2003

High-throughput genetic analysis and combinatorial chiral separations based on capillary electrophoresis

Wenwan Zhong
Iowa State University

High-throughput genetic analysis and combinatorial chiral separations based on capillary electrophoresis

by

Wenwan Zhong

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Analytical Chemistry

Program of Study Committee:
Edward S. Yeung, Major Professor
Donald C. Beitz
Robert S. Houk
Marc Porter
Keith Woo

Iowa State University

Ames, Iowa

2003

Graduate College

Iowa State University

This is to certify that the doctoral dissertation of

Wenwan Zhong

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

To my dear families

TABLE OF CONTENTS

ABSTRACT

Capillary electrophoresis (CE) offers many advantages over conventional analytical methods, such as speed, simplicity, high resolution, low cost, and small sample consumption, especially for the separation of enantiomers. However, chiral method developments still can be time consuming and tedious. We designed a comprehensive enantioseparation protocol employing neutral and sulfated cyclodextrins as chiral selectors for common basic, neutral, and acidic compounds with a 96-capillary array system. By using only four judiciously chosen separation buffers, successful enantioseparations were achieved for 49 out of 54 test compounds spanning a large variety of pKs and structures. Therefore, unknown compounds can be screened in this manner to identify optimal enantioselective conditions in just one run. In addition to superior separation efficiency for small molecules, CE is also the most powerful technique for DNA separations. Using the same multiplexed capillary system with UV absorption detection, the sequence of a short DNA template can be acquired without any dye-labels. Two internal standards were utilized to adjust the migration time variations among capillaries, so that the four electropherograms for the A, T, C, G Sanger reactions can be aligned and base calling can be completed with a high level of confidence. The CE separation of DNA can be applied to study differential gene expression as well. Combined with pattern recognition techniques, small variations among electropherograms obtained by the separation of cDNA fragments produced from the total RNA samples of different human tissues can be revealed. These variations reflect the differences in total RNA expression among tissues. Thus, this CE-based approach can serve as an alternative to the DNA array techniques in gene expression analysis.

CHAPTER 1. GENERAL INTRODUCTION

Dissertation organization

Starting with a general introduction about capillary electrophoresis and related gene analysis techniques, this dissertation continues with three complete scientific manuscripts in the following chapters. A final chapter summarizes the work and provides some prospects for future research.

Capillary Electrophoresis

Since its debut in early 1980s, capillary electrophoresis (CE) has become a more and more essential separation technique in academic research and industrial applications. Attracted by its unique characters, such as extremely fast separation speed, extraordinary separation efficiency and resolution, low sample and buffer consumption, simplicity of instrumentation, and feasibility in automation, research scientists have devoted enormous efforts to broaden the horizon of CE for the last two decades.

A History of CE

The pioneering studies on the theories for the mechanism of electrophoresis, the role of electroosmosis, and mobility of ions in an electric field were started in the 19th century. In 1930, Tiselius carried out electrophoresis experiments on macromolecules --- proteins --- in free solution, which was the first practical demonstration of the use of electrophoresis as a

separation technique.[1] However, further utilization of this technique was hindered by severe band broadening resulted from electrical heating. Fast circulation of water maintained at 4° C was not enough for efficient cooling. Supporting media such as paper,[2] starch,[3] agarose,[4] and polyacrylamide[5] were employed to restrict diffusion of bands and at the same time to provide molecular sieving effect to separate molecules based on their sizes. However, the thermal effect was still not significantly reduced. A big contribution was made by Hjerten[6] in 1967, who recognized that carrying out electrophoresis in narrow diameter tubes can diminish the Joule heating. A capillary of 300 μm internal diameter was employed and ultraviolet absorbance detection was used for separation. Following his idea, some progresses were made over a short period of time, such as using polyacrylamide gel-filled glass tubes for protein analysis[7] and using smaller diameter Pyrex tubing (i.d. 200 to 500 μm)[8] or Teflon tubing (i.d. 200 μm)[9] for free-zone electrophoresis applied to separation of both inorganic and organic molecules. An even more important breakthrough happened in 1981. Jorgenson and Lukacs first demonstrated that the use of even narrower capillaries (i.d.< 100 μm) produced highly efficient electrophoretic separations in which 4×10^5 theoretical plates were achieved.[10,11] The large surface-to-volume ratio of the capillary allowed adequate heat dissipation so that voltages as high as 30 kV could be used. The work of Jorgenson and Lukacs initiated the rapid development of CE techniques in the last two decades.

Basics about CE

Capillary electrophoresis separates compounds in an open tubular capillary under an electric field. The experimental set-up for CE is simple. A schematic diagram of a typical CE system is given in Figure 1. The typical i.d. of the capillary column is from 10 μm to 75 μm.

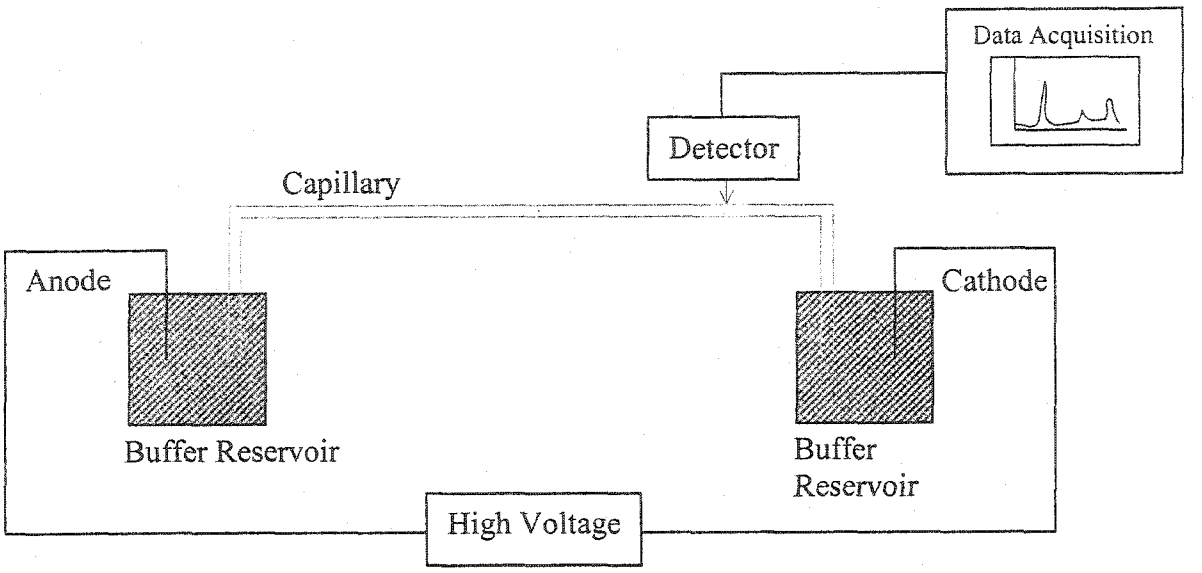


Fig 1. A schematic diagram of CE system.

Each end of the capillary is dipped in a buffer container with an electrode, which is connected to one output of a high-voltage power supply. Up to 8 kv/cm electric field can be applied. Close to the end of the capillary, the polyimide coating is removed to create an optical window for optical detection. Normal detection methods in CE include ultraviolet (UV) absorption, laser-induced-fluorescence (LIF), electrochemical detection, and mass spectrometry (MS). For proteomic materials, MS becomes a more important detection method, which allows chemical structure determination and/or identification of proteins after separation. The entire operation, from capillary conditioning, sample analysis, data acquisition and processing, can take less than 10 minutes depending on the length of the capillary, and can be easily automated, as is the case for most commercial CE instruments.

The major driving force in CE is the electroosmosis flow (EOF). The inner surface of the fused-silica capillary is covered with negatively charged silanol groups, which attract the cationic species in the buffer solution to form a double-layer at the silica/water interface. The double-layer has one static layer (stern layer) and one mobile layer (Helmholtz layer). The mobile layer contains a radial decreasing positive charge density, and the cations would move to the cathode end carrying the water of hydration with them. Due to the cohesive nature of the H-bonding of the hydrated water, the whole buffer solution would be dragged to the cathode to form electroosmosis flow. Different from the parabolic flow profile in the pressure driving system, the EOF has a flat profile, which reduces the radial diffusion of the sample plug, and thus improves the separation resolution. The magnitude of the EOF can be calculated from the dielectric constant, viscosity of the solution, and the zeta potential of the mobile layer. So that modifying the property of the inner capillary surface, or changing the buffer parameters, such as ionic strength, pH, and composition, can tailor the EOF to the

need of separation. Besides the movement under EOF, ionic compounds in the electric field have their own electrophoretic (EP) mobilities. Electrophoretic mobility of an ion is proportional to its charge to size ratio and inversely proportion to the viscosity of the buffer. That different ions have different electrophoretic mobility is the major mechanism behind CE separation. Thus, any molecule in the CE buffer moves at a velocity that is the vector sum of the electrophoretic velocity and the velocity imparted by the electroosmotic flow. The presence of EOF allows the separation of both negatively and positively charged species in the same run. And EP separates compounds relying on their different charge to size ratios. That is why CE is orthogonal to other separation techniques, in particular reverse-phase high performance liquid chromatography (HPLC), in which the separation of molecules is mostly based on their absorption or distribution in the stationary phase.

Currently, there are five modes of CE, capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC), capillary gel electrophoresis (CGE), capillary isoelectric focusing (CIEF), and capillary isotacophoresis (CITP). Due to the rapid development of these various operational modes, the versatility and range of applications of CE are improved, and CE has become a more and more mature analytical technique providing practical and innovative solutions to challenging separation problems.

CZE

Free-zone capillary electrophoresis is the most common and the most basic CE mode. The separation matrix is free buffer solution only. Sample ions are separated into discrete zones based on their different apparent mobilities. Fine tuning the buffer conditions

(e.g., buffer composition, ionic strength, pH value, viscosity) and running conditions (e.g., electric field, column dimension and length, injection) is the key to achieving efficient separations in CE. Jorgenson and Lukacs studied the separation efficiency and resolution between two adjacent eluting species in their first demonstration of CE.[9] A term used in chromatography, a number of theoretical plates (N), can be employed to describe the efficiency of separation in CE. They found that N is proportional to the apparent mobility and applied voltage but inversely proportional to the diffusion coefficient of the molecule. So that higher electric field not only speeds up the separation but also increases separation efficiency. The Joule heating effect caused by high running current can be minimized by reducing the radius of the separation column and lowering the ionic strength of running buffer. Both of them decrease the current flowing in the system. Also, the smaller the i.d. of the capillary, the more efficient the heat dissipation of the column. But there are limitations on how fast we can perform the separation. Because the resolution between two adjacent eluting species is inversely proportional to the square root of their average total mobility. Thus, maintaining the same level of resolution should be considered while choosing the separation voltage.

Other parameters of CZE that significantly affect separation efficiency and resolution include buffer pH, ionic strength, control of EOF, and size of injection. Most of the analytes for CE contain either acidic or basic groups. Optimizing the pH value of the buffer solution helps to change their dissociation situations, which may result in increasing the mobility difference between two species. Modification of pH value of the separation buffer is the first step in optimization of CZE separation. Also the ionic strength of the buffer would affect the EOF. As the buffer pH goes up from 2 to 12 at constant ionic strength, the EOF of silica

capillary first increases gradually, and then reaches a plateau after pH 8.[12] Higher ionic strength lowers the zeta potential of the double-layer, and decreases EOF.[13] And some organic solvents, like methanol, increase the pK values of the surface silanol groups and produce a corresponding decrease in the electroosmosis velocity when added to the running buffer.[14] Since the resolution is inversely proportional to the total velocity of the molecules, a smaller EOF could be beneficial to the separation. In some cases, the capillary wall coating material, such as cetyltrimethyl-ammonium bromide (CTAB[15]) and polyethylene glycol, can be used to control quantity and direction of EOF and prevent absorption of negatively charged species, like proteins at pH values higher than their pI, onto the wall.[16] The length of injection plug and detection zone are also important parameters affecting separation efficiency. Zare and coworkers found out that only when the length of the injection zone occupied only a small portion of the total length of the capillary would the high efficiency be achieved.[17] Usually only a few nanoliters of sample are injected hydrodynamically or electrokinetically.

CZE has been applied to the separation of a broad range of analytes, including organic acids,[18,19] inorganic ions,[20,21] amino acids,[22,23] and carbohydrates.[24] While difference in mobility is not enough to distinguish slightly different molecules, complexing agents are needed in the running buffer. For example, addition of chiral molecules like cyclodextrins and their derivatives,[25,26] crown ethers,[27,28] and macrocyclic antibodies[29,30] are critical in the separation of enantiomers.

