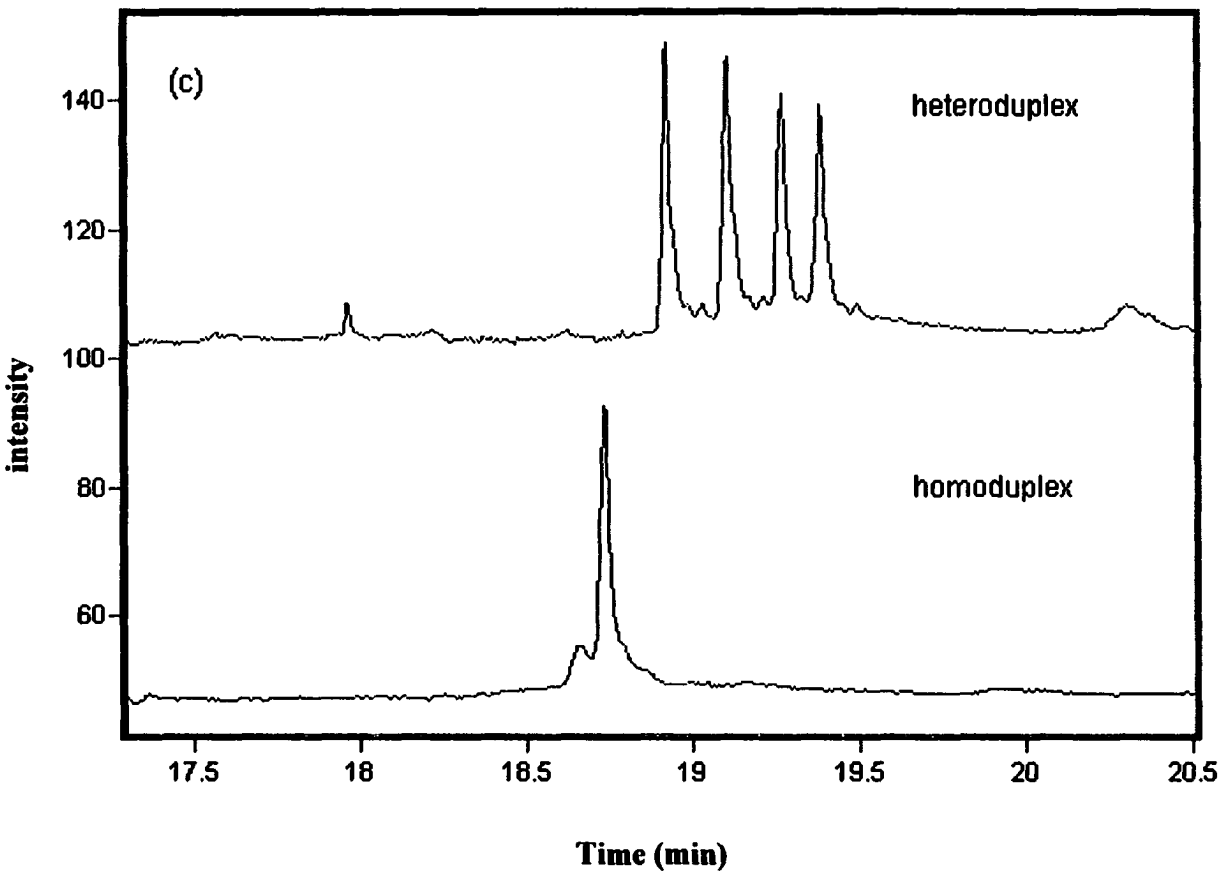


Figure 5 (continued)

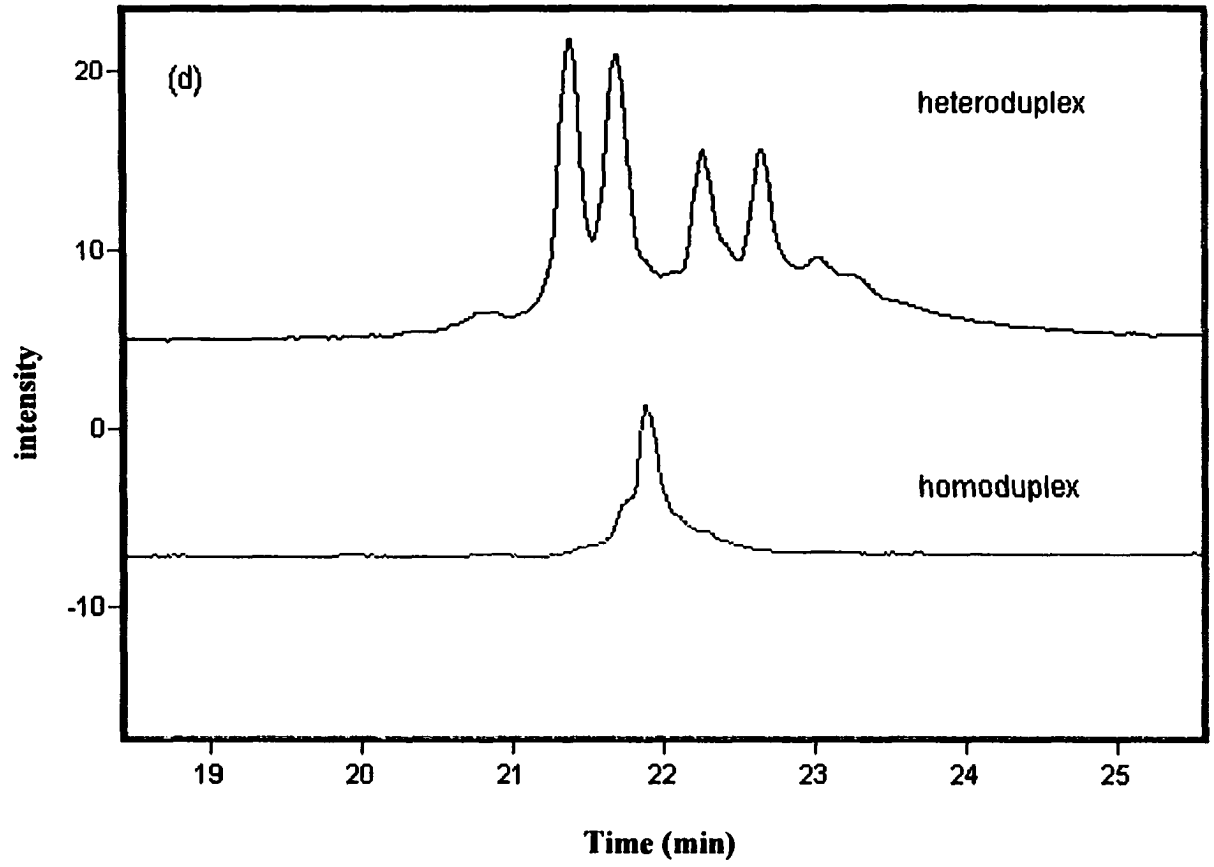


Figure 5 (continued)

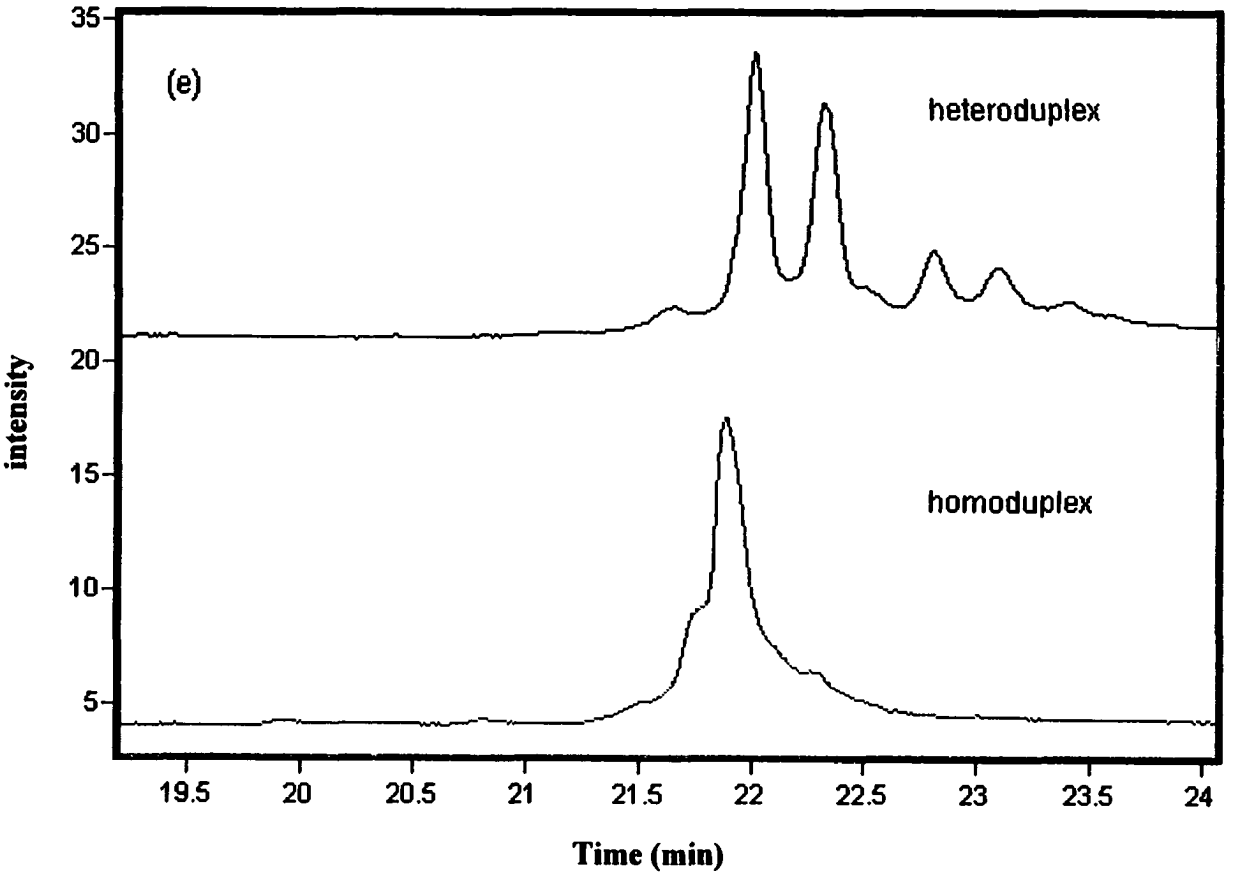


Figure 5 (continued)