MEKC

Based on the above discussion, we know that free-zone electrophoresis is only good for charged solutes. So, in order to expand the application range of CE to neutral molecules, Terabe and co-workers introduced a new mode of CE, micellar electrokinetic chromatography (MEKC) in 1984.[31] It allows the separation of neutral and charged solutes simultaneously by adding ionic surfactant to the CZE running buffer at a value above its critical micelle concentration (CMC). Surfactant monomers form aggregates, called micelles, at this value. These micelles act as pseudo-stationary phase, so that separation can be achieved based on the differential partitioning of analytes between the micelle phase and the aqueous phase. Because movement of solutes in MEKC is also driven by electric field for charged solutes, their separations are the combinations of chromatographic and electrophoretic separation mechanisms. This results in a difference in the rate of migration of cationic, neutral, and anionic analytes that does not normally exist in conventional chromatographic techniques such as HPLC and gas chromatography (GC) or could not be satisfactorily separated by CZE in which the separation is solely based on the differences in electrophoretic mobility. Also MEKC has advantages over HPLC with regard to the efficiency of the separation system, separation speed, cost, and tolerance to matrix constituents. No wonder MEKC has become more and more popular and attracted so much attention since its introduction. However, compared to free-zone electrophoresis, MEKC has extra band broadening, which impacts its efficiency. It is the resistance to mass transfer from aqueous phase to the pseudo-stationary phase and vice versa. It is due to the different velocities for the same molecule before and after its partitioning into or adsorption onto the micelles. So, the partition or adsorption kinetics should be fast to prevent severe band

broadening. There are additional band broadening sources in MEKC, including polydispersity of micelles and relatively high conductance of the ionic micellar running buffers. Theoretical studies on the band broadening effects and improvement of resolution in MEKC have been reported.[32,33,34] The most important issue in the optimization of MEKC separation conditions is the selection of surfactant or mixed surfactant system. The commonly used and widely studied surfactant systems are SDS, polyethyleneglycol ethers, tetra-alkyl-ammonium salts, bile salts, and fatty acids. MEKC has also inspired the use of other types of pseudo-stationary phases that aid in the separation of solutes, including those formed from cyclodextrins (CD-MEKC),[35] a variety of polymers,[36] microemulsions,[37] and suspensions.[38] As a separation technique for charged and non-charged species, MEKC has become a very successful CE mode and has been applied to a wide range of disciplines, such as analysis of drugs,[39] food and beverages,[40] forensic analysis,[41] environmental studies,[42] and bioanalytical separations.[43] The analytes for MEKC include proteins, amino acids, amines, phenolic compounds, carboxylic acids, and so on.

CGE

CE is very powerful in the separation of ions and even neutral molecules. But when it comes to the separation of biopolymers, such as DNA, RNA, and SDS-denatured proteins, which have constant charge to size ratios, free buffers with or without additives are ineffectual. A sieving matrix that can separate those biopolymers based on their sizes should be employed. This mode of CE is called capillary gel electrophoresis(CGE). The most successful application of it is on the separation of DNA and RNA molecules. Agarose or crosslinked polyacrylamide were the first sieving matrixes for CGE.[44,45] Polymerization took

place inside the capillary. But introduction of air bubbles and impurities during injection and polymerization of gels, degradation of gels under alkaline pH normally used to separate biopolymers, and clogging problems prevents the extensive usage of this kind of gels. Instead, replaceable linear polymer solutions were developed to take the place of gels in CGE. Most of the polymers used are linear polyacrylamide,[46] hydroxypropyl methyl cellulose (HPMC),[47] polyvinylpyrrolidone (PVP),[48] poly(ethylene oxide) (PEO)[49] and so forth. Recently, a new kind of sieving matrix, monomeric non-ionic surfactant solutions, was developed in our group.[50] This kind of surfactant forms self-assembly long chain micelles in solution, which behave like dynamic polymers. The use of replaceable sieving matrix make it possible to reuse the capillary for many runs and still maintain the same separation efficiency. The crosslinked-gel or semi-dilute polymer solutions are characterized by pores formed from a network of covalently bound polymers or dynamically entangled polymer chains. DNA molecules reptate inside the network like a snake seeking its way out in thick grass, crawling through the "tubes" formed by the section of polymer chains surrounding them. Larger fragments meet more obstacles when traveling through the "pores" and they are eluted slower than are the shorter fragments. This migration model is called Bias Reptation Model (BRM), capturing the essential migration features of DNA in polymer matrices.[51,52] Based on this model, the mobility of DNA is inversely proportional to the DNA size but would reach plateau mobility above critical size. This model also fits to other biopolymers, like RNA and SDS-denatured proteins. Since "pore size" and polymer chain properties play important roles in the size-dependent separation of biopolymers, optimization of CGE methods boils down to the choice of the appropriate polymers. We need to choose a polymer that is far above its entanglement threshold at the concentration needed to achieve the

appropriate mesh size. At the same time, we still have to keep viscosity low to ease the refilling of polymer solution in the capillary. Strong entanglement and low viscosity are contradictory requirements that are only met by few polymers. That is why viscosity-adjustable matrices are highly favorable. The continuous discovery of suitable polymers and rapid development of high throughput capillary array systems resulted in the splendid accomplishment of the Human Genome Project, which sequenced the whole genomes of human and other important organisms two years ahead of planned.

CITP and CIEF

The other two modes of CE, capillary isotachphoresis and capillary isoelectric focusing, have been least used in recent years. The most important applications of these two techniques are on the on-column sample preconcentration coupled to other CE modes. CIEF can also be used for rapid determination of the isoelectric point of a protein.

Capillary Array Electrophoresis (CAE)

If CE can only be operated individually, the sample throughput is very low and CE may not be this popular any more. It is the small dimension of the CE column and the simplicity of the CE set-up that make it relatively easier than the other separation techniques, such as HPLC, to run in a multiplexed manner to improve the sample throughput tens or even hundreds of times more. Dozens or hundreds of capillaries are bundled together and serve as individual separation channels. Only one detector is employed to monitor all the columns. So, the most challenging part to design in a CAE instrument is its detection system that can the most challenging part to design a CAE instrument is its detection system that can

illuminate all the capillaries and collect signals from each of them. Since laser-induced-fluorescence (LIF) is the most sensitive detection method for CE, most of the commercial CAE instruments are designed for LIF detection. Zugurski and McCormick presented the first report of capillary array electrophoresis (CAE) in 1990.[53] A detector scanned across an array of capillaries, recording the fluorescence signal sequentially from each capillary. This scanning detection method was further developed by Mathies *et al.*, and a commercial CAE instrument was constructed based on their design by Amersham Bioscience (Piscataway, NJ).[54] The scanning confocal detection system of this instrument consists of a microscope objective to focus the laser light inside the capillaries and at the same time collect the emitted light from the center of the column. The confocal system utilizes a set of four photomultiplier tubes with proper filters and dichroic beam splitters. However, this scanning method suffers from a limited duty cycle, and fast sampling rates were not achieved. To evenly distribute the light among all capillaries and continuous monitor each capillary, a post-column detection with liquid sheath flow was developed independently by two groups, Kambara *et al.* and Dovichi *et al.*[55,56] Briefly, the capillary bundle is aligned inside a quartz cuvette. Along the dead space between the columns and the walls of the cuvette, a buffer solution is pumped through the cell. The liquid sheath flow outside the capillary drags down the electrophoretic bands eluting from the column, tapering them to a smaller diameter. A laser beam crosses all flow streams and excites the fluorescent molecules. The fluorescent light is collected at 90 degrees from the plane of laser by being dispersed through a diffraction grating and imaged onto a cooled charge-couple device (CCD). This is the most successful commercial CAE instrument constructed by Applied Biosystem (Foster City, CA). Another on-column detection scheme was developed in our group, and the commercial instrument is built by

SpectruMedix (State College, Pennsylvania).[57,58] Ninety-six capillaries are immersed in a refractive index-matching liquid between glass plate, so that the laser beam crossing the capillary array is confined and energy loss of laser is minimized. The fluorescent light is collected at a right angle from the laser axis to reduce scattered light and detected by a CCD camera.

Other than these commercial capillary arrays, many different versions of arrays were custom-built by different research groups to further improve the illumination and detection. Kambara's group studied the side illumination with detection on column to make the replacement of polymer solution from both ends of capillaries easier[59] and advanced the lens effect approach by intercalating each capillary with a glass rod of the same external diameter as the capillaries.[60] Another elegant approach for a multiple capillary instrument was the use of optical fibers for illumination and collection of the signal in a 90 degree arrangement.[61] To break the actual limit of 96 capillaries, Mathies' group developed a confocal system with the capillaries aligned in a circular array.[62] The microscope objective spins inside a drum, interrogating the columns one by one. A larger number of capillaries, about 1000 capillaries, were easily accommodated in this geometry. In our group, a fast step-scanning 96-capillary system was built, which employs a single photomultiplier tube for fluorescence detection and a high-performance galvanometer scanner for sequential and independent illumination of the center of each capillary.[63] Step scanning maximizes the duty cycle and guarantees that the capillary walls are not irradiated.

Even though LIF detection has very high sensitivity, the sophisticated design of the detection scheme along with the expensive laser, detector and optical components increases the instrumental and operational cost of the CAE system. Also, applications of this kind of

instrument are limited to fluorescent analytes. A CAE instrument with UV absorption detection was developed in our group and has been commercialized recently.[64] The approach involves simultaneous illumination of all the capillaries by the light source, collection of transit lights after capillaries by a camera lens, and signal detection by a linear photodiode array. Since most of the molecules have absorption in the ultraviolet or visible range, this system has been applied to a wide range of applications, including combinatorial screening of enzyme activity,[65] genotyping,[66] comprehensive peptide mapping,[67]screening of homogeneous catalysis and reaction optimization,[68] DNA sequencing,[69] and multiplexed chiral separation.[70]

Microfluidic Chips for Capillary Electrophoresis

The CE technique can be further miniaturized due to its total elimination of mechanical pumps for liquid movement and no need of fixed stationary phases in the separation column. Microfluidics for integrated chemical processing and analysis under electrokinetically driven force had its official debut in the early 90s.[71] A great deal of research has focused on optimizing these devices for genetic and other analyse in aspects like fabrication, design and fluid manipulation. The manufacture of CE and CAE microchips is based on microfabrication techniques originally developed in the semiconductor industry. Microchannels and other features are defined on glass wafers using photolithography and etched by wet chemical techniques.[72] A main separation channel crosses the intersection of two side channels for sample introduction at one end. Electrical contact is made via wires immersed in reservoirs. Buffer loading, sample injection, separation and detection can be controlled through a computer program without any interpretation. While having the same

advantages belonging to CE techniques, such as high efficiency, high separation speed, low sample consumption, and ease for automation, microfabricated CE expands those effects contributing to the extremely short injection plugs, short separation distances and high electric fields. The first CE separation on a chip was demonstrated in 30μm channels etched into a glass slab with dimensions of 148 mm x 39 mm x 10 mm.[73] Fluorescein and calcein were separated in 7 minutes generating up to 35,000 theoretical plates. It can be done even more rapidly nowadays. It was proved separation of a binary mixture with baseline resolution in a 0.9 mm separation length within 150 ms with a field strength of 1.5kV/cm.[74]

More importantly, microfabrication can integrate sample preparation, all basic fluidic manipulations from aliquoting to dilution to addition and mixing of reagents, and separation on one single chip, and can arrange multiple channels in parallel on the same chip to improve sample throughput. Reaction chambers can be included on the microchip for sample separation in genetic analysis. For example, eight 280 nL PCR chambers interfaced directly with the CE channels were integrated on a chip.[75] Single-molecule of DNA amplification was amplified and analyzed in less than 20 min with results similar to those obtained by conventional capillary gel electrophoresis. An integrated sample preparation system comprised of microfluidic valves and vents, nanoliter PCR amplification chambers, and CE separation channels, all integrated in a glass sandwich structure has been developed in Mathies' laboratory.[76] From the same group, a high-throughput 96-channel DNA sequencing microfabricated CAE chip was constructed and performance of DNA sequencing was proved with average read length of 430 bp in less than 25 min.[77]

The optimized whole implementation of microfabricated devices will provide the next generation of electrophoretic technology. As with CE, microchips can be applied to the

analysis of drugs, small molecules and ions, proteins and peptides, and DNA. The integration of sample preparation onto the microchip increases the utility of microchip and will allow the true benefits of miniaturization, including smaller reagent volumes and costs, to be fully realized. More channels are being incorporated on one chip to obtain higher throughput.

Related DNA Techniques

CE plays a very important role in microscale biomolecule analysis, and the majority of CE work of the last decade is on DNA study. DNA is made up of four nucleotides that are repeated millions or billions of times in an organism. These nucleotides are similar in structure with a furanose bridge linking a phosphate group and a base. The bases in the four nucleotides are adenine, thymine, cytosine, and guanine. They are abbreviated using the first letters of the names of the bases, A, T, C, and G. A single DNA strand contains a sugar-phosphate backbone with the bases as side chains. In nature, DNA molecule consists of two complementary strands that are wound together through H-bonds between bases of form a double helix structure. The pairing between bases is highly specific, always happening between C – G and A – T. The particular order of the bases in DNA encodes the instruction on how to make protein. Proteins are the major components and functional molecules in cells. A protein's chemistry and behavior are fundamentally specified by the gene sequence. So, learning about DNA is the first step in understanding the secret of life. DNA studied can lead to insight understanding of difference among species, new knowledge about human biology, new ways to diagnose, treat, and someday prevent the thousands of disorders that affect us, solutions to challenges in health care, energy sources, agriculture, and

environmental cleanup. The most basic knowledge about DNA should be its base sequence, the genetic codes for making proteins. The Human Genome Project completed the sequence of the whole human genome which contains about 3.3 billion base pairs and results in an unprecedented understanding of the basic biochemical processes of living organisms.[78] This human genome sequence will underpin human biology and medicine in the next century, providing a single, essential reference to all genetic information.

Sanger's method and DNA sequencing

There are four best-known DNA sequencing techniques. They are: the Sanger enzymetic method, the Maxam and Gilbert chemical method, the Pyrosequencing method – DNA sequencing in real time by the detection of released pyrophosphate, and single molecule sequencing with exonuclease.[79-82] Sanger's method is the most important and commonly used one. This method, which revolutionized the field of genomics in the last two decades, is also known as the chain termination method or the dideoxynucleotide method. It consists of a catalyzed enzymetic reaction that polymerizes the DNA fragments complementary to the template DNA of interest (unknown DNA). Briefly, a primer (short oligonucleotide with a sequence complementary to the template DNA) is annealed to a specific known region on the template DNA, and provides a starting point for DNA synthesis. In the presence of DNA polymerases, catalytic polymerization of deoxynucleoside triphosphate (dNTP) onto the DNA occurs. The polymerization is extended until the enzyme incorporates a modified nucleoside, the terminator dideoxynucleoside triphosphate (ddNTP), which lacks the 3'-OH group necessary for further extension of the strand into the growing chain. This method is performed in four different tubes, each containing the appropriate

amount of one of the four terminators. All the generated fragments have the same 5'-end, whereas the residue at the 3'-end is determined by the ddNTP used in the reaction. In this way, random complementary strands with all possible stop points (i.e. all possible lengths with an integer number of nucleotides) are synthesized. The length of the fragment indicates the position of the 3'-end base on the strand complementary to the template. Usually the primers or the ddNTP are labeled for the identification of that last base of the fragments. These newly synthesized ssDNAs are mixed and then separated by gel electrophoresis based on their sizes. The sequence of the complementary DNA strand can be determined by the elution order of the synthesized fragments.[79] For a better understanding of the Sanger reaction, see Fig. 2.

Because of the extremely large size of the genomic DNA, ranging from 50 million to 3 billion base pairs, the whole genome DNA is broken down into fragments 2 or 3 kb long by enzyme digestion, prior to doing the chain-termination reaction. The fragments, inserted into a vector, are replicated in a bacterial culture. Several positive amplifications are selected, and the DNA is extensively sequenced. This random approach normally produces a high level of redundancy (the same segment is sequenced 6-10 times, in different reactions) and generates overlap in many regions.[83] A new variation of the method introduced by Venter *et al.* in 1996 is to shotgun a whole genome at once and depend enormously on computational resources to align all generated sequences.[84] This method has become well established and highly automated during its utilization in the Human Genome Project.

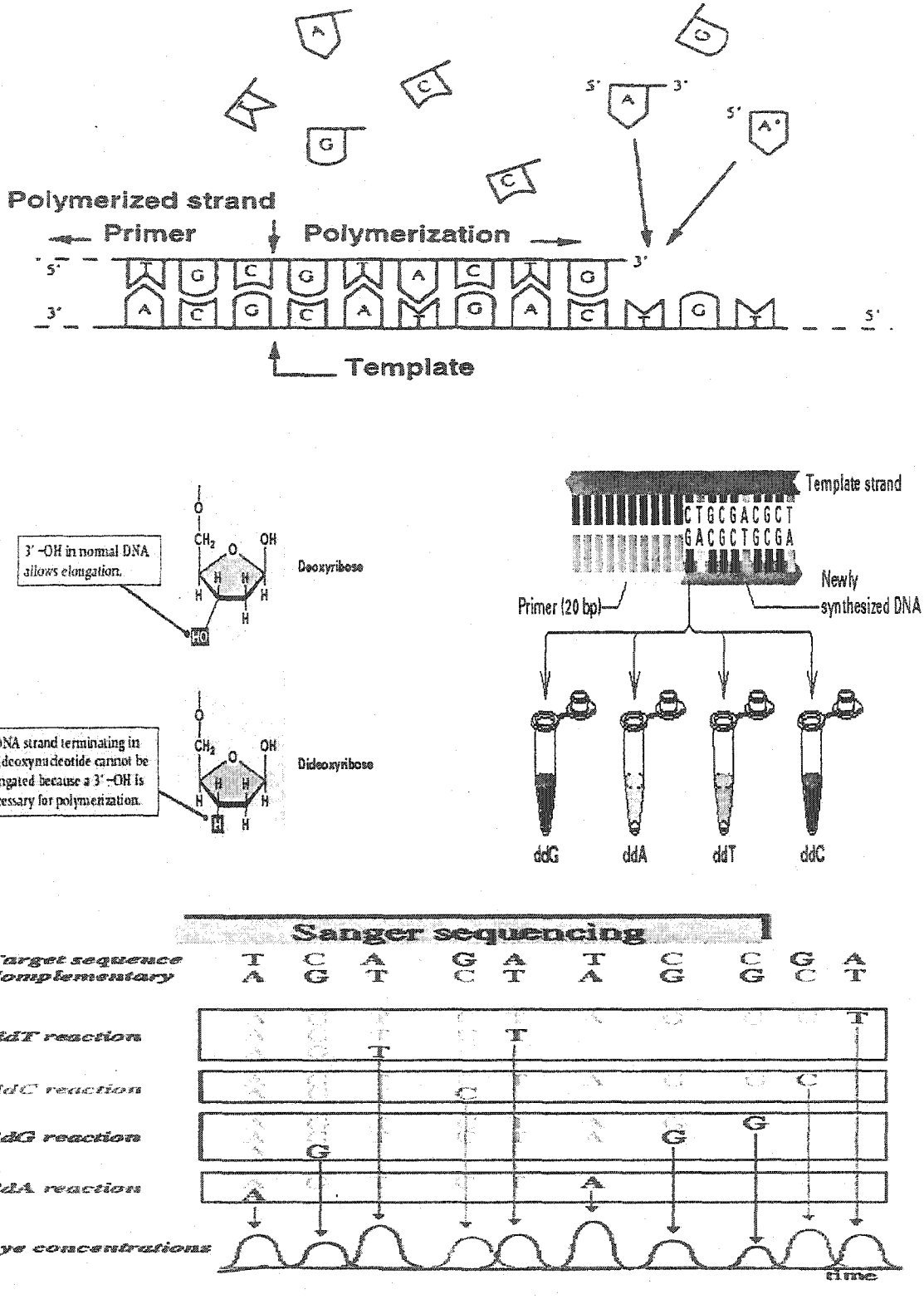


Fig. 2 Schematic representation of a sequencing process

Gene Expression Profiling Techniques

As our knowledge of gene sequences and sequence variation in populations increases, we will pinpoint more and more of the genes and proteins that are important in common, complex diseases. The expression status of genes in particular tissues can be monitored in parallel. Comparing corresponding gene expression changes under different circumstances can provide assistance in defining the functions of some gene products, establish the association of specific genes with disease phenotypes as well as help to establish the cause of the disease, and develop new therapy and new drugs. This kind of study is called functional genomics. Gene expression profiling is the most useful approach in this field. It examines the cell's use of sequence information to relate function and structure by examining the cellular RNA content. Many biotechniques were developed for this purpose.

The DNA microarray method developed by researchers at Stanford is the most powerful one. It allows the accumulation of large amounts of functional-genomic information in a parallel fashion.[85] This method is to use cDNA tags deposited onto a glass slide in known locations by high speed printing methods. Reverse transcription of mRNA from different cell populations produces cDNA that can be labeled with different fluorescent tags. These tagged cDNA fragments are hybridized to the cDNA on the chip, and differences in mRNA expression between the cell populations can be examined. The DNA chip assay allows true parallelism, miniaturization, multiplexing, and automation, and these key features provide a set of performance specifications that cannot be achieved with other techniques. Efforts have been made to reduce the cell population sizes required for analysis, to decrease the cost of the procedure, and to improve the precision of the method.[86-88] As these micro-

array methods become faster and more affordable, mRNA expression levels for cell populations exhibiting different phenotypes can be investigated more readily.

Serial analysis of gene expression (SAGE) is another high-throughput method very different from microarray technology. cDNA made from cellular mRNA is treated to create a single tag from each cDNA. The tag sequence (10-14 basepairs) can uniquely identify each transcript, and the concentration of each tag sequence is proportional to the level of mRNA in the original sample.[89] The tag sequences are ligated to long multimers, cloned, and sequenced. SAGE analysis is relatively fast, straightforward and cost-effective. But SAGE produces many redundant sequences as a consequence of its protocol, so several hundreds to several thousands of sequences need to be investigated to sample the whole cDNA population. A high level of initial mRNA is required for this method. For these reasons, SAGE is better suited for examinations of specific cell types or tissue with little cellular diversity.[90]

For organisms with little or no genomic information, microarray and SAGE are powerless, and differential display (DD) should be the technique of choice.[91] DD is a combination of three frequently used molecular biology techniques that allow one to visualize and compare gene expression patterns between two or more samples. It requires reverse transcription of polyadenlylated RNA, polymerase chain reaction (PCR) of the reverse-transcribed cDNA using special DD primers and polyacrylamide gel electrophoresis on the PCR products to get the expression fingerprint. DD is most useful when looking for low to medium-abundance, novel genes. However, a drawback of the DD protocol is that the analysis of large populations of mRNA with DD would require about 240 different primer combinations per sample to cover 95% of the genome.[92]

The success of these methods has increased our understanding of transcriptional regulation of the cell. Advances in these methods promise to reduce complexity and cost, allowing comparison of gene expression patterns for diverse cell populations on a nucleotide level.

Our Goal

Nowadays, analytical techniques are more and more important in helping people exploring the world and searching for the secret of life. Researchers keep working on the development of new techniques, hoping to find the most ideal methods. This is also the goal of our research: to find simpler, less expensive, faster and more universal approaches for DNA sequencing, chiral separation screening, and gene expression profiling relying on high efficiency and high throughput CE.

REFERENCES

1. Tiselius, A., *Nova Acta Regiae Societatis Scientiarum Upsaliensis*, **1930**, Ser IV, Vol. 7, Number 4.
2. Consden, R., Gordon, A.H., and Martin, A.J.P., *J. Biochem.*, **1944**, 38, 244.
3. Smithies, O., *J.Biochem.*, **1955**, 61, 629.
4. Hjerten, S., *Biochem. Biophys. Acta*, **1961**, 53, 514.
5. Raymond, S., and Weintraub, L., *Science*, **1959**, 130, 711.
6. Hjerten, S., *Chromtogr. Rev.*, **1967**, 9, 122.
7. Neuhoff, V., Schill, W.B., and Sternbach, H., *Biochem. J.*, **1970**, 117, 623

8. Virtanen, R., *Acta Polytechnica Scandinavia*, **1974**, Vol. 123

9. Millers, F.E.P., Everaerts, F.M., and Verheggen, T.P.E.M., *J. Chromatogr.*, **1979**,169, 11.

10. Jorgenson, J. and Lukacs, K.D., *Anal. Chem.*, **1981**, 53, 1298.

11. Jorgenson, J. and Lukacs, K.D., *Science*, **1983**, 222, 266.

12. Lukacs, K.D. and Jorgenson, J., *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **1985**, 8, 407.

13. Fujiwara, S. and Honda, S., *Anal. Chem.*, **1986**, 58, 1811.

14. Fujiwara, S. and Honda, S., *Anal. Chem.*, **1987**, 59, 487.

15. Altria, K.D. and Simpson, C.F., *Anal. Proc.*, **1986**, 23, 453.

16. Lauer, H.H. and McManigill, D., *Anal. Chem.*, **1986**, 587, 166.

17. Hugang, X., Coleman, W.F., and Zare, R.N., *J. Chromatogr.*, **1989**, 480, 95.

18. Jones, W.R., Jandik, P., *J. of Chromatogr.*, **1992**, 492, 608, 385.

19. Jones, W.R., Soglia, J., McGlynn, M., Haber, C., Reineck, J., Krstanovic, C., *Am. Lab.*, **1996**, 28, 25

20. Weston, A., Brown, P., Jandik, P., Jones, W., *J. Chromatogr.*, **1992**, 608, 395

21. Honda, S., Iwase, S., Makino, A., and Fujiwara, S., *Anal. Biochem.*, **1989**, 176 (1), 72.

22. Nickerson, B. and Jorgenson, J.W., *J. High Res. Chromatogr. Chromatogr. Comm.*, **1988**, 11, 533.

23. Nickerson, B. and Jorgenson, J.W., *J. High Res. Chromatogr. Chromatogr. Comm.*, **1988**, 11, 878.

24. Liu, J., Shirota, O., and Novotny, M., *Anal. Chem.*, **1991**, 63, 413.

25. Hempel, G. and Blaschke, G., *J. Chromatogr.*, **1996**, 675, 139.

26. Yoshinaga, M. and Tanaka, M., *J. Chromatogr.*, **1994**, 679, 359.

27. Schmid, M.G. and Gubitz, G., *J. Chromatogr.*, **1995**, 709, 81.

28. Walbroehl, Y. and Wagner, J., *J. Chromatogr.*, **1994**, 680, 253.

29. Armstrong, D.W., Rundlett, K., and Reid, G.L., *Anal. Chem.*, **1994**, 66, 1690.

30. Armstrong, D.W., Gasper, M.P., and Rundlett, K.L., *J. Chromatogr.*, **1995**, 689, 285.

31. Terabe, S., Otsuka, K., Ichikawa, K., Tsuchuya, A., and Ando, T., *Anal. Chem.*, **1984**, 56, 111.

32. Terabe, S., Otsuka, K., and Ando, T., *Anal. Chem.*, **1989**, 61, 251.

33. Little, E.L. and Foley, J.P., *J. Microcol. Spe.*, **1992**, 4, 145.

34. Rasmussen, H.T., Gobel, L.K., and McNair, H.M., *J. HRC & CC*, **1991**, 14, 25.

35. Shamsi, S.A., *Electrophoresis*, **2002**, 23(22-23), 4036.

36. Palmer, C.P., *Electrophoresis*, **2000**, 21, 4054.

37. Altria, K.D., *J. Chromatogr. A*, **2000**, 892(1+2), 171.

38. Bachmann, K., Gottlicer, B., Haag, I., Han, K.Y., Hensel, W., and Mainka, A., *J. Chromatogr. A*, **1994**, 688, 283.

39. Nishi, H., Terabe, S., *J. Chromatogr. A*, **1996**, 735, 3.

40. Boyce, M.C., *J. Chromatogr. A.*, **1999**, 847, 369.

41. Lurie, I., *J. Chromatogr. A*, **1997**, 780, 265.

42. Song, L., Xu, Z., Kang, J., Cheng, J., *J. Chromatogr. A*, **1997**, 780, 297.

43. Strege, M.A., Lagu, A.L., *J. Chromatogr. A*, **1997**, 780, 285.

44. Motsch, S.R., Kleemib, M.H., and Schomburg, G., *J. High Res. Chromatogr.* **1991**, 14, 629.

45. Cohen, A.S., Najarian, D.R., Paulus, A., Guttman, A. Smith, J.A., and Karger, B.L., *Proc. Natl. Acad. Sci. USA*, **1988**, 85, 9660.