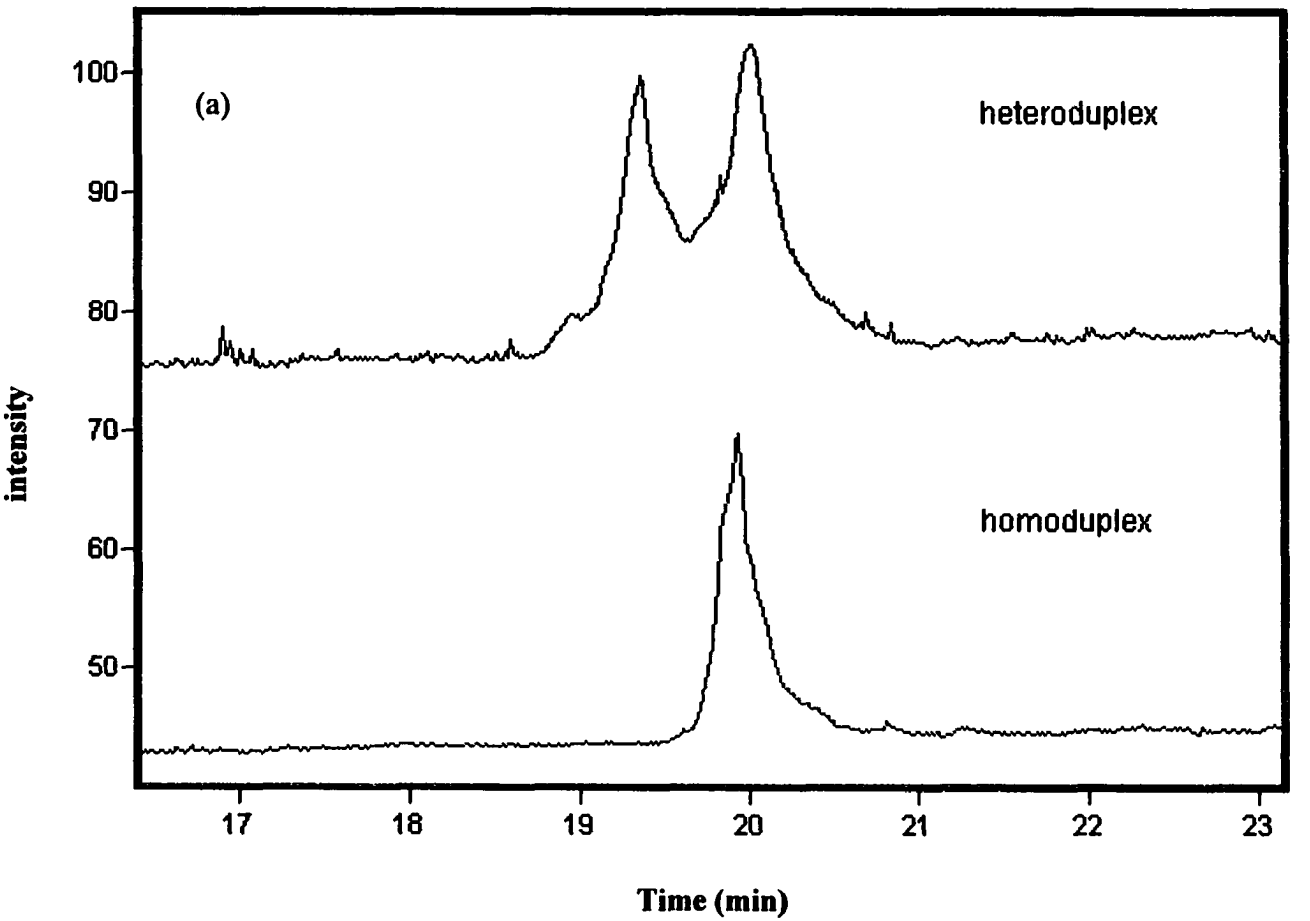

(a)
heteroduplex
homoduplex
intensity
100
90
80
70
60
50
17
18
19
20
21
22
23
Time (min)

Figure 6

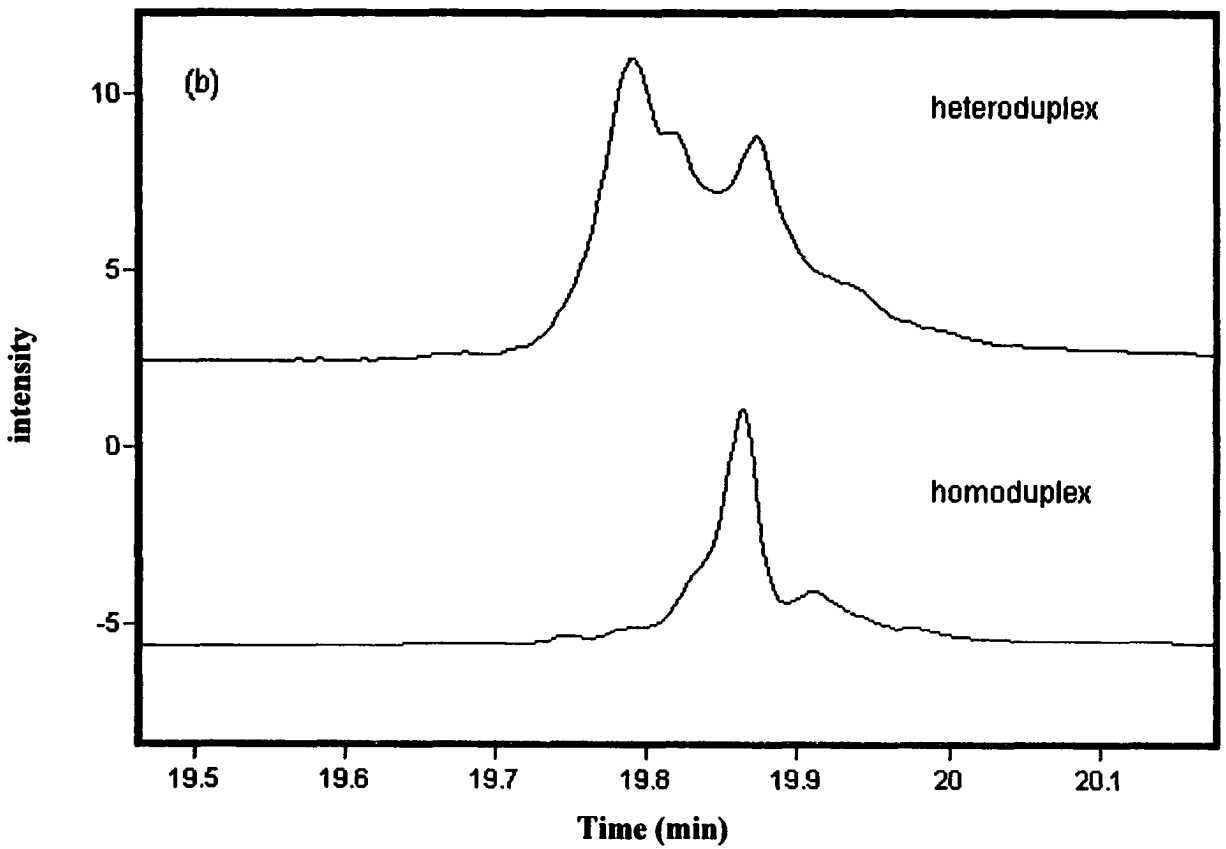


Figure 6 (continued)

CHAPTER 5. SIMULTANEOUS GENETIC TYPING FROM MULTIPLE SHORT TANDEM REPEAT LOCI USING 96-CAPILLARY ARRAY ELECTROPHORESIS SYSTEM

A paper accepted by Electrophoresis

Qiufeng Gao, Ho-ming Pang and Edward S. Yeung

Abstract

Short tandem repeats (STR) markers are highly polymorphic and widely used in human identification and genetic mapping. We demonstrate fast and reliable genotyping based on the four STR loci vWF, THO1, TPOX, CSF1PO by multiple-capillary array electrophoresis. Extracted human genomic DNAs were amplified by polymerase chain reaction (PCR). The PCR products were mixed with pooled allelic ladder as an absolute standard and coinjected from a 96-vial tray. Separations were performed in poly(vinylpyrrolidone) (PVP) sieving matrix with a one-hour turnaround time, with no degradation over 27 runs. Simultaneous one-color laser-induced fluorescence detection was achieved by using a CCD camera. The allele peaks for the unknown sample were identified by comparing the normalized peak intensities of the mixtures to those of the pooled ladder by using a straightforward algorithm. An extremely high level of confidence in matching the bands was indicated with negligible cross-talk (< 0.89%) between adjacent capillaries. This scheme is applicable for STR genotyping with high resolution, high speed and high throughput.