46. Heiger, D.N., Cohen, A.S., and Karger, B.L., *J.Chromatogr.*, **1990**, 516, 33.

47. Baba, Y., Ishimaru, N.,Samata, K., and Tsuhako, M., *J. Chromatogr. A*, **1993**, 653, 329.

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

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

50. Wei, W., Xue, G., Yeung, E.S., *Anal. Chem.*, **2001**, 73(8), 1776.

51. Lumpkin, O.J., Dejardin, P., and Zimm, B.H., *Biopolymers*, **1985**, 24, 1573.

52. Viovy, J.L., *Rev. Mod. Phy.*, **2000**, 72 (3), 816

53. Zagursky, R.J., and McCormick, R.M., *Biotechniques*, **1990**, 9, 74.

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

55. Bay, S., Strake, H., Zhang, J.Z., Elliot, L.D., Dovichi, N.J., *J. Capil. Electrophor.*, **1994**, 1, 121.

56. Takahashi, S., Murakami, K., Anazawa, T., Kambara, H., *Anal. Chem.*, **1994**, 66, 2149.

57. Ueno, K., and Yeung, E.S., *Anal. Chem.*, **1994**, 66, 1424.

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

59. Anazawa, T., Takahashi, S., Kambara, H., *Anal. Chem.*, **1996**, 68, 2699.

60. Anazawa, T., Takahashi, S., Kambara, H., *Electrophoresis*, **1999**, 20, 539.

61. Quesada, M.A., Zhang, S.P., *Electrophoresis*, **1996**, 17, 1841.

62. Scherer, J.R., Kheterpal, I., Radhakrishnan, A., Ja, W.W., Mathies, R.A., *Electrophoresis*, **1999**, 20, 1508.

63. Xue, G., Yeung, E.S., *Electrophoresis*, **2001**, 22, 3490.

64. Gong, X., Yeung, E.S., *Anal. Chem.*, **1999**, 71, 4989.

65. Ma, L., Gong, X., Yeung, E.S., *Anal. Chem.*, **2000**, 3383.

66. Gong, X., Yeung, E.S., *J. Chromatogr.*, **2000**, 741, 15.

67. Kang, S.H., Gong, X., Yeung, E.S., *Anal. Chem.*, **2000**, 72, 3084

68. Zhang, Y., Gong, X., Zhang, H., Larock, R., Yeung, E.S., *J. Com. Chem.*, **2000**, 2, 450.

69. Zhong, W., Yeung, E.S., *J. Chromatogr.*, **2002**, 960(1-2), 229.

70. Zhong, W., Yeung, E.S., *Electrophoresis*, **2002**, 23(17), 2996.

71. Manz, A., Graber, N., Widmer, H.M., *Sens. Actuator B*, **1990**, 1, 244.

72. Manz, A., Harrison, D.J., Verpoorte, E.M.J., Fettinger, J.C., Paulus, A., Ludi, H., Widmer, H.M., *J. Chromatogr.*, **1992**, 593, 253.

73. Harrison, D.J., Manz, A., Fan, Z., Ludi, H., Widmer, H.M., *Anal. Chem.*, **1992**, 64, 1926.

74. Jacobson, S.C., Hergenroder, R., Koutny, L.B., Ramsey, J.M., *Anal. Chem.*, **1994**, 66, 1114.

75. Lagally, E.T., Medintz, I., Mathies, R.A., *Anal. Chem.*, **2001**, 73, 565.

76. Lagally, E.T., Simpson, P.C., Mathies, R.A., *Sens. Actuator B*, **2000**, 63, 138.

77. Paegel, B.M., Hutt, L.D., Simpson, P.C., Mathies, R.A., *Anal. Chem.*, **2000**, 72, 3030.

78. Collins et al., *Science*, **2003**, 300, 286.

79. Sanger, F., Nicklen, S., Coulson, A.R., *Proc. Natn. Acad. Sci. USA*, **1977**, 74, 5463.

80. Maxam, A.M., Gilbert, W., *Proc. Natn. Acad. Sci. USA*, **1977**, 74, 560.

81. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., Nyren, P., *Analyt. Biochem.*, **1996**, 242, 84.

82. Jett, J.H., Keller, R.A., Martin, J.C., Marrone, B.L., Moyzis, R.K., Ratlife, R.L., Seitzinger, N.K., Shera, E.B., Stewart, C.C., *J. Biomolec. Struct. Dyn.*, **1989**, 7, 301.

83. Adams, M.D., Fields, C., Venter, J.C., *Automatic DNA Sequencing and Analysis, San Diego: Academic Press*, **1996**.

84. Venter, J.C., Smith, H.O., Hood, L., *Nature*, **1996**, 381, 364.

85. DeRisi, J.L., Iyer, V.R., Brown, P.O., *Science,* **1997**, 278, 680.

86. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., Lockhart, D.J., *Nat. Biotechnol.*, **1997**, 15, 1359.

87. Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F., *Nat. Biotechnol.*, **1999**, 17, 974.

88. Mahadevappa, M., Warrington, J.A., *Nat. Biotechnol.*, **1999**, 17, 1134.

89. Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., Kinzler, K.W., *Science*, **1997**, 276,1268.

90. Stollberg, J., Urschitz, J., Urban, Z., and Boyd, C.D., *Genome Res.*, **2000**, 10, 1241.

91. Liang, P., and Pardee, A.B., *Science*, **1992**, 257, 967.

92. Laing, P., and Pardee, A.B., *Methods Mo. Gen.*, **1994**, 5, 3.

CHAPTER 2. MULTIPLEXED CAPILLARY ELECTROPHORESIS FOR DNA SEQUENCING WITH UV ABSORPTION DETECTION

A paper published in Journal Chromatography A

Wenwan Zhong and Edward S. Yeung

ABSTRACT

DNA sequencing is performed in a multiplexed capillary electrophoresis system by UV absorption detection. Four individual electropherograms are obtained by simultaneously running the unlabeled DNA products of the four ddNTP-terminated reactions in the capillary array. The sequence of the template used in the cycle-sequencing reaction can be determined by overlaying the four electropherograms. Two internal standards are employed to adjust for the variance in migration times among the capillaries. After applying the correction algorithm, base calling can be done at a high level of confidence.

---

*Reprinted from Journal of Chromatography A, 960, 229-239.

INTRODUCTION

Capillary gel electrophoresis (CGE) has become an important technique in DNA sequencing because of its high speed, high resolution, flexibility, and possibility of building integrated and automated system. In order to decipher the DNA sequence, radioactive or fluorescent labeling of the DNA fragments created by Sanger's chain termination reaction [1] is required for applying standard detection methods. Since autoradiography is labor intensive and can pose safety concerns, laser-induced fluorescence (LIF) has replaced it as the main detection method in DNA analysis. Extensive research work has been done to develop better detection schemes in fluorescence-based DNA sequencing, such as four-channel or two-channel detection using four dyes and single-color, intensity-based detection [2-5]. Wavelength selection in the emission or excitation spectra and fluorescence lifetime measurements in time domain or frequency domain are the major methods to discriminate among the different colored labels on the DNA fragments [6]. However, fluorescence detection has its drawbacks too. It requires expensive equipment and reagents. Furthermore, a fluorescence label affects the fractional size and charge of the DNA fragments, which results in mobility shifts among the labeled fragments in CGE [7]. Mobility shifts make data processing more complicated and lead to errors in base calling. Compared to LIF, UV absorption detection is not as sensitive, but the instrumental setup is much more simple and less expensive. It is easier to operate and maintain because of the use of a UV lamp rather than a laser system. No dye-label is required when using UV absorption detection in DNA analysis, since DNA has strong absorption at 254 nm. A 100-bp DNA has essentially 100 absorbers per fragment. Finally, mobility shift should not be a problem because no labels are present.

Because there are no dye labels, the DNA products from the four individual termination reactions must be run at four different capillaries in order to assemble the sequence. This is analogous to radioactive labeling and infrared single-label sequencing in slab-gel sequencing. The use of four separate capillaries at a time means that a multiplexed capillary system is the only way to achieve high-speed, high-efficiency, and high-throughput DNA sequencing. A novel absorption detection method has been applied to multiplexed capillary electrophoresis in our group [8]. The system has proven to be very reliable and efficient in many applications, such as screening of enzyme activity, peptide mapping of proteins and genetic typing [9-11]. Here, we will demonstrate that the same instrumentation can be used in DNA sequencing. Because detection is based on UV absorption, sieving matrixes like polyacrylamide [12] or poly(vinyl pyrrolidone) [13] cannot be employed. Instead, we use a new dynamic sieving matrix based on the self-assembly of monomeric surfactants into large aggregates under certain conditions [14].

MATERIAL AND METHODS

Chemicals

All chemicals for preparing running buffer solutions were purchased from Sigma (St. Louis, MO). The buffer contained 89 mM Tris, 50 mM TAPS, 20 mM histidine, 2 mM EDTA and 7 M urea in deionized water and was filtered with a .22 μm cellulose membrane filter from Corning (Corning, NY). The chemicals for cycle-sequencing buffer ($MgCl_2$ and Tris) were purchased from Fisher (Fair Lawn, New Jersey). Ten base pair DNA ladder was obtained from Life Technologies (Frederick, MD). The internal standards, 40-bp and 80-bp

fragments, and selected cycle-sequencing primers were prepared at the Nucleic Acid Facility (Iowa State University, Ames, IA). The 323-bp template was prepared using reagents in PCR Core System II from Promega (Madison, WI) and its Positive Control PCR Protocol. The PCR product was purified using QIAquick PCR Purification Kit from Qiagen (Valencia, CA). ThermoSequenase (32 U/μl), dNTPs (100 mM) and ddNTPs (10 mM) were obtained from USB/Amersham Life Sciences (Arlington Heights, IL).

Sequencing reaction

In order to generate enough quantities of sequencing fragments for UV detection, we used the cycle-sequencing protocol introduced by Cohen et al. [15]. The reaction mixture was combined in a microcentrifuge tube and put on ice: 200 pmol primer, 0.2 pmol template, 10 μl Tris pH 9 (250 mM), 10 μl $MgCl_2$ (50 mM), 10 μl dNTP mix (10 mM) and 32 U ThermoSequenase. Autoclaved and deionized water was added to obtain a total volume of 90 μl. Twenty microliter reaction mixture was added to each of the four 0.2 ml PCR reaction tubes (Molecular BioProducts, San Diego, CA) containing 1.25 μl of the appropriate ddNTP (1 mM). The samples were kept on ice before they were put onto the preheated block (95 °C) of GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). The total cycle number is 200. Each cycle contains three consecutive steps: 95 °C, 30 s; 52 °C, 30 s; 72 °C, 30 s. The product of cycle-sequencing reaction was purified by a spin column (Princeton Separation, Adelphia, NJ), and dried in vacuum. Before injection, the DNA samples were dissolved in 3 μl deionized water and transferred to a 96-well 0.2 ml micro-tube plate (Marsh Biomedical Products, Rochester, NY), spiked with 10 pmol internal standards (40-bp and 80-

bp DNA fragments). After heating the plate at 95 °C for 3 min for denaturing, the sample plate was put onto ice for injection.

DNA separation

Sieving matrix was prepared by dissolving polyoxyethylene-6-cetyl-ether (Sigma) in the running buffer while gently heating and stirring [14]. Then the low viscosity gel was forced into a 24-capillary array from the ground end. Before injection, the matrix-filled capillary array was pre-run for 5 min at 32 °C. Injection was performed at 2 kV for 2.5 min. During the run, the temperature was kept at 32 °C. The running voltage of 8.8 kV was applied by a power supply from Glassman High Voltage, Inc. (Whitehorse Station, NJ). After each run, the capillaries were regenerated by washing with 0.1 M hydrochloric acid for a few minutes, then rinsed with deionized water for half an hour.

Instrumentation

The basic setup of the multiplexed capillary electrophoresis system is similar to that described in ref. [8]. Thirty capillaries, 75-μm i.d. and 365-μm o.d., were packed side by side with 70-cm effective length and 90-cm total length. Fifty centimeter of each capillary was enclosed in a water circulation system (Fisher Scientific, Pittsburgh, PA) to control the running temperature. A 254-nm mercury lamp was used for UV absorption detection. The transmitted light from the capillary array passed through an interference filter (Oriel) and a quartz lens (Nikon; focal length = 105 mm; F = 4.5). A linear photodiode array detector (PDA) (Hamamatsu model S5964, Hamamatsu, Japan) was used to collect data, and a National Instrument PCI E Series multifunction 16-bit I/O board was employed to transfer data

to a computer (233-MHz Pentium, Packard Bell). The raw data sets were converted into single-diode electropherograms by a Labview program. Data treatment and analysis were performed using Microsoft Excel 97 and GRAMS/32 5.05 (Galactic Industries).