Introduction

The analysis of short tandem repeat (STR) loci provides a reliable way for forensic applications and genetic mapping. Typically tetranucleotide repeat loci are the choice for most applications because of their high degree of polymorphism in human populations.[1] Genotyping with STR loci involves polymerase chain reaction (PCR) amplification of human genomic DNA, separation and size determination of the PCR products. The separation of the PCR fragments is typically performed by slab gel electrophoresis. Recently, capillary electrophoresis (CE) has become an attractive alternative to slab gels in a wide range of DNA separations including DNA sequencing[2-7] and genotyping.[8-13] Although capillary electrophoresis has shown a great improvement in terms of separation speed, there is still the limitation of throughput since population-based human identification studies require large numbers of samples. A scheme to greatly increase the overall throughput is to use an array of capillaries in a single instrument. For STR analysis, 48-capillary array electrophoresis with a confocal fluorescence scanner has been reported.[14-16] A mechanical stage was used to translate the capillary array across the optical region. Our group has developed a 96-capillary system with on-column detection for DNA separation.[17] We used a modified version of this system for the present study.

Other multiplexed CE schemes have been suggested recently. An interesting format is the use of machined channels.[18-23] Although low-resolution genotyping can be achieved in a 26-mm channel length,[23] it will be necessary to have 30-cm long channels to provide adequate resolution of fragments up to 1000 bp.[24] This will be desirable for DNA sequencing or to include more loci for multiplex genotyping. The glass or silicon surfaces of these channels will need new surface treatment procedures to avoid electroosmotic flow or other

interactions with the DNA. Sample injection and sieving-matrix introduction still need to be implemented. Single-base resolution for the larger DNA fragments has yet to be demonstrated. So, with today's technology, multiple capillaries are still the mature format for multiplexed CE runs.

Despite the promising outlook on the use of multiple capillaries for genotyping and for DNA sequencing and the imminent release of commercial prototypes, all reports in the literature and at conferences to date show only fragmentary results. Reconstructed image files resembling slab-gel data were exhibited without full data analysis. Typically only one capillary was selected from the array and a few PCR products were displayed together with a size standard after software smoothing. Although statistics on genotype identification were presented regarding uniformity and ruggedness over a 48-capillary array,[16] the raw data were not available to judge the variations in migration times and intensities. The total turnaround time over many runs (true throughput) was not reported. Technology optimized for multiple runs in single capillaries or for single runs in large arrays are not sufficient for large-scale screening applications.

The sieving matrix plays an important role in high-resolution separations of DNA fragments. Linear polymer solutions are widely used in CE in consideration of reliability and ease of column regeneration. In our previous work,[25] we reported a sieving matrix composed of linear poly(vinylpyrrolidone) (PVP). This sieving matrix has very low viscosity, excellent self-coating effect in bare capillaries and good performance in separations of DNA fragments.[26] These characteristics are favorable for applications in capillary array systems.

In slab gels, the size determination of PCR fragments can be performed by comparing the migration distance of a PCR fragment to that of a standard size marker run either in the

same lane or in an adjacent lane. However, in CE, adjacent lanes cannot be employed because the nature of the capillary walls, internal diameter, temperature, and sieving matrix all contribute to migration variations among the capillaries. Instead, size determination is usually performed by coinjecting an internal size marker along with the PCR fragments. For fluorescence detection, the size marker can be a pooled allelic ladder or other known DNA fragments labeled with a different dye. Multiwavelength detection is employed and the actual size of each PCR fragment is determined by matching the DNA bands in two channels.[9] However, mobility shifts can be present due to structural differences of the dyes used.[27,28] The size marker can also be two DNA fragments labeled with the same dye as that on the PCR fragments if overlap is absent. The two marker fragments should be able to bracket the allelic ladder. The actual size of each PCR fragment is then calculated by interpolation from the relationship between the fragment size and its migration time (or mobility).[29] There can still be mobility variations caused by different base compositions even though the DNA fragments have identical sizes.[30,31] The experimental conditions must also be strictly controlled through the whole experiment. Zhang and Yeung[10] reported a size determination method based on using the coinjected pooled allelic ladder as an absolute standard. This method enables simple data interpretation and accurate typing from one-color detection.

In this paper, we demonstrate accurate genotyping of several individuals in a 96-capillary array system with the pooled allelic ladder as an absolute standard. A straightforward algorithm allows highly confident genotyping to be performed without human intervention. Complete statistics are presented to show reliability. Fast and high-resolution separations were achieved in this system. Column regeneration in-between runs

was not necessary with the PVP sieving matrix, and a large number of consecutive runs can be performed in the array without degradation in performance.

Experimental Section

Chemical Reagents

1xTBE buffer was prepared by dissolving 8.5 g pre-mixed TBE buffer powder (Amerosco, Solon, OH) and 240 g urea (ICN Biomedicals, Aurora, OH) into deionized water to form 500 ml solution. Poly(vinylpyrrolidone) (PVP) was obtained from Sigma (St. Louis, MO). The sieving matrix was made by dissolving 6.2% (w/v) of 360,000 M_W PVP and 1x TBE buffer. The mixture of the polymer powder and buffer solution was first shaken for 2 min and left standing for 1 h to get rid of bubbles.

DNA Samples

InstaGene™ whole blood extraction kit was obtained from Bio Rad (Hercules, CA). Whole blood samples were drawn from 3 individuals. Genomic DNA extraction was performed by following the manufacturer's protocol. TAMRA labeled primers for THO1, TPOX, and CSF1PO were purchased from Perkin-Elmer Applied Biosystems Division (PE-ABD, Foster City, CA) along with PCR Amplification kit. TAMRA labeled primer pair for vWF were custom-synthesized by Integrated DNA Technology (Coralville, IA). The sequences of the synthesized oligonucleotides are: Forward, 5'-TAMRA-CCCTAGTGGAT-GATAAGAATAATCAGTATG-3'; Reverse, 5'-GGACAGATGATAAATACATAGGATG-GATGG-3'. The PCR reactions were performed according to manufacturer's protocol. For THO1, TPOX, and CSF1PO, the reaction parameters are as follows: hold at 93 °C for 3 min;

30 cycles of denaturing at 94 °C for 1 min, annealing at 66 °C for 1 min, and extension at 72 °C for 1 min; hold at 60 °C for 30 min. The reaction parameters for vWF were modified as follows: hold at 93 °C for 3 min; 30 cycles of denaturing at 94 °C for 45 s, annealing at 59 °C for 1 min, and extension at 72 °C for 1 min; hold at 70 °C for 10 min. The four loci of each individual were amplified separately and then combined together. The TAMRA labeled CTTv ladder was obtained by amplifying the diluted (5000x) fluorescein-labeled pooled allelic ladder from Promega (Madison, WI).

The PCR products were purified by ethanol precipitation[32] and dried by vacuum. 300 • 1 denaturing reagent composed of 50% (v/v) saturated urea in D.I. water and 50% (v/v) formamide was used to resuspend each sample. The individual samples were mixed with the pooled ladder at a ratio of 4:1 (v/v) before injection. A 96 (8 x 12)-well microtiter plate was used as the sample tray. 4 μl of sample was pipetted into each tube. The samples were arranged in order so that capillary 1 is for the pooled ladder, capillary 2 is for mixture 1, capillary 3 is for mixture 2, and capillary 4 is for mixture 3, and so on.

Instrumentation

Figure 1 shows the experimental setup. 96 capillaries (75 μm i.d. 150 μm o.d., Polymicro Technologies Inc.) with 35 cm effective length and 55 cm total length were packed side by side at the detection windows. An Ar^+ ion laser (Coherent I-90) with 80-mW output at 514 nm was used for excitation. A 5 cm focal length cylindrical lens was used to expand the laser beam to ~4 cm upon the detection window to cover the entire array. Another 10-cm focal length cylindrical lens was used to focus the laser beam vertically on the detection windows. A CCD (Princeton Instruments, Inc., model TE/CCD-512-TKB) with a 28 mm

Nikon camera lens was used for image capture. The CCD was arranged in a way such that ~300 pixels were used to view the 1.5 cm detection region. To simplify alignment, a 5 x 300 pixels image was used on the detection windows to ensure all capillaries were covered. Winspec software (version 1.3B) from Princeton Instruments was used to control the CCD camera. A holographic notch filter (Kaiser Optical System, Inc.) was located in front of the CCD chip to filter out the laser scatter. At the ground end for the capillary array, the capillaries were bundled together to allow simultaneously matrix filling. At the sample injection end, the capillary array was spread out and mounted on a copper plate to form an 8 x 12 format with dimensions that fit a 96-well microtiter plate for sample injection. In addition, 96 gold-coated pins (Mill-Max Mfg. Corp.) were located next to the capillary tips to serve as electrodes. A high-voltage DC power supply from Spellman provided power for electrophoresis.