RESULTS AND DISCUSSION

Principles of normalization

Even though in a capillary array system, all the capillaries are run under exactly the same conditions (voltage, temperature, injection time, and buffer pH), the surface chemistry and geometry of the capillaries are different. Also, the gel matrix cannot be exactly the same after being pushed into the capillaries. These variations can cause substantial variations in the migration times of DNA fragments, which precludes calling bases by simply overlapping the four individual electropherograms. We already demonstrated that the use of two internal standards provides normalization of migration times in micellar electrokinetic chromatography (MEKC) and capillary zone electrophoresis (CZE) [16]. This normalization method should also be useful in unlabeled DNA sequencing. Without labeling dyes, there should be no mobility shift among the Sanger fragments. Also, since sufficient denaturant (7 M urea) has been included in the buffer, compressions in GC-rich regions are minimized. Non-uniform migration times among different capillaries are therefore only caused by the variations mentioned above. So, we used two DNA fragments of known lengths to adjust the migration times of each capillary for base calling. Another reason that makes the internal standardization method suitable for CGE is that the relationship between migration time and base number is linear over a narrow range [17]. In entangled polymer solutions, the best

model that describes DNA movement is "the biased reptation with fluctuations" (BRF) model [18]. According to this model, for small molecules (below a critical size), the mobility of the DNA fragment μ is inversely proportional to its size, represented by base number N ("reptation without orientation") [19]:

$$\mu/\mu_0 \sim 1/N \tag{1}$$

where μ_0 is the mobility in free solution. From the definition of electrophoretic mobility:

$$\mu = v/E = x/tE \tag{2}$$

where v is the average velocity, E is the externally applied electric field strength and x is the distance traveled in time t [17]. We can thus change Eq. (1) to:

$$t \sim N \tag{3}$$

Based on the migration times of the two internal standards in each capillary, linear equations of $t \sim N$ for the corresponding electropherograms can be determined. Then, we can adjust the migration times of the DNA fragments in every capillary using one capillary as the migration time standard. After the normalization process, we can call the sequence of the DNA template according to the order of the adjusted migration times. For example, there are two different equations for the electropherograms of ddATP and ddCTP termination reactions:

$$N_A = a_1 t_A + b_1 \tag{4}$$

and

$$N_C = a_2 t_C + b_2 \tag{5}$$

Using Eq. (4) as the standard equation, we can manipulate the terms in Eq. (5) to give the identical value for N_C:

$$N_C = a_1((a_2 t_C)/a_1 - (b_1 - b_2)/a_1) + b_1 \tag{6}$$

So the migration time of the C fragments can be normalized by:

$$t_C' = (a_2/a_1)t_C - (b_1 - b_2)/a_1 \quad (7)$$

Normalization of dsDNA migration times

First, we tested the normalization principle with dsDNA using 10-bp DNA ladder. The ladder was injected into selected capillaries without dilution. The raw data are shown in Fig. 1. Even though the samples were the same and all the capillaries were operated under the same conditions at the same time, the migration times of the same length fragment in different capillaries are different. Here, 4 or 5 capillaries were bundled together at the ground end to facilitate filling with gel with a 100-μl glass syringe. #2, #5, and #6 indicate the group of capillaries each numbered capillary belongs to. We can see that the migration times in the capillaries are very different no matter whether the capillaries are from the same bundle or not. Using the three largest peaks (10 bp, 100 bp and 330 bp) as internal standards, we can align these electropherograms (Fig. 2). Fig. 2A shows the result of using 10-bp and 100-bp fragments as standards to align the other peaks. After normalization, all the peaks of the same size fragments falling in the range of 10-bp to 100-bp have the same migration times. But for the peaks larger than 100-bp, the 330-bp must be used together with the 100-bp fragment as standards to achieve the proper result (Fig. 2B). This confirms that the linear relationship only fits in a narrow range.

DNA sequencing

The next experiment was for ssDNA, the Sanger fragments derived from cycle-sequencing reaction. The four chain-termination reactions created four sets of DNA frag-

ments, corresponding to the four bases in DNA, A, C, G, and T. In order to obtain enough signal for UV absorption detection, we dissolved the dried sample in deionized water to implement stacking in injection. A layer of silicone oil (Life Technologies, Rockville, Maryland) was put on top of the vials to avoid evaporation during heating. It is known that the efficiency of CGE separation of oligonucleotides dissolved in water or other low ionic strength solvents is affected by the injection field strength and duration [20]. To achieve high resolution, a low injection field and a long injection time should be utilized [21]. We found that 2 kV injection voltage and 2.5 min injection time worked best in our experiment. The longer injection time did not degrade the separation performance because of stacking. This is confirmed by examining the resolution among the small DNA fragments, which would have been affected the most by electrokinetic injection.

Fig. 3 shows the four electropherograms of A, C, G, T for pGEM DNA in four capillaries of the array. The known base number N was plotted as a function of migration time t (Fig. 4). In an individual capillary, the base number is proportional to the migration time t. However, we cannot obtain the right order of bases from Fig. 4 because of migration variations among capillaries. Some fragments can be off by 5 bp. Two internal standards, 40-bp and 80-bp DNA fragments, were co-injected with the DNA samples. Based on the migration times and base number of these internal standards, all the t ~ N equations of the capillaries were determined. After normalization of the migration times, the four electropherograms were aligned to call the sequence of the template (Fig. 5). Fig. 6 shows more clearly that after normalization, all the peaks are on the same line, and the sequence can be read directly from the data in these two figures (Table 1). In the worst case, there exists only a 0.5 bp error. The standard fragments added prevent base calling at those specific locations. However, stag-

gered sizes can be used to span a large normalization range and to recover any missing information.

The protocol we used compensates for the amount of product by using short template, high concentration of primer and high number of reaction cycles. So, the read length is short in this experiment. Especially for the A reaction, no peaks can be seen in the electropherogram after 89 bp. For the C, G, T reactions, even though the read length was up to 150 bp, we can only call the right sequence from 26 bp to 93 bp because of the limited useful range of two internal standards. Clearly, such short sequences will not allow the present scheme to compete with current instrumentation for genomic sequencing. However, for diagnosis [9] or antisense characterization [17], read lengths of 100 bp are adequate. The present scheme therefore offers an alternative to mass spectrometric analysis of short fragments [22]. The fact that capillary arrays are eventually scalable to 384 or even 1536 formats means that using 4 lanes at a time is not unreasonable. The unusually large number of amplification cycles implies a longer sample preparation time. However, that is always performed off-line in an automated system and has little effect on the throughput.

In addition to studying the four electropherograms for four different termination reactions, we also studied the electropherograms for a single termination reaction. Fig. 7 shows the result of overlaying two electropherograms of two G reactions after normalization. The G sample was run with the other four A, C, G, T samples (Fig. 6) in the same array. After migration time correction, the peaks for the same length fragments were aligned exactly. We thus confirmed that, after migration time correction, there is no mobility shift among the capillaries. We found some small ghost peaks, which only showed up in one of the electropherograms, presumably due to the loss of fidelity of the cycle-sequencing reaction.

CONCLUSION REMARKS

This experiment demonstrated that DNA sequencing could be performed in a multiplexed capillary electrophoresis system with UV absorption detection using internal standards. Because the products of four ddNTP reactions are run at four different capillaries, no additional information is needed to distinguish them in detection. Recently, a commercial version of the 96-capillary absorption instrument has become available [23]. It makes the detection scheme much simpler and more straightforward than autoradiography or LIF detection, and also lowers the operation cost. By using two internal standards, about 100 bp can be read. For longer reads, we expect that an additional one or two internal standards will suffice.

ACKNOWLEDGEMENT

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

REFERENCES

[1] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. (USA)* **1977**, 74,5463.

[2] K. Ueno, E. S. Yeung, *Anal. Chem.* **1994**, 66,1424.

[3] S. Takahashi, K. Murakami, T. Anazawa, H. Kambara, *Anal. Chem.* **1994**, 66,1021.

[4] R. A. Mathies, X. C. Huang, M. A. Quesada, *Anal. Chem.* **1992**, 64,2149.

[5] H. Kambara, S. Takahashi, *Nature* **1993**, 361,565.

[6] E. S. Yeung, Q. Li, in *High Performance Capillary Electrophoresis*; Khaledi, M. A., Ed., John Wiley & Sons, Inc., **1998**, Vol. 146, Chapter 22.

[7] H. Tan, E. S. Yeung, *Electrophoresis* **1997**,18, 2893.

[8] X. Gong, E. S. Yeung, *Anal. Chem.* **1999**,71, 4989.

[9] X. Gong, E. S. Yeung, *J. Chromatogr. B* **2000**,741,15.

[10] S. H. Kang, X. Gong, E. S. Yeung, *Anal. Chem.* **2000**,72,3014.

[11] L. Ma, X. Gong, E. S. Yeung, *Anal. Chem.* **2000**,72, 3383.

[12] S. Carson, A. S. Cohen, A. Belenkii, M. C. Ruiz-Martinez, J. Berka, B. L. Karger, *Anal. Chem.* **1993**,65,3219.

[13] Q. Gao, E. S. Yeung, *Anal. Chem.* **1998**,70,1382.

[14] W. Wei, E. S. Yeung, *Anal. Chem.* **2001**, 73,1776.

[15] D. Froim, C. E. Hopkins, A. Belenky, A. S. Cohen, *Nucl. Acids Res.* **1997**, 25, 4219.

[16] G. Xue, H.-M. Pang, E. S. Yeung, *Anal. Chem.* **1999**, 71,2642.

[17] A. Belenky, D. L. Smisek, A. S. Cohen, *J. Chromatogr. A* **1995**, 700,137.

[18] T. Duke, J. L. Viovy, A. N. Semenov, *Biopolymers* **1994**, 34, 239.

[19] C. Heller, *Electrophoresis* **1999**, 20,1962.

[20] D. Demorest, R. J. Dubrow, *J. Chromatogr.* 1991, 559,43.

[21] O. Salas-Solano, M. C. Ruiz-Martinez, E. Carrilho, L. Kotler, B. L. Karger, *Anal. Chem.* **1998**, 70,1528.

[22] Z. Fei, T. Ono, L. M. Smith, *Nucl. Acids Res.* **1998**, 26,2827.

[23] www.combisep.com

Table 1. Adjusted migration times (frame numbers), the corresponding base number, and the read sequence obtained from Fig. 6.

Base No.	A(t)	C(t')	G(t')	T(t')	Sequence
26				23044.8	T
27		23343.18			C
28				23443.01	T
29			23669.29		G
30		23696.49			C
31		23796.85			C
32	23957.9				A
33			24130.79		G
34				24131.26	T
35			24376.79		G
36				24455.01	T
37				24595.22	T
38	24804.5				A
39		24816.53			C
40	24963.9				A
41	25055.5				A
42		25060.04			C
43		25145.64			C
44	25288.5				A
45	25403.5				A
46				25412.33	T
47				25523.66	T
48	25704.5				A
49	25773.5				A

Table 1. (continued)

Base No.	A(t)	C(t')	G(t')	T(t')	Sequence
50		25808.86			C
51		25879.7			C
52	26009.5				A
53	26078.5				A
54				26132.64	T
55				26222.16	T
56		26305.72			C
57				26388.89	T
58			26489.61		G
59	26626.5				A
60				26690.83	T
61				26780.96	T
62	26864.9				A
63			26898.78		G
64	27040.5				A
65	27100				A
66	27160				A
67	27211.5				A
68	27275				A
69		27279.75			C
70				27393.22	T
71		27426.74			C
72	27526.9				A
73				27628.37	T
74		27650.57			C

Table 1. (continued)

Base No.	A(t)	C(t')	G(t')	T(t')	Sequence
75			27699.77		G
76			27765.24		G
77	27816.9				A
78		27865.05			C
79	27969.7				A
80				28082.4	T
81		28094.3			C
82	28169.4				A
83	28230				A
84	28295.9				A
85				28414.95	T
86			28423.98		G
87	28466.5				A
88	28502				A
89	28562.9				A
90		28579.84			C
91				28738.19	T
92			28756.18		G
93		28788.91			C

FIGURE CAPTIONS

Figure 1. Extracted UV electropherograms of 10-bp dsDNA ladder separation in the capillary array system. Capillaries No. 8, 10 are from bundle #2; No. 19, 21 are from bundle #5; No. 23, 26 are from bundle #6.

Figure 2. The 10-bp ladder electropherograms of Fig. 1 after migration time adjustment. Peaks of same size fragments in different capillaries show up at the same adjusted times. A. 10-bp and 100-bp fragments were used as internal standards in normalization. B. 100-bp and 330-bp fragments were used as internal standards in normalization.

Figure 3. Electropherograms of four individual sequencing reactions. The peaks with "*" on top are the internal standards at 40 bp and at 80 bp.

Figure 4. Migration times plotted as a function of base number from the original electropherograms in Fig. 3.

Figure 5. Overlay of the four individual UV electropherograms after normalization of Fig. 3 using 40-bp and 80-bp fragments as internal standards.

Figure 6. Adjusted migration times plotted as a function of base number.

Figure 7. Overlay of two G electropherograms. The peaks without matching peaks in the other electropherogram are ghost peaks.