Experimental Procedure

The capillary array was first flushed with methanol and then water for cleaning. 2 ml of PVP sieving matrix was used to fill the capillary array under 100 psi pressure while the injection end was immersed into the buffer. After filling, the ends of the capillary array were immersed into the buffer. The sample tray was cooked at 95 °C for 3 min and was immediately inserted into an aluminum tray at –20 °C so that the DNA samples were cooled down rapidly to prevent renaturing. 30 s at 220 V/cm field strength was used for sample injection. Then the buffer tray was shifted under the injection end of the array for electrophoresis at 220 V/cm. 300 ms exposure time was used for the CCD camera. After each run, fresh PVP matrix was used to simply flush out the old one to maintain the separa-

tion efficiency in the subsequent run. During idle periods, the capillary array was stored in deionized water.

Results and Discussion

Separation Performance of Multiple Capillaries

Figure 2 shows the CE separation of PCR products derived from three individuals by using a single capillary. Multiplexed amplification has been reported to generate PCR products in STR analysis.[33] In this study, we adopted separate amplification of each locus. This method has higher amplification efficiency so that a minimum number of reactions were needed to produce samples for 96 capillaries. It is clear that the three individuals possess different genotypes. To identify the alleles based simply on migration times is not reliable, since run-to-run variations are expected. Even a 0.5% change in migration time will be detrimental to identification, since this will create > 1 bp shift in the calculated fragment size.

With PVP as the sieving matrix, a large number of runs can be achieved in a multi-capillary array system without loss of resolution. Figure 3 shows the sixteenth run of a set of pooled CTTv ladder and mixtures of unknown samples injected together with the pooled ladder in the 96-capillary array. Other runs in a series of 27 experiments show similar results. The electropherogram channels have been converted into a 2-D image. The horizontal direction represents the capillary array arrangement while the vertical direction represents the migration time. Out of 96 capillaries, two capillaries (#30 and #86) showed no signal due to clogging from prior repeated use over a one-month period. Four capillaries with the pooled standard ladder showed lower resolution but the individual peaks could still be identified. All others showed uniform resolution and intensity distributions. It should be noted that the

capillaries with lower resolution were not the same ones in consecutive runs. Therefore, it seems unlikely that this is related to degradation of the capillary columns. Because of the manual operation, occasional problems with matrix filling and sample preparation may have contributed to the variations. Since the standard samples have a higher concentration of DNA, the likelihood of inadequate denaturation prior to injection is higher.

The time scale in Figure 3 is from 26 min to 50 min. This leads to a turnaround time of 1 h (with 5 min for sieving matrix replacement). At this speed, over 750 samples can be processed each day (8 h) in this instrument. This number may be increased by 10-fold if a larger scale instrument, e.g. with 1000 capillaries, is built or if automation to 24-h operation is implemented. In our 96-capillary system, the overall separation and detection performance in each capillary is very similar to that in a single capillary (Figure 2).

Figure 4 more clearly shows representative electropherograms that were randomly picked from each sample group. Single-base resolution was achieved between allele 9.3 (single-base deletion) and allele 10 at the THO1 locus. The S/N ratio is more than sufficient for precise peak-height measurements. With simple dilution or float dialysis of PCR products,[15] the S/N ratio may be improved further. This suggests that one may ultimately need even less sample in such experiments. The tolerance to lower PCR efficiency can also be enhanced.

Genotype Indication

Although mobility shifts may exist in CE and prevent precise sizing of the PCR fragments using other types of internal standards, a peak from the PCR products here should match the corresponding peak from the pooled allelic ladder perfectly since they have

identical structure. The relative peak intensities of the pooled ladder will be altered by the addition of the unknown PCR products. Therefore, by comparing the relative intensities of the sets of peaks associated with each locus, one can identify the genotype. Visual comparison between each sample (Figure 4a through c) and the pooled ladder (Figure 4d) readily allows identification of the unknown alleles depicted in Figure 2. This indicates one can establish an objective computer algorithm to identify unknown alleles.

It is quite difficult, even impossible, to control the absolute peak intensities of the injected pooled ladders among capillaries and between runs (see Figure 4). A number of factors may contribute to the variation, such as capillary diameter, ionic strength caused by the co-added PCR matrix, laser power distribution, light collection efficiency and sensitivities among CCD pixels, etc. Thus, using an internal standard to normalize the signal intensities within each locus is necessary to detect intensity changes in the peak patterns. Previous work has demonstrated that peak height is a suitable parameter for quantitation.[10] Any significant increase in the relative peak heights in the mixture compared to those in the pooled ladder alone will allow identification of the unknown alleles. We will assume that the pooled allelic ladder includes all alleles from the PCR products. If there is an abnormal genotype, extra peak(s) will show up in the electropherograms due to the excellent resolving power of this system. Extra peaks can be picked out by counting the peak number with standard software before applying our automated algorithm. Then the odd sample can be reanalyzed manually, very much like editing in DNA sequencing. Naturally, abnormal genotypes will be missed if they are at low concentrations, which will be true for any separation/detection scheme.

Semi-Automated Algorithm

90 electropherograms in one run (Figure 3) were extracted as ASCII files. Spikes were removed from each data set by interpolation based on the fact that these are much narrower features (one or two data points) than real allele peaks (7 or more data points). Electropherograms like those in Figure 4 were then subjected to standard chromatographic software (GRAMS/32, Galactic Industries, Salem, NH) for baseline correction and for peak picking. The data reduces to a set of 90 x 33 peak heights for 90 capillaries and 33 total alleles in the 4 loci.

The group of 20 capillaries corresponding to the pooled allelic ladder was analyzed first. In each of the locus, the most intense allelic peak was selected for intensity normalization. Less error should be introduced because of the higher S/N ratio and the reduced influence from contaminants. The peak heights of the other peaks in the allelic region were simply divided by the selected peak height to account for any variations in injection and excitation among the capillaries. The relative standard deviation (RSD) for each allelic peak is thereby reduced from 30% in the raw data to 4% after normalization. The RSD of the normalized data actually ranged from 1.3% to 7.0% in this experiment, depending on the S/N of each peak.

Next the peaks for each locus of the unknown mixtures are normalized to the selected peak in that capillary. To allow comparison among the many alleles, residuals are calculated as the fractional changes in peak heights between the unknown sample and the pooled ladder. By definition, the peak selected for normalization will always show a residual of zero. Most of the other allele peaks will give residuals close to zero since their relative intensities should remain constant in the absence of unknown alleles. Residuals significantly larger than zero

will indicate the presence of additional material in the mixture of unknown and pooled ladder. For example, a residual of 1 would indicate a 100% increase in the normalized intensity. Unknown alleles can thus be identified.

An objective and straightforward way to decide whether a residual is significantly larger than zero is to use the Q-test in statistics. Outliers in a data set can be determined at any given confidence level in this manner. Here, since one or two alleles can be present in each locus (homozygous versus heterozygous), we modified the Q-test to interrogate separately each of the two largest residuals in each locus. The Q value is calculated as the distance between each residual in question and the *third* largest residual in that locus divided by the range of residual values. The Q-test is then performed as usual to identify the unknown alleles.

A visual representation of our allele identification algorithm is given in Figure 5. Each frame represents one of the 90 capillaries. The data was analyzed twice, once normalized with the most intense peak in the pooled ladder (4 loci on the left) and once normalized with the least intense peak in the pooled ladder (4 loci on the right). As discussed above, zero means no change in relative intensity in the sample mixture, 1 means a 100% increase in relative intensity, and so on. Using the Q-test, all of the alleles in all capillaries were identified with an error rate of only 5 out of 720 (< 1%) at the 99% confidence level of the Q-test. At the 96% confidence level, one fewer error was found. At the 99.9% confidence level, two more errors were present.

A unique situation occurs when the allele selected for normalization coincides with an unknown allele in the sample. Such is the case in the left group in Figure 5a III. The alleles that are *absent* in the sample will take on *negative* values. This however does not affect the

Q-test, since the rank ordering remains unchanged.

Before drawing conclusions about the ruggedness of our genotyping algorithm, it is necessary to examine every situation where such allele identifications were in error. First, spikes or contamination peaks are present occasionally in the electropherograms. These are shown in Figure 6a. Judging from their intensities, they are likely to be dust particles or bubbles that crossed the laser beam. Spikes (second asterisk in Figure 6a) are narrow features that are readily removed digitally before applying the algorithm above and do not affect allele identification. Contamination peaks (first asterisk in Figure 6a) that fall right on top of an allele peak can be problematic. In Figures 5a and 6a, allele 4 in locus III is clearly contaminated. The peak height (off scale) is 9 times the normal value. By chance, this occurrence did not cause any allele identification errors because allele 4 was in fact present in that particular sample. However, errors (false positive) were evident in Fig. 5b (allele 5 in locus II) and 5c (allele 7 in locus IV) due to contamination peaks highlighted in Fig. 6b and 6c respectively. We note however that a contamination peak will also be mistaken as an unknown allele by any other genotyping algorithm.

Second, we are treating the 9.3 and 10 alleles in locus II as separate components. The electropherogram must therefore be able to resolve these two peaks. Figure 6d shows one capillary (out of 90) where this is not the case. Figure 5d shows a negative value for allele 6 as a result of the failure of our chromatography software to recognize that peak in the electropherogram. The 3 unknown samples in this study did not contain either allele 9.3 or allele 10, so no miscalls were registered. However, it is interesting that the Q-test failed to identify any alleles here. The unusually large range of residuals in the group distorted the Q value. This algorithm therefore provides useful diagnostics for abnormal peak heights within

the locus. Fig. 6d shows that if either allele is present, we will have a positive call but will not be able to immediately attribute that to one or the other allele. Manual examination of the electropherogram in such a situation will reveal a broad feature overlapping the two alleles. Presumably either manual recognition of the presence of a shoulder on either side or formal peak deconvolution will resolve any ambiguities. These single-base deletions in allelic loci are not common but need to be considered in genotyping. This is particularly true if the deleted allele was not present in the pooled ladder. These and other PCR artifacts such as a-addition should be detectable since the separation is adequate for single-base resolution.[25] Other potential problems such as stutter and weak signals will affect this approach and other schemes equally.

Third, we chose the allele in each locus with the best S/N and therefore the smallest RSD among the standard runs for normalization (left 4 groups in Figure 5). This should minimize error propagation to the other allelic peaks. For example, when we chose the allele with the lowest intensity in each locus for normalization (right 4 groups in Figure 5), one additional miscall (false negative) was introduced in the entire data set. The lower S/N led to larger uncertainties in the normalized peak heights, which is evident in Fig. 5d, right vs. left groups. However, most of the Q values here are well above the threshold even at the 99.9% confidence level that essentially the same identification accuracy was achieved.

Fourth, contamination peaks that do not appear right on top of an allele peak can change the peak order within a locus to cause misidentification. This is illustrated in Figure 6e in locus III. An extra peak was picked up by the chromatography software. Peaks 3-9 actually correspond to alleles 2-8. However, simply verifying the peak count will eliminate any confusion. The correct allele peaks (peaks 1 and 3) can be distinguished from the

contamination peak (peak 2) based on the migration time relative to those of the other alleles. The overlap among peaks 1-3 did however affect the peak height determination (see Figure 5e), and allele 2 was miscalled (false positive) based on the Q-test.

To summarize, the 5 misidentifications in the entire experiment are due to contamination peaks (Fig. 5b locus II allele 5, Fig. 5c locus IV allele 7, Fig. 5e locus III allele 2), fluctuating baseline (affecting peak-height measurements in locus IV of one capillary, data not shown), and inadequate separation (Fig. 5d). A sixth misidentification would have occurred if a different sample were being analyzed in Figure 5a.

Alternatively, from Figure 5, it is clear that if one simply chooses 0.5 as the threshold for a positive call and –0.5 as the threshold for a negative call for each allele, the accuracy for allele identification in our entire data set will be identical to above. For the case of negative calls among N possible alleles in a locus (Fig. 5a and 5e, left group locus III), one would need to find at least N–2 negatives in a locus to confirm the presence of the other alleles in the unknown sample. This method of allele identification can also be associated with statistical confidence limits. For example, the RSD of the negative alleles within that locus in that experiment can be calculated. The t-test can then be applied based on how many times RSD the heights of the positive alleles are larger than the average heights of the negative alleles. By applying this alternative allele identification scheme, confidence levels similar to the Q-test are found.

It should be noted that several checks for internal consistency exist in this scheme. For pure samples (single individual), there can be no more than 2 positive alleles. The same data set can be recalculated N times using each allele in the locus for normalization. The genotype calls must all agree each time. If not, or if no alleles are found, it is possible to

lower the confidence limit and recalculate. Identification at a lower confidence limit is not as satisfying, but we note that this scheme, unlike the use of other types of internal standards, is free from the effects of migration or injection variances. Naturally, failure to identify any alleles can also mean that there is not enough PCR product to alter the intensity pattern significantly in that locus. A rough rule of thumb given the variations found in these experiments is that in the mixture the PCR products should be at least 50% of the corresponding amounts in the pooled ladder for accurate genotyping.

Acknowledgments

The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract W-7405-Eng-82. This work was supported by the Director of Energy Research, Office of Health and Environmental Research, and by the National Institutes of Health under grant HG-01385.

References

1. Edwards, A., Civitello, A., Hammond, H. A., Caskey, C. T. *Am. J. Hum. Genet.* **1991**, 49, 746-756.
2. Cohen, A. S., Najarian, D. R., Paulus, A., Guttman, A., Smith, J. A., Karger, B. L. *Proc. Natl. Acad. Sci. (USA)* **1987**, 85, 9660-9663.
3. Smith, L. M. *Nature* **1991**, 349, 812-813.
4. Drossman, H., Luckey, J. A., Kostichka, A. J., D'Cunha, J., Smith, L. M. *Anal. Chem.* **1990**, 62, 900-903.

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

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

7. Fung, E. N., Yeung, E. S. *Anal. Chem.* **1995,** 67, 1913-1919.

8. Butler, J. M., McCord, B. R., Jung, J. M., Allen, R. O. *Biotechniques* **1994,** 17, 1062-1070.

9. Wang, Y., Ju, J., Carpenter, B. A., Atherton, J. M., Sensabaugh, G. F., Mathies, R. A. *Anal. Chem.* **1995**, 67, 1197-1203.

10. Zhang, N. Y., Yeung, E. S. *Anal. Chem.* **1996,** 68, 2927-2931.

11. Mitchell, J. W., Walsh, D. *Electrophoresis* **1998,** 19, 80-85.

12. Lazaruk, K., Walsh, P. S., Oaks F., et al., *Electrophoresis* **1998**, 19, 86-93.

13. Isenberg, A. R., Allen, R. O., Keys K. M., et al., *Electrophoresis* **1998**, 19, 94-100.

14. Wang, Y., Wallin, J. M., Ju, J., Sensabaugh, G. F., Mathies, R. A. *Electrophoresis* **1996,** 17, 1485-1490.

15. Vainer, M., Enad, S., Dolnik V., et al., *Genomics* **1997**, 41, 1-9.

16. Mansfield, E. S., Robertson, J. M., Vainer, M., et al., *Electrophoresis* **1998,** 19, 101-107.

17. Ueno, K., Yeung, E. S. *Anal. Chem.* **1994,** 66, 1423-1431.

18. Mathies, R. A., Ju, J., Zhu, H., Clark, S. M., Woolley, A. T., Wang, Y., Benson, S. C., Glazer, A. N. *DOE Human Genome Workshop IV*, Santa Fe, NM, November 13-17, **1994**, Abstract #133.

19. Balch, J., Davidson, C., Gingrich, J., Sharaf, M., Brewer, L., Koo, J., Smith, D., Albin, M., Carrano, A. *DOE Human Genome Workshop IV*, Santa Fe, NM, November 13-17, **1994**, Abstract #134.

20. Woolley, A. T., Mathies, R. A. *Proc. Natl. Acad. Sci. (USA)* **1994**, 91, 11348-11352.

21. Woolley, A. T., Sensabaugh, G. F., Mathies, R. A. *Anal. Chem.* **1997**, 69, 2181-2186.

22. Simpson, P. C., Roach, D., Woolley, A. T., Thorsen, T., Johnston, R., Sensabaugh, G. F., Mathies, R. A. *Proc. Natl. Acad. Sci. (USA)* **1998,** 95, 2256-2261.

23. Schmalzing D., et al. *Proc. Natl. Acad. Sci. (USA)* **1997,** 94, 10273-10278.

24. Schmalzing, D., Adourian, A., Koutny, L., Ziaugra, L., Matsudaira, P., Ehrlich, D. *Anal. Chem.* **1998**, 70, 2303-2310.

25. Gao, Q., Yeung, E. S. *Anal. Chem.* **1998**, 70, 1382-1388.

26. Madabhushi, R. S., Menchen, S. M., Efcavitch, J. W., Grossman, P. D. *U.S. Patent #5,567,292*, October 22, **1996**.

27. Tan, H., Yeung, E. S. *Electrophoresis* **1997**, 18, 2893-2900.

28. Ju, J., Kheterpal, I., Scherer, J. R., Ruan, C., Fuller, C. W., Glazer, A. N., Mathias, R. A. *Anal. Biochem.* **1995,** 231, 131-140.

29. Butler, J. M., McCord, B. R., Jung J. M., et al. *Electrophoresis* **1995,** 16, 974-980.

30. Wenz, M. *Nucleic Acids Res.* **1994**, 22, 4002-4008.

31. Chang, H-T., Yeung, E. S. *J. Chromatogr. B* **1995**, 669, 113-123.

32. Baskin, J., Marsh, M., Barker, D., Johnston, R., *Appl. Theor. Electrophor.* **1996,** 6, 23-28.

33. Kimpton, C. P., Gill, P., Walton, A., Millican, E. S., Adams, M. *PCR Methods Appl.* **1993**, 2, 13-22.

Figure Captions

Figure 1. Schematic of experimental setup of the 96-capillary array system.