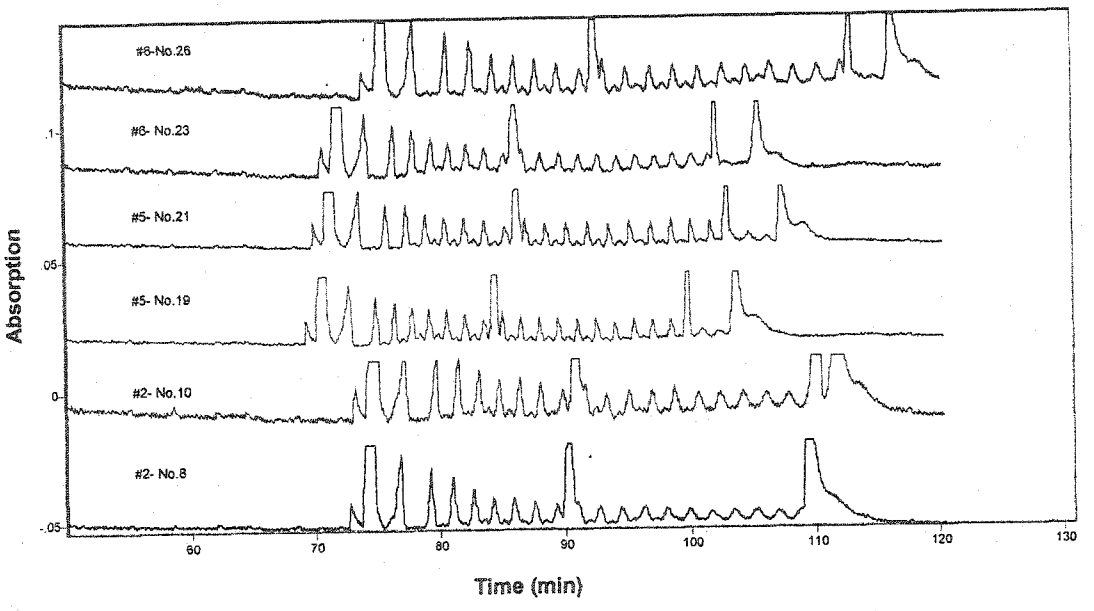

#6-No.26
#6- No.23
#5- No.21
#5- No.19
#2- No.10
#2- No.8
Absorption
.1
.05
0
-.05
60
70
80
90
100
110
120
130
Time (min)

FIGURE 1

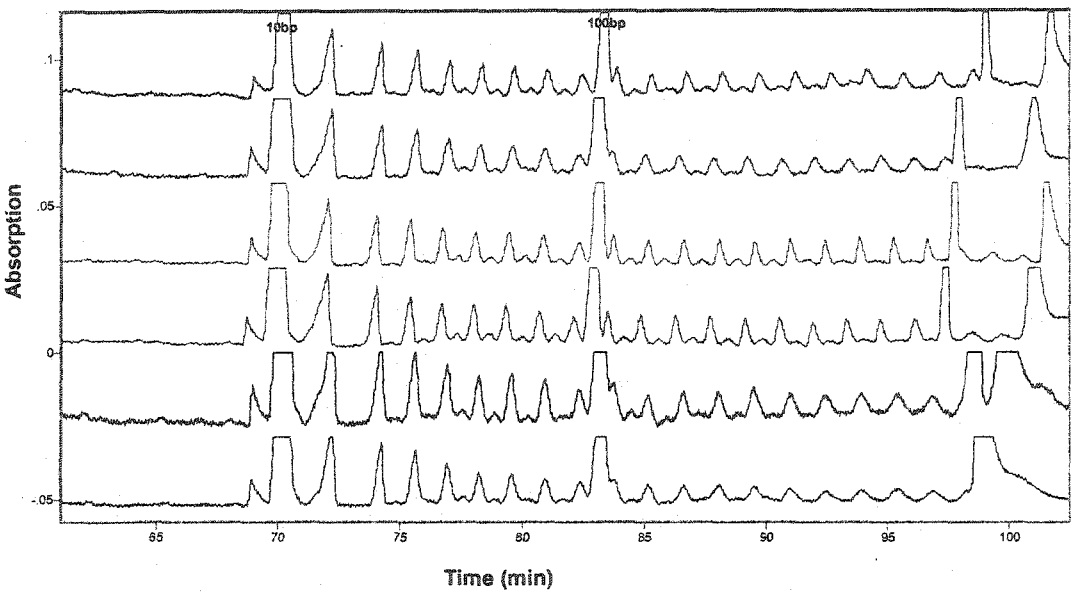


FIGURE 2A

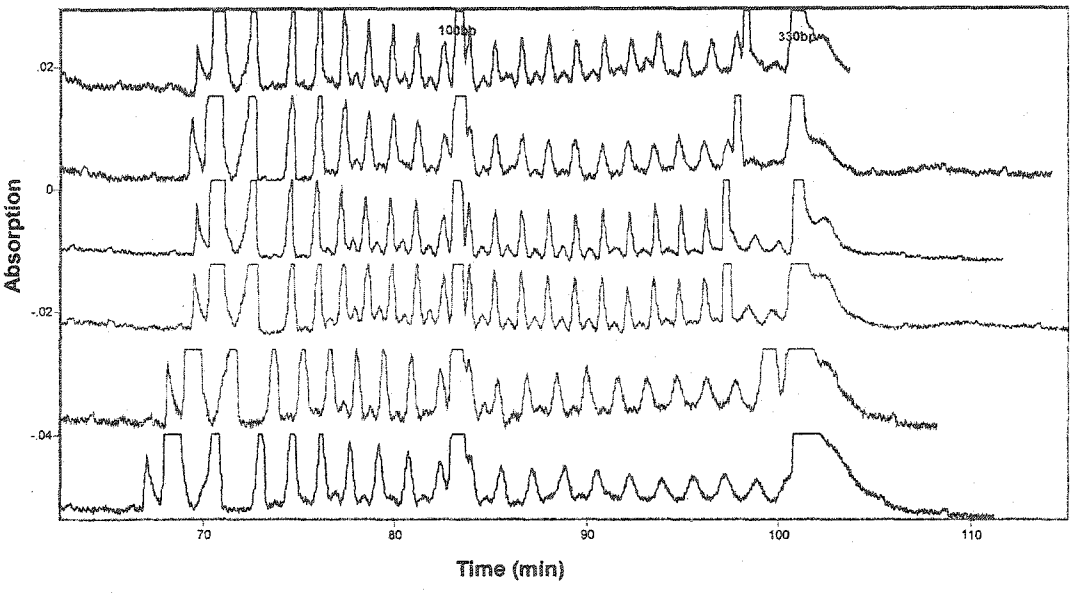


FIGURE 2B

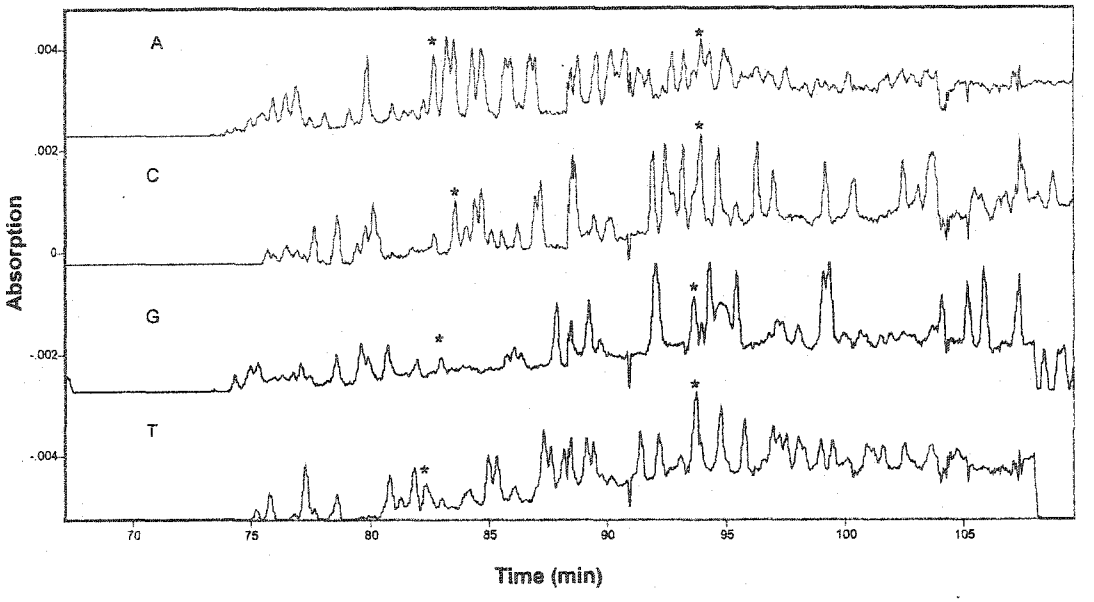

A
C
G
T
.004
.002
0
-.002
-.004
Absorption
70
75
80
85
90
95
100
105
Time (min)

FIGURE 3

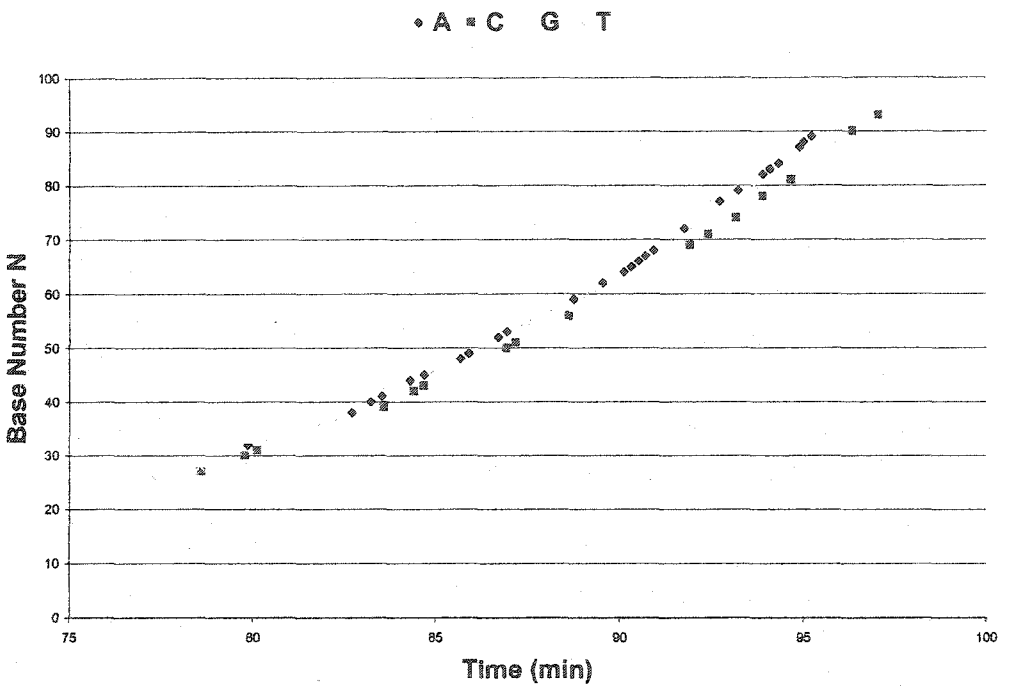

A C G T
Base Number N
Time (min)
0
10
20
30
40
50
60
70
80
90
100
75
80
85
90
95
100

FIGURE 4

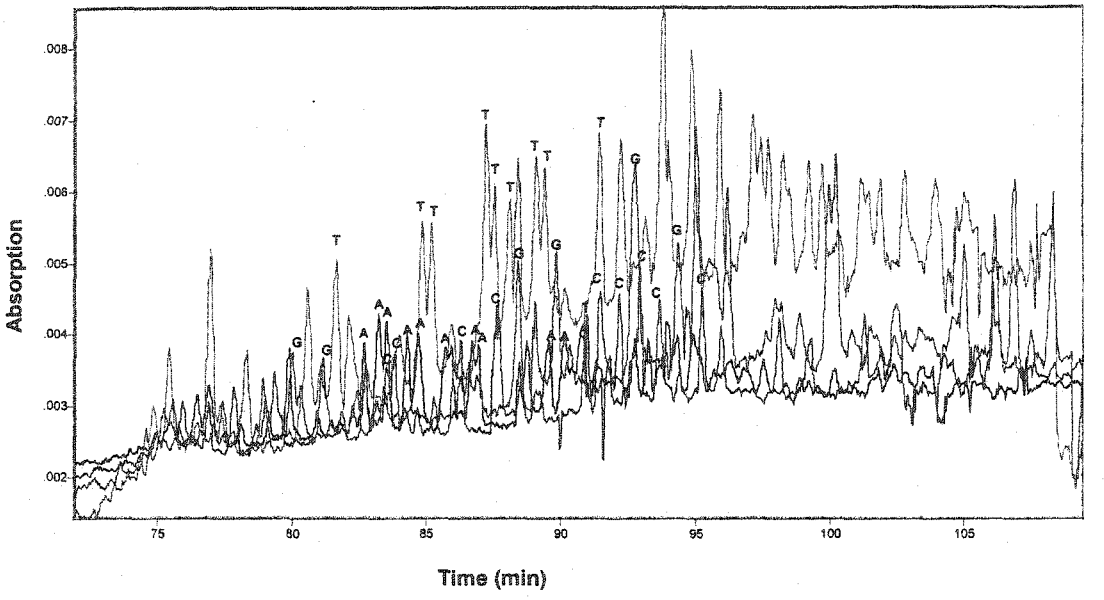

Absorption
.008
.007
.006
.005
.004
.003
.002
75
80
85
90
95
100
105
Time (min)

FIGURE 5

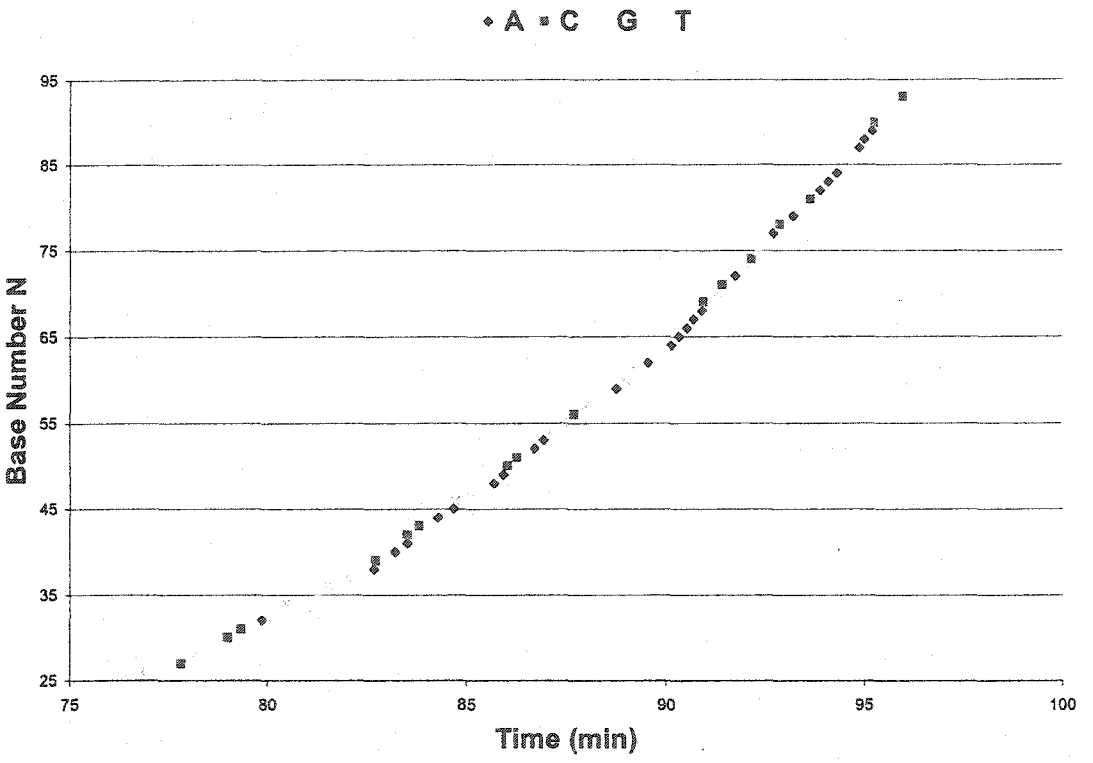

A C G T
95
85
75
65
55
45
35
25
Base Number N
75
80
85
90
95
100
Time (min)