Figure 2. Electropherograms of PCR products from unknown samples in the CTTv region. Traces (a), (b), and (c) refer to samples from individuals A, B and C, respectively. Separation matrix: 6.2% 360K M_W PVP in 1xTBE with 8 M urea. Running buffer: 1x TBE with 8 M urea. Running voltage: −12 kV. Injection: −12 kV for 15 s. Capillary: 75 • m i.d., 55 cm total length, 35 cm effective length. Detection: 543.5 nm HeNe laser with 5 mW output. Fluorescence detection was accomplished by using a PMT with 610 nm long-pass filter.

Figure 3. 2-D image of the electropherograms of the pooled ladder and the mixtures of individual samples with pooled ladder in 96 capillaries. The horizontal direction is capillary number; the vertical direction is migration time. Different samples are injected into consecutive capillaries. The time scale is from 26 min to 50 min.

Figure 4. Representative electropherograms of the four groups of samples. Each electropherogram is randomly picked up from the corresponding groups in the 96-capillary array. Traces (a), (b), (c), (d) are for the pooled ladder, the mixtures of pooled ladder and individual samples A, B and C, respectively. Traces (b), (c), (d) were shifted to the time scale of trace (a) for comparison.

Figure 5. Relative peak height residuals of selected traces in the array. Calculations were performed twice with different peaks selected for normalization. Left group: normalized to the most intense peak; right group: normalized to the least intense peak. Loci I through IV are vWF, THO1, TPOX and CSF1PO, respectively.

Figure 6. Electropherograms corresponding to the data presented in Figure 5. Asterisks refer to features that can affect allele identification.

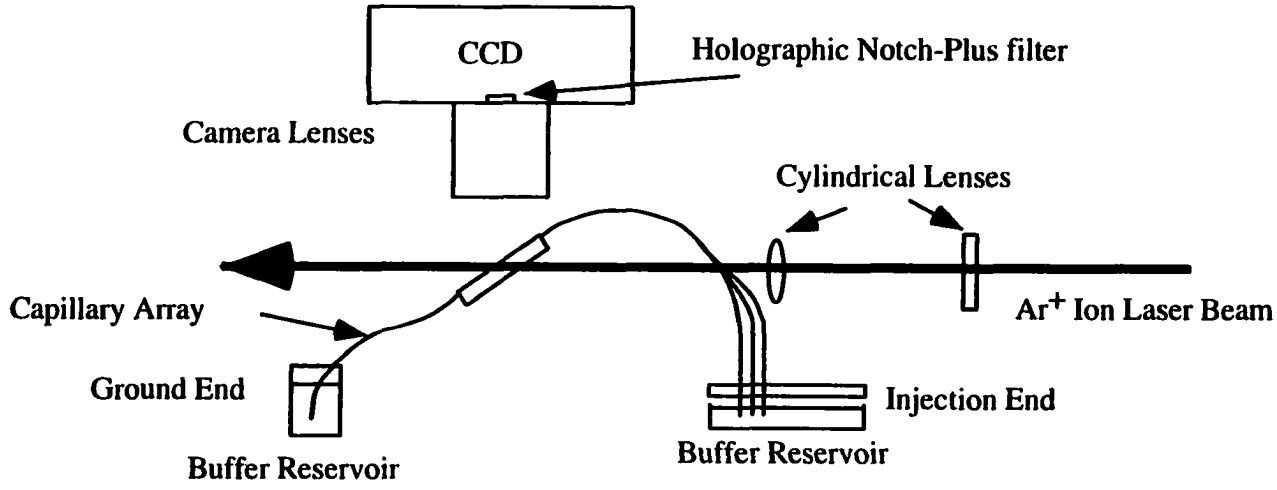

CCD
Holographic Notch-Plus filter
Camera Lenses
Cylindrical Lenses
Capillary Array
Ar+ Ion Laser Beam
Ground End
Injection End
Buffer Reservoir
Buffer Reservoir

Figure 1

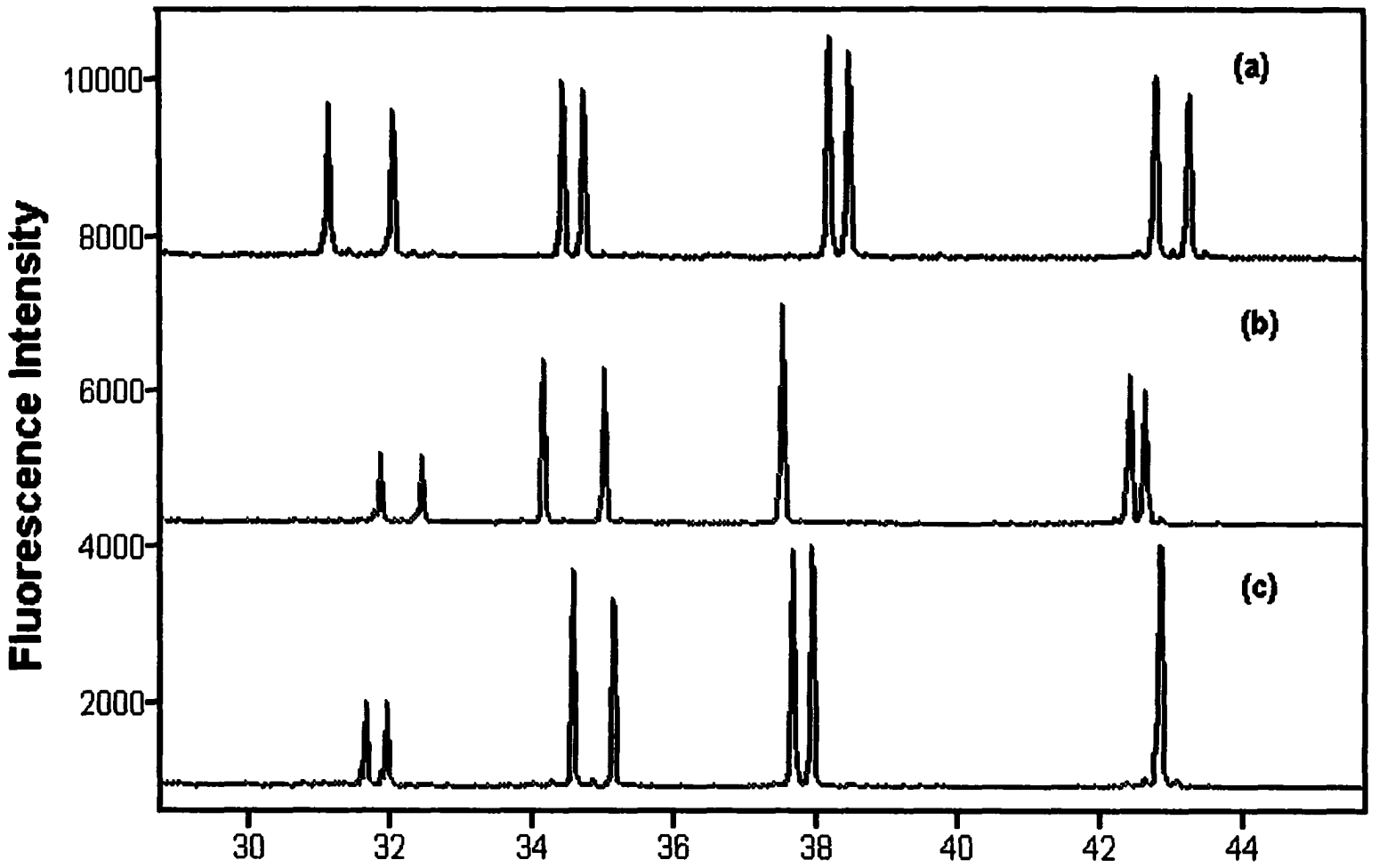

Fluorescence Intensity
10000
8000
6000
4000
2000
(a)
(b)
(c)
30
32
34
36
38
40
42
44

Figure 2

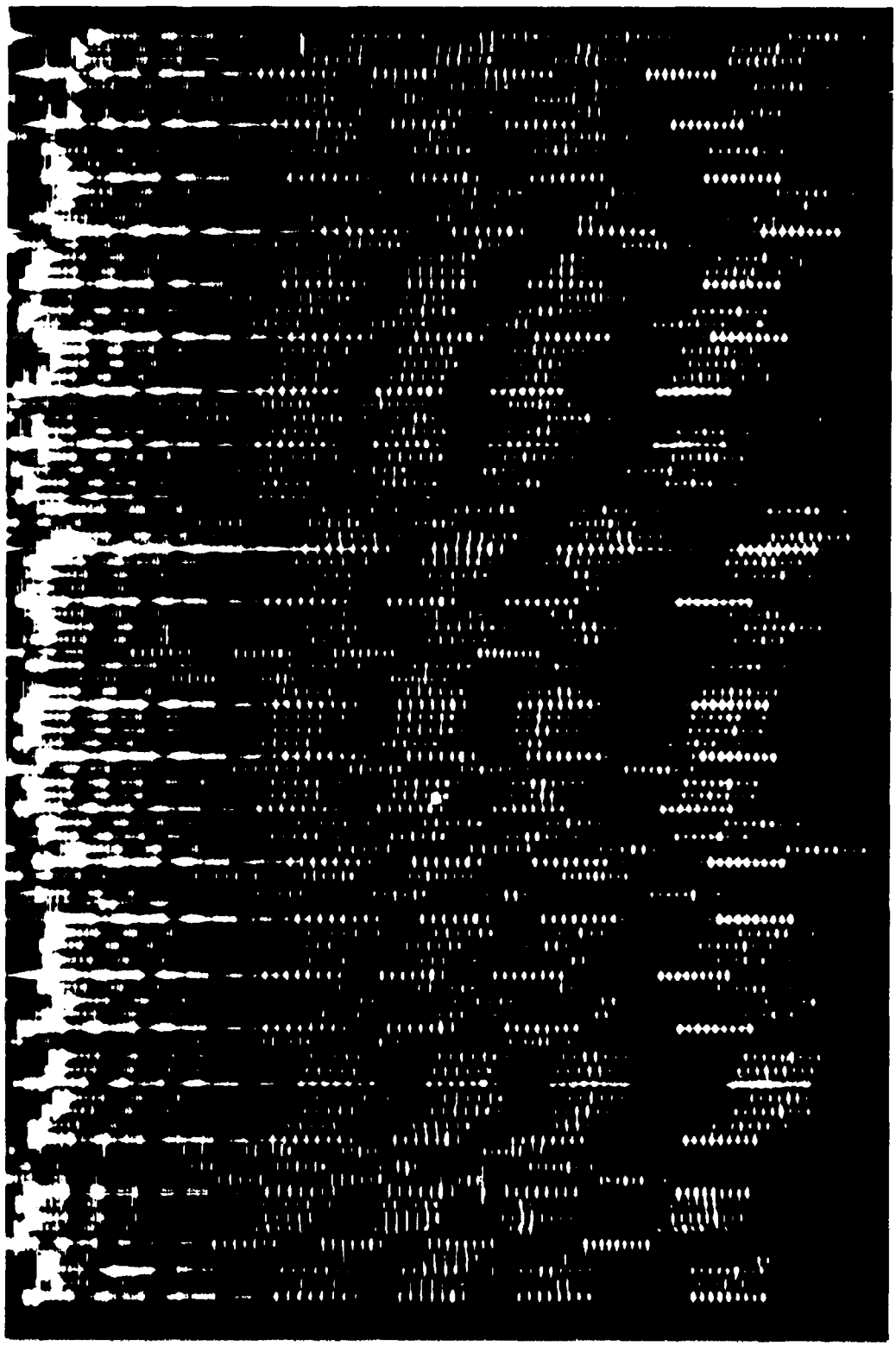


Figure 3

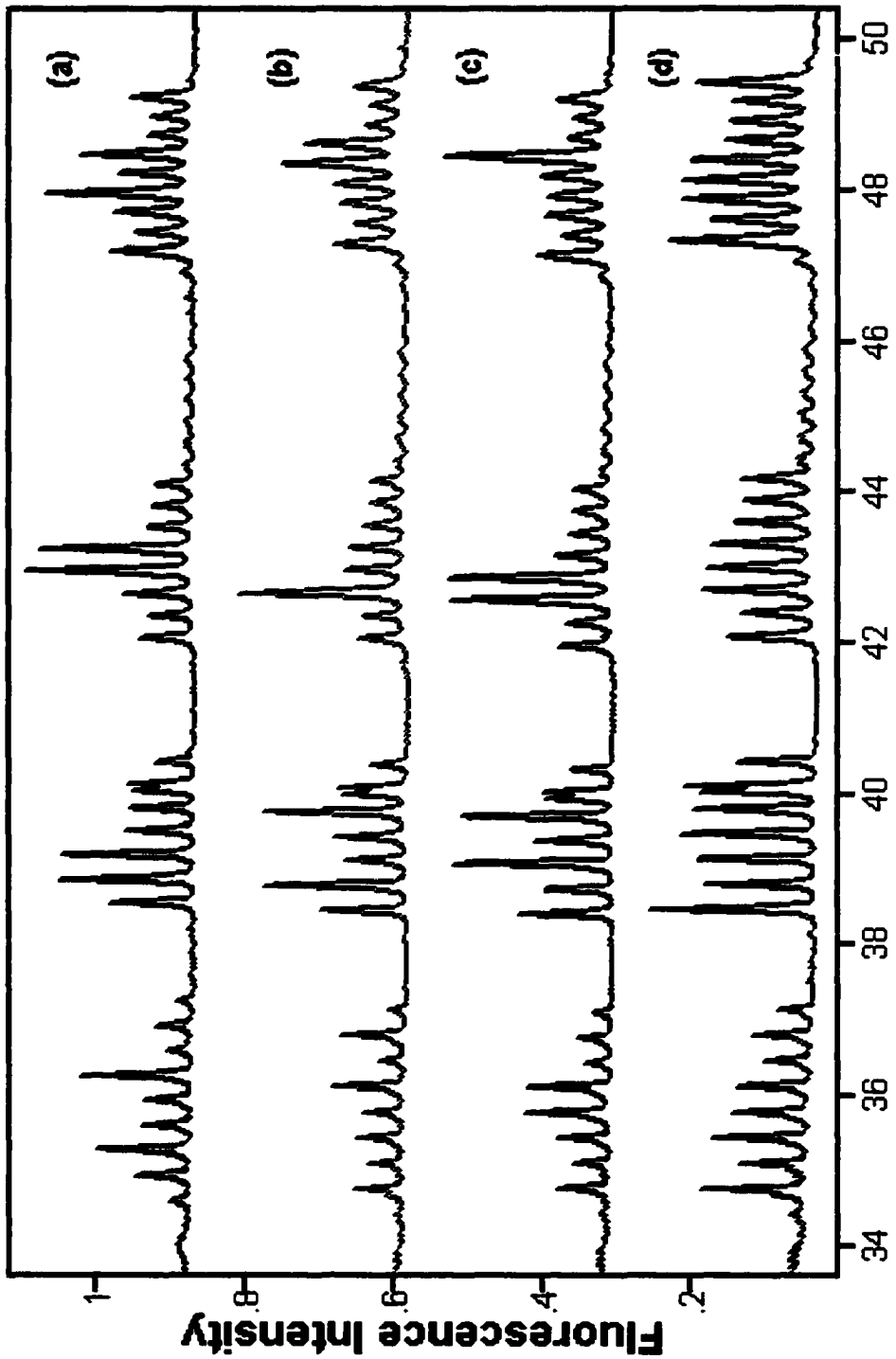


Figure 4

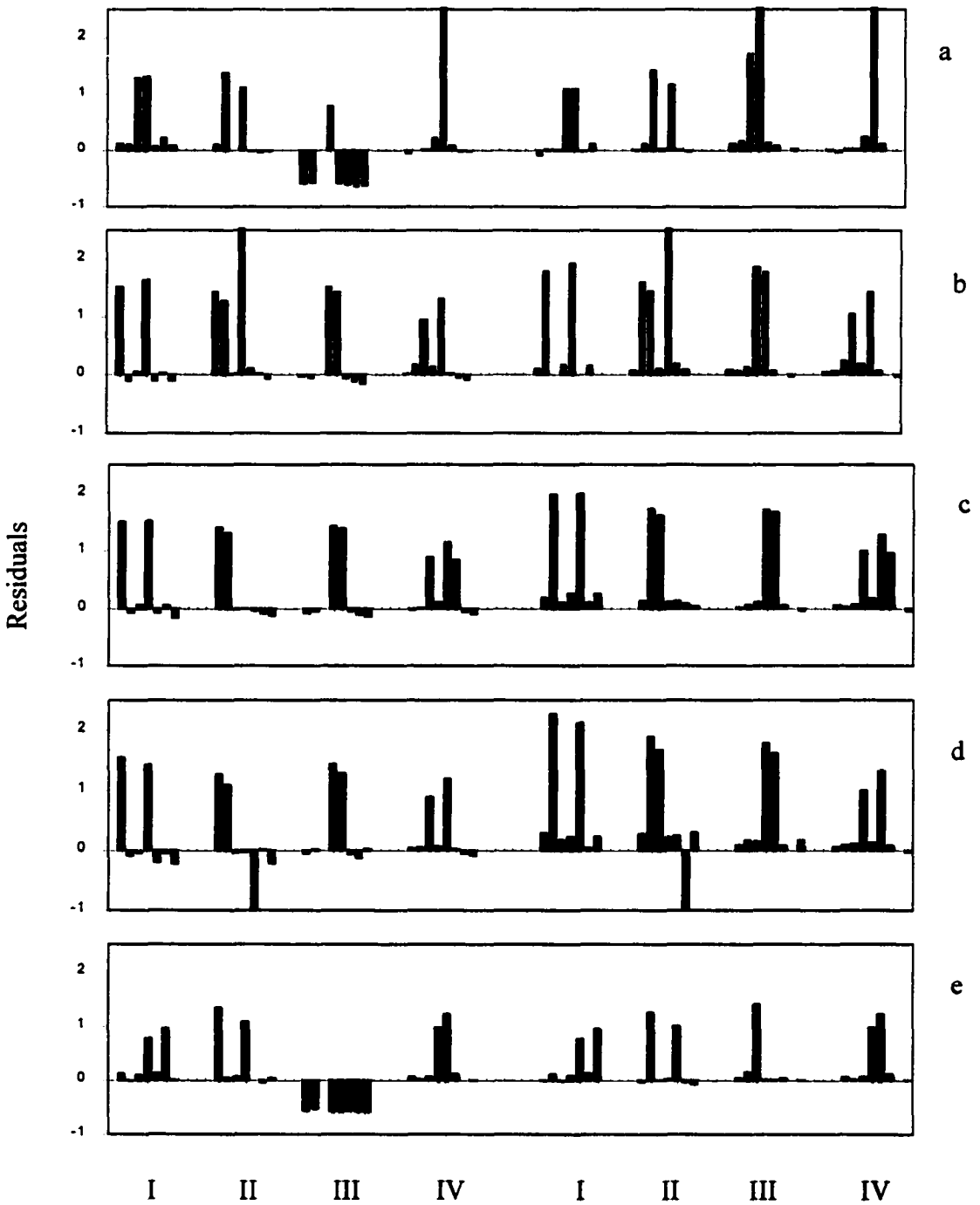


Figure 5

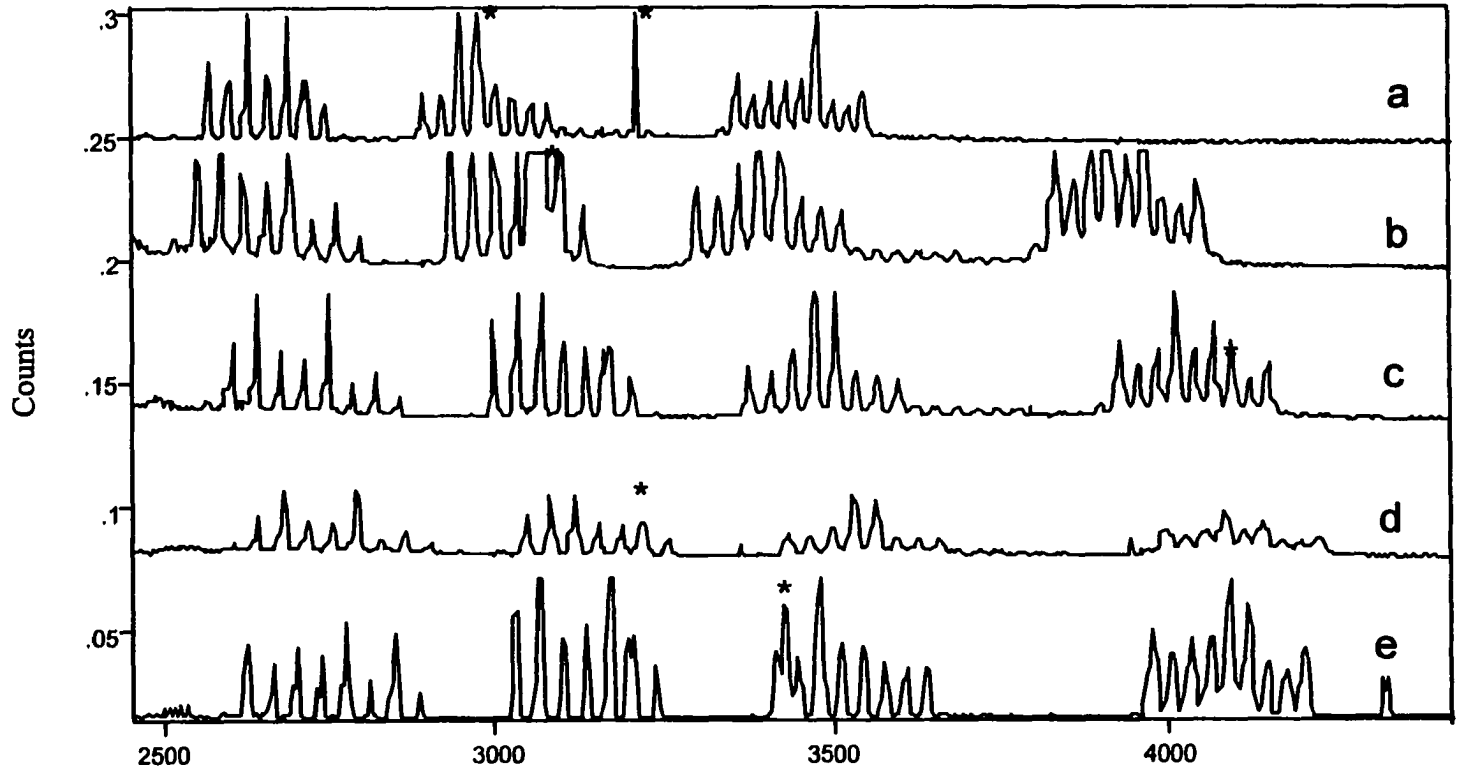


Figure 6

CHAPTER 6. GENERAL CONCLUSIONS

DNA analysis has become more and more important in the fields of genetic mapping, forensic analysis and clinical diagnosis. Capillary electrophoresis provides a fast way for the separation of DNA fragments. It needs a small amount of sample and is easy for automation. With capillary array electrophoresis, much higher throughput can be achieved. In both capillary electrophoresis and capillary array electrophoresis, the sieving matrix plays an important role to the performance. Modern capillary electrophoresis systems still demand sieving matrices with high performance and ease of operation.

The work in this dissertation has described the development of a new sieving matrix for DNA analysis, including sequencing, mutation detection and genotyping. First of all, the influence of capillary wall coating on the DNA sequencing performance was evaluated. Secondly, Polyvinylpyrrolidone was introduced for DNA analysis. It has very low viscosity, excellent EOF suppressing effect so that it is quite friendly to the capillary. The application of PVP on the separation of DNA sequencing fragments and genotyping fragments has been validated. Further work with PVP was performed in Mutation detection. A fast mutation detection scheme with temperature gradient capillary electrophoresis using PVP as sieving matrix was accomplished. Finally, A high-throughput genotyping system with 96-capillary array and PVP sieving matrix was successfully demonstrated. Short turnaround time and tens of runs of capillary array lifetime made this system practically applicable.

The combination of 96-capillary array electrophoresis and PVP sieving matrix should be expandable to other high-throughput DNA separations, e.g. mutation detection. The appealing features of this system will draw more interest in similar applications such as protein mapping, etc.

APPENDIX

CDCE	Constant Denaturing Capillary Electrophoresis
CDGE	Constant Denaturing Gel Electrophoresis
CSGE	Conformation sensitive Gel Electrophoresis