FIGURE 6

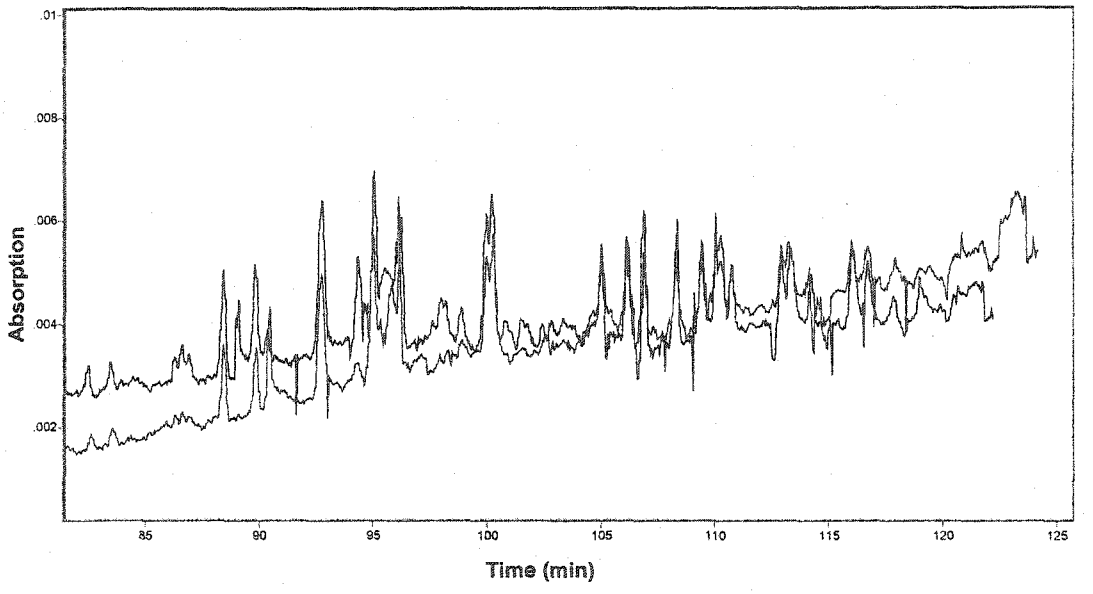

.01
.008
.006
.004
.002
Absorption
85
90
95
100
105
110
115
120
125
Time (min)

FIGURE 7

CHAPTER 3. COMBINATORIAL ENANTIOMERIC SEPARATION OF DIVERSE COMPOUNDS USING CAPILLARY ARRAY ELECTROPHORESIS

A paper published in Electrophoresis *

Wenwan Zhong and Edward S. Yeung

ABSTRACT

Combinatorial chiral separations were performed on a 96-capillary array electrophoresis system. A comprehensive enantioseparation protocol employing neutral and sulfated cyclodextrins as chiral selectors for common basic, neutral and acidic compounds was developed. By using only four judiciously chosen separation buffers, successful enantioseparations were achieved for 49 out of 54 test compounds spanning a large variety of pKs and structures. Therefore, unknown compounds can be screened in this manner to identify the optimal enantioselective conditions in just one run.

---

* Reprint with permission from Electrophoresis 2002, 23, 2996

INTRODUCTION

Enantiomers of biologically active compounds interact differently with living systems [1]. Usually one enantiomer of a drug has desired pharmacological effect while the other is inactive or even toxic. In order to provide single enantiomeric pharmaceutical products, which are typically safer and more efficient than mixtures of enantiomers, pharmaceutical companies put substantial efforts in the development of both asymmetric synthesis and enantioseparation [2]. The role of chiral separation includes preparative isolation of the desired enantiomer from the racemate, quality control of the enantiomerically pure drugs, and pharmacokinetic studies of each enantiomer in the human body [3].

High-performance liquid chromatography (HPLC) is the dominant technique in chiral separation. However, for unknown compounds, e.g., a new drug candidate in the pharmaceutical industry, it often takes days of trial-and-error before a suitable condition is found for effective enantioselective separation. For developing suitable separation conditions, capillary electrophoresis (CE) may be a better technique. CE has higher separation efficiency and resolution so that selectivity is less of a problem; it is easy to employ different chiral selectors in CE, which involves just changing the running buffer and not switching to a new column; it allows lower operating cost and reduced solvent consumption, which are environmentally friendly features; and it dramatically shortens the method-development time and increases throughput by operating in a highly multiplexed mode. Capillary array electrophoresis (CAE) has already played a major role in the success of the Human Genome Project [4]. Recent work has demonstrated the potential of CE in high-throughput screening in drug discovery [5] proteomics [6], and combinatorial synthesis [7].

Amongst the various modes of CE demonstrated for direct chiral separation, capillary

zone electrophoresis (CZE) is the most widely used because of its simplicity—chiral selectors are simply added into the running buffers to perform enantioseparation [8]. In CE, the two enantiomers have identical electrophoretic mobilities owing to their common physical properties. The chiral selector is used to interact with the enantiomers to form reversible and labile diastereoisomeric or inclusion complexes. A minimum of three simultaneous spatial interactions should occur between the selector and the analyte [9]. In order to effect enantioseparation in CE [10], either the affinities of each enantiomer to the chiral selector are different, or the sizes or shapes of the selector-enantiomer complexes are not the same.

There are many kinds of selectors employed in CE, for example, cyclodextrins (CDs), crown ethers, macrocyclic antibiotics, proteins, linear polysaccharides, ligand-exchange type selectors and chiral micelles [11]. CDs are the most popular selectors in CE, because they are UV transparent, soluble, stable in aqueous solution and commercially available at a relatively low cost [12]. Native CDs are cyclic oligosaccharides composed of 6, 7 or 8 D-glucopyranose units, corresponding to α-, β-, or γ-CD respectively. In order to improve the solubility and enantioselectivity of CDs, many derivatized CDs have been synthesized by simple chemical modifications of the hydroxyl groups on the parent cyclodextrins. Thus, other than hydrophobic inclusion of the enantiomer to the CD cavity, there are other interactions between the selector and the analyte, including hydrophilic, polar and electrostatic interactions. In addition to using single chiral selectors to achieve enantioseparation, many groups have demonstrated that mixing two or more chiral selectors together can improve the resolution and offer more flexibility [13-16].

Since many chiral recognition mechanisms are not known in detail and therefore difficult to control, method development in enantioseparation is laborious and time consuming.

Many factors should be considered in the optimization of the experimental conditions, such as type and concentration of the selector, buffer pH, choice of other buffer components and buffer concentration. Each parameter may need to be varied at several levels before the optimal separation can be found. With the help of a capillary array system, many separation conditions can be tested in just one run. Therefore, the time needed to find the best condition can be dramatically reduced. Furthermore, these conditions can be chosen based on the wealth of method-development strategies suggested by previous results [17-21].

The purpose of our work is to design a universal screening system containing the most useful chiral separation conditions for common chemicals. Several different separation conditions, such as pH, different kinds of CDs, and different CD concentrations, are tested in a combinatorial manner for each compound. By applying this concept to a capillary array, robust separation conditions can be found for an arbitrary analyte in one single run. 54 chiral compounds of different structures were selected as analytes to represent common amines, acids, and neutral chemicals of medium size. The goal is to identify a minimal set of separation conditions that is suitable for base-line separation of most analytes.

EXPERIMENTAL SECTION

Reagents and Chemicals

Racemic test compounds were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO, U.S.A.), except that chlormezanone was from TCI America (Portland, OR, U.S.A.) and bupivacaine, dobutamine, and prilocaine were from ICN Biomedicals (Aurora, OH, U.S.A.). These were all dissolved in 10:1 H_2O/MeOH mixture to yield samples solutions of 0.2 mM. O-phosphoric acid, Trisma-base(tris[hydroxymethyl] aminomethane), so-

dium phosphate-monobasic, and sodium phosphate-dibasic were from Fisher Scientific (Fairlawn, NJ, U.S.A.). Didodecyldimethylammonium bromide was from Fluka (St. Louis, MO, U.S.A.). The plain running buffer of 25 mM phosphoric acid (titrated to pH 3.0 with Tris) was filtered through a filter system from Corning (Corning, NY, U.S.A.) containing a .22 μm cellulose membrane. Sulfated-β-CD (S-β-CD), heptakis(2,3,6-di-O-methyl)-β-CD (DM-β-CD), heptakis(2,6-tri-O-methyl)-β-CD (TM-β-CD), and hydroxypropyl-β-CD (HP-β-CD) were purchased from Sigma. β-CD sulfobutyl ether 4 was from Cydex (Overland Park, KS, U.S.A.). Calculated amounts of CDs were added to the plain buffer to obtain separation buffers. Other tested sulfate-group-containing CDs are heptakis-(2,3-diacetyl-6-sulfato)-β-CD (HDAS-β-CD) and heptakis-(2,3-dimethyl-6-sulfato)-β-CD (HDMS-β-CD) (Regis Tech., Morton Grove, IL, U.S.A.). Other running buffers with 25 mM phosphate, at pH 5.0, 7.0, and 9.0, were prepared by dissolving sodium phosphate-dibasic (pH 9.0) or sodium phosphate-monobasic (pH 5.0 and 7.0) into D.I. water and adjusted to the desired pH with orthophosphoric acid or sodium hydroxide.

Instrumentation

All separations were done on a multiplexed capillary array with UV absorption detection [22]. 96 capillaries (o.d. 150 μm, i.d. 75 μm) of 80 cm total length and 60 cm effective length were packed on a capillary holder. The outlet ends of every 8 capillaries were bundled together. Thus the whole array has 12 bundles and can test 12 different separation conditions for 8 compounds in the same run. The inlet ends of the 96 capillaries were mounted on a copper plate with 96 gold electrodes, which makes independent injection from a 96-well mi-

crotiter plate into each capillary possible. The light source was a zinc lamp (UVP, Upland, CA, U.S.A.) and the 213.9-nm wavelength was selected through an interference filter placed after the array and in front of a quartz lens (Nikon; focal length = 105 mm; f/4.5), which focuses the light onto a linear photodiode array (PDA) (Hamamatsu model S5964, Hamamatsu, Japan). A National Instrument PCI E Series multifunction 16-bit I/O board was employed to transfer data to a computer (1 GHz Pentium, Dell). The data were analyzed and converted into single-diode electropherograms by in-house Labview programs.

Before each run, the capillary array was rinsed with the corresponding running buffers containing 0.005% DDAB for 15 min to reverse the electroosmotic flow [23]. A 96-well microtiter plate (Daigger, Vernon Hills, IL, U.S.A.) filled with fresh running buffers but without DDAB at the injection end and twelve 1.7 ml centrifuge tubes at the detection end were filled with different running buffers with 0.0005% DDAB for replenishing the dynamic coating. Samples were loaded by 10 s of gravity flow (9 cm height). Separations were performed at –8 kV (100 V/cm) and the running current was kept at 20 μA per capillary to limit joule heating. Two cooling fans were used to circulate air around the array to maintain a running temperature of around 25 °C.

RESULTS AND DISCUSSION

Theoretical consideration

The goal of this work is to develop a universal chiral separation system for most common compounds. It is therefore very important to select representative test compounds. Important considerations are the functional groups carrying different charges at different pH and the property of the groups proximal to the asymmetric center that can lead to stereospeci-

fic interactions. Our collection of 54 analytes includes basic compounds with different amine groups (primary, secondary, tertiary, and aromatic vs. aliphatic nitrogens), carboxylic acids, phosphoric acid, and neutral compounds. Some typical structures are shown in Fig. 1.

Optimization of the chiral separation of such a large collection of compounds should start with a theoretical understanding of the various selectivity models [24-26]. The simplest system to consider is one in which both the chiral selector and the enantiomers are in pure form, i.e., they are either fully ionized or neutral. In such a system, the expression for the effective electrophoretic mobility would be [27]

$$\mu_{A,eff} = \frac{\mu_A + \mu_{AC} K_A [C]}{1 + K_A [C]} \tag{1}$$

μ_A is the ionic mobility of analyte A, μ_{AC} is the mobility of the selector-analyte complex, K_A is the equilibrium constant of the complex, and [C] is concentration of the selector. Then the separation selectivity S becomes [28]

$$S = \frac{\mu_{(R)A,eff} - \mu_{(S)A,eff}}{\mu_{ave,eff} + \mu_{eo}} = \frac{\Delta\mu_{eff}}{\mu_{ave,eff} + \mu_{eo}} \tag{2}$$

Where μ_{eo} is the mobility of electroosmosis flow (EOF). The difference between the two effective mobilities depends on [C] and K_A [29],

$$\Delta\mu_{eff} = \frac{(\mu_A - \mu_{AC})(K_{R,A} - K_{S,A})[C]}{1 + (K_{R,A} + K_{S,A})[C] + K_{R,A} K_{S,A} [C]^2} \tag{3}$$

If both the selector and analyte are ionizable, the situation is more complicated since K_A would become the apparent stability constant and would vary with pH [25]. So the type and concentration of the chiral selector and the pH of the running buffer are the most important

factors in selectivity tuning. Since cyclodextrins are the most commonly used chiral selectors, we chose CDs for the present scheme. Different CDs at different concentrations were investigated for the separations of the collection of 54 compounds. Also, we evaluated the separations under different pH conditions using one CD—sulfated-β-CD.

Enantioseparations with sulfated-β-CD

In separations based on CDs, the size of the CD cavity has an important influence on selectivity. Complexation occurs only when the hydrophobic part of the analyte should be at least partially included into the hydrophobic cavity of the CD. Then, chiral recognition depends on the protruding side chains having conformational interactions with the groups on the rim of the cavity. Such interactions include hydrogen bonding, dipole interaction, hydrophobic interaction, electrostatic force and so on. β-CD has 7 glucose units, and its cavity size matches the sizes of most common pharmaceutical compounds with one aromatic ring. So β-CD and its derivative forms are the most useful CDs in chiral separation.

From Eq. 3, it is easy to see that $\Delta\mu_{eff}$ would reach a maximal value when [C] = $1/(K_{R,A}K_{S,A})^{0.5}$ [30, 31]. A lower stability constant requires a higher concentration of selector to achieve the maximal mobility difference. That is why charged CD is more favorable for separation. In addition to the ability to mobilize neutral compounds, if the CD has charge opposite to that of the analyte, the strong coulombic attraction promotes complexation. Furthermore, if μ_{AC} and μ_A have different signs, the value of $\Delta\mu_{eff}$ increases even if other values are kept constant [32]. Therefore, usually lesser amounts of charged CDs are needed for chiral separation compared to that of neutral CDs. Negatively charged S-β-CD alone is a

good selector for a wide variety of compounds because of its high number (between 7 and 11) of sulfate groups [33], especially for positively charged compounds. The sulfate groups give the CD a permanent negative charge at any pH above 2.0, which would make optimization simpler. Acidic pH is favored for the separation of basic enantiomers using negatively charged CD because the selector and analyte carry opposite charges. The acidic condition is also good for separating acidic compounds. It has been observed that enantiomeric resolution is difficult for acidic compounds in their anionic form with negatively charged CDs [34, 35]. This is because when the pH is higher than the pKa of the weak acids, they would be repelled by the sulfated-β-CD.

Initially, we tested the selectivity of S-β-CD under acidic condition, pH = 3.0. Seventeen of the 54 compounds were successfully separated and 3 (trimipramine, chlorpheniramine, and cloperastine) were partially resolved using 1.5% S-β-CD in 25 mM pH 3 tris-phosphate buffer. Fig. 2 shows the enantioseparations of some selected compounds and Table I lists the migration times and separation factors of the 17 compounds. The triangular peak shapes in Figure 2 are characteristic of mismatches in the electrophoretic mobilities between the analyte and the buffer ions [36, 37]. We compared the separation results for 1.5% S-β-CD, 2% S-β-CD and 3% S-β-CD and found that the lower concentration already provided sufficient separation. Increasing the S-β-CD concentration did not improve the separation much but increased the current to over 40 μA per capillary. Also, we tried several single-isomer β-CD derivatives with sulfate groups, such as HDAS-β-CD and HDMS-β-CD. The advantage of using single-isomer CDs is to eliminate the differences in composition and resolving characteristics from batch to batch [33, 38]. HDAS- and HDMS-β-CD required con-

centrations as high as 15 to 30 mM for good separation because of their relatively low sulfate substitution numbers. Furthermore, these two single isomers did not show as broad an enantioselectivity as S-β-CD. The latter is also much cheaper yet showed no obvious selectivity variations between runs. 5 mM SBE-β-CD alone did not provide the same performance as the S-β-CD for many of the 17 compounds. But, by mixing with either DM- or TM-β-CD, *vide infra*, SBE-β-CD can separate most of the remaining compounds tested, including neutral and acidic compounds. In contrast, S-β-CD could not resolve any of the chiral acids or neutral compounds tested here. Therefore, S-β-CD appears to be a good selector for compounds with a protonated amine group or a hydroxyl group linked directly to the asymmetric carbon atom. It is reasonable to assume that the aromatic ring fits into the cavity while the –OH group or the amine group forms hydrogen bonds with the oxygens at the rim of the CD cavity to cause chiral recognition. Electrostatic interaction between the positively charged amino group and the negatively charged sulfate group enhances the stability of the complex.

Other than acidic conditions, we also tried the separations using 1.5% S-β-CD at pH 5.0, 7.0, and 9.0. Because EOF is quite large under these pH conditions, no DDAB coating was needed. The applied electric field was in the normal direction (positive to negative). Oxyphenylclimine, trimiprazine, and nafronyl (with 2 chiral centers, partially separated into three peaks) were separated at pH 5.0, promaquine was resolved at pH 7.0, and promethazine and chlorpheniramine were resolved at pH 9.0 (Fig. 3). However, the goal here is to use as few separation conditions as possible to cover as many enantiomers with different properties as possible. Since the basic pH conditions do not appear to be as versatile as the acidic ones, they are not included in subsequent experiments.

According to the general inclusion model of analytes to CDs, at higher pH conditions like 5.0, 7.0 and 9.0, most amino groups are only partially protonated. The coulombic attraction is thus weak and less favorable hydrogen bonding exists without the protonated amino group. On the other hand, the use of those CDs in which some hydroxyl groups on the rim are substituted by hydrophobic ones might be helpful to the separation under acidic conditions. HP-β-CD, DM-β-CD, and TM-β-CD are three neutral CD derivatives previously shown to be highly effective in separating a large number of chiral compounds [25, 39]. We therefore included these three kinds of CDs in our study. However, since using neutral CDs alone does not impart any electrophoretic mobility to the non-charged analytes, we employed a dual CD system by mixing charged CDs with these neutral CDs.

Dual buffer system

A dual CD system usually contains one neutral CD and one charged CD. Such a dual system offers more flexibility [16]. For the neutral compounds, the charged CD provides the necessary electrophoretic mobility while complexation with the neutral CD might give higher selectivity. It has been suggested that for the dual system higher selectivity can be achieved if the two selectors have opposite effects on both the mobility and the affinity patterns [15, 34, 40-42]. For example, if one selector accelerates the enantiomer and the other decelerates it, better separation while using both selectors would only occur if one has stronger binding with the S form and the other has stronger binding with the R form. It is thus important to first characterize each selector before using the two selectors together. Another advantage of mixing the neutral and negative CDs together is that the concentration of the charged CD need not be as high as in the single-CD system. Selectivity can be improved by increasing

only the concentration of the neutral CD.

We mixed each neutral CD with the two sulfated CDs, S-β-CD and SBE-β-CD, respectively, to create the dual systems. The concentrations of S-β-CD were 1.2% or 1.5% and those of SBE-β-CD were 3 mM or 5 mM. The concentrations of the neutral CDs were varied from 5 mM to 30 mM at 5 mM intervals for evaluation. The CDs were dissolved in the same buffer, 25 mM tris-phosphate (pH 3.0). After screening all 54 compounds under all conditions, three mixtures were found out to be most effective. They are 10 mM HP-β-CD with S-β-CD, 5 mM DM-β-CD with 5 mM SBE-β-CD, and 20 mM TM-β-CD with 5 mM SBE-β-CD. Trimipramine and cloperastine, as well as the other 4 chiral amines including nafronyl and trimiprazine (Fig. 3 and Table II), were resolved in 10 mM HP-β-CD with 1.2% S-β-CD. Four carboxylic acids and one tertiary amine were separated in 5 mM SBE-β-CD with 20 mM TM-β-CD mixtures (Fig. 4 and Table III). The other 20 compounds of bases, acids and neutrals were resolved using 5 mM DM-β-CD with 5 mM SBE-β-CD (Fig. 5 and Table IV). Three of these compounds (promaquine, promethazine, and chlorpheniramine) were also resolved under basic conditions using 1.5% S-β-CD only. But, the three dual-CD buffer systems under acidic conditions provided good enantioseparations for a wider range of compounds, especially those with bulky size (two aromatic rings close to each other like suprofen and fenoprofen, or containing three rings like thioridazine, promethazine, and oxyphenylclimine), or with more hydrophobic groups around the chiral carbon atom (warfarin, dobutamine, and chlorpheniramine), or acidic and neutral compounds without any charge under acidic conditions.

As we discussed above, it is necessary to test the condition of using 5 mM SBE-β-CD

alone to see whether mixing two CDs together improved or degraded the separation. Very few analytes could be resolved and even then only partially under this condition. They were atenolol, indapamide, promaquine, dobutamine, and homatropine. Many others bases produced broad peaks that reflect some selectivity while not much distortion in peak shape was seen for the acidic and neutral compounds. We conclude that here SBE-β-CD acts only as a carrier of the non-charged compounds. Mixing it with neutral CDs enhanced the overall chiral selectivity. Other attempts like mixing S-β-CD with DM- or TM-β-CD, and mixing SBE-β-CD with HP-β-CD did not improve the separations. Actually very few successful separations based on this type of dual system were reported in the literature [13, 15, 43].

There are five compounds in our set of analytes which were not resolved under these four buffer conditions or under other conditions tested during these experiments. They are acebutolol, menadione sodium bisulfite, 4-bromomandelic acid, phenylpropanol, and methyltetradone (Figure 7). Menadione sodium bisulfite carries a negative charge at pH 3.0, so it does not benefit from the use of negatively charged CD as selector. Successful enantioseparation of the two acidic racemates, menadione sodium bisulfite and 4-bromomandelic acid, were achieved using cationic β-cyclodextrin, quaternary ammonium-β-CD, or amphoteric-β-CD [44]. Acebutolol has a long chain substituted at the 4- position on the aromatic ring, making it difficult to fit into the cavity of β-CD. So only a broad peak was observed under all conditions but inadequate separation was achieved. For the remaining two (neutral) compounds, since interaction between the charged CD and the non-charged analytes is weak, higher sulfated CD concentrations have been shown to be suitable for enantioseparation [45]. That can be accomplished by including this as a fifth buffer in the present combinatorial set.

Even though this fifth buffer requires a higher operating current, the total current for the array should still be tolerable. For completeness, structures of the 35 test compounds not shown in Figures 1 and 7 are shown in Figure 8.

Other factors

There are factors other than pH and CD concentration that affect enantioselectivity, e.g., organic modifier, background buffer ions, and ionic strength. In our experiments, we used a common buffer component, phosphate buffer at a low concentration (25 mM) because we already have sulfated CD as an additive. The electrical current of the capillary array would be too high if we used high phosphate concentrations. A high operating temperature can degrade the enantioseparation because of joule heating and changes in complexation [25]. Adding an organic modifier to the running buffer would improve the solubility of the samples and lower the running current, but would decrease the effective mobility of the analytes and weaken the complexation with cyclodextrin. We tried several buffers containing different concentrations of MeOH, but did not obtain better results since MeOH also destroyed the DDAB coating. The DDAB coating was needed to create reversed EOF to accelerate elution of the sample under acidic pH and reversed electric field [23]. So we did not include organic modifiers in the selected buffers.

CONCLUSIONS

Four CE buffers were found to be sufficient for effective enantioseparations of 49 out of the 54 test compounds. The use of a 96-capillary array system dramatically decreased the screening time and enhanced the efficiency of method development. The separation times

here were limited by the lengths of the capillaries used in the array. There is no inherent reason why shorter capillaries or higher field strengths cannot be used to reduce analysis times to 10-15 min. Based on the diversified structures (with or without basic or acidic functional groups) and size (containing two to three benzene rings or other heterocyclic rings) of these compounds, it can be concluded that these conditions should be suitable for most common pharmaceutical products. A more comprehensive system can be derived from these four conditions by further varying the concentrations of the CD selectors and employing CDs with similar substitutions but different cavity sizes (α-CD or γ-CD instead of β-CD). For example, the use of a high concentration of highly charged CDs of different sizes allowed the separation of most compounds tested [45]. However, the running current is about 7 times higher than that observed here. For an array system, these high currents will be problematic for reliable operation. Also, several optimization steps requiring operator interference were required to provide broad coverage [45]. For further refinement of the universal chiral separation system, experimental variables like the type of selector, organic modifier and other CE separation modes should be included, at the cost of a lower throughput. In the combinatorial approach suggested here, after initial screening with the four conditions, some knowledge about the appropriate chiral selector will be obtained even if baseline separation is not achieved. For the development of pharmaceuticals, this represents a substantial savings in time and effort.

ACKNOWLEDGEMENT

The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract W-7405-82. This work was supported by the Director of Science,

Office of Basic Energy Sciences, and the National Institutes of Health.

REFERENCES

[1] Batra, S., Seth, M., Bhaduri, A. P., *Prog. Drug Res.* **1993**, *41*, 192-248.

[2] Stinson, S. C., *C&E News* **1995**, *October 9*, 44.

[3] Maier, N. M., Franco, P., Lindner, W., *J. Chromatogr. A* **2001**, *906*, 3-33.

[4] Collins, F., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., *Science* **1998**, *282*, 682-689.

[5] Ma, L., Gong, X., Yeung, E. S., *Anal. Chem.* **2000**, *72*, 3383-3387.

[6] Kang, S. H., Gong, X., Yeung, E. S., *Anal. Chem.* **2000**, *72*, 3014-3021.

[7] Zhang, Y., Gong, X., Zhang, H., Larock, R. C., Yeung, E. S., *J. Comb. Chem.* **2000**, *2*, 450-452.

[8] Koppenhoefer, B., Zhu, X., Jakob, A., Wuerthner, S., Lin, B., *J. Chromatogr. A* **2000**, *875*, 135-161.

[9] Dalgliesh, C. E., *J. Chem. Soc.* **1952**, 3940-3942.

[10] Chankvetadze, B., Blaschke, G., *J. Chromatogr. A* **2001**, *906*, 309-363.

[11] Rizzi, A., *Electrophoresis* **2001**, *22*, 3079-3106.

[12] Gubitz, G., Schmid, M. G., *J. Chromatogr. A* **1997**, *792*, 179-225.

[13] Izumoto, S., Nishi, H