DGCE	Denaturing Gradient Capillary Electrophoresis
DGGE	Denaturing Gradient Gel Electrophoresis
DHPLC	Denaturing High Performance Liquid Chromatography
LIF	Laser Induced Fluorescence
PCR	Polymerase Chain Reaction
PVP	Polyvinylpyrrolidone
SNP	Single Nucleotide Polymorphism
SSCP	Single Strand Conformational Polymorphism
TGCE	Temperature Gradient Capillary Electrophoresis
TGGE	Temperature Gradient Gel Electrophoresis
ds-DNA	Double-stranded DNA
ss-DNA	Single-Stranded DNA

ACKNOWLEDGEMENTS

First, I feel indebted so much to Dr. Yeung, my major professor. He introduced me the exciting area of bioanalysis after I came to Iowa State University. Our first conversation led me to the decision of electing him as my major professor and exploring this challenging field. Over the past few years, his curiosity and dedication to research has inspired me. Meanwhile, he has shown great guidance, patience, and encouragement towards me. His amiable personality and fairness makes the research a pleasant job.

I am grateful to my committee members, Drs. R. S. Houk, P. A. Thiel, W. S. Trahanovsky, W. C. Black and D. C. Johnson for their precious time and advice through the study program.

My sincere thanks go to Drs. Ho-Ming Pang, Hongdong Tan and Nanyan Zhang. I appreciate the chance to cooperate with Dr. Pang. I learned quite a lot from our discussion. Drs. Tan and Zhang have given me valuable suggestions and encouragement on my experiments.

I also want to thank all the Yeung's group members, both previous and present. Qingbo, Eliza, Changsheng, Jan, Michael, Yonghua, Xiaoyi, Yeongsong, Smiley, Nancy, Huisu have provided me their help in all aspects. I would also like to thank Ms. Vickie Johnson, our secretary, for editing my papers' formats for publication.

I dedicate this dissertation to my wife, my parents and parents-in-law. Their endless love and support made this work possible. My wife, Lei Liu, has gone through the happiness and hardship with me and comforted me when I was stressed over the years. My mother and mother-in-law have taken the duty of taking care of my little daughter so that I can concentrate on my research.

This work was performed at Ames Laboratory under contract no. W-7405-Eng-82 with the U. S. Department of Energy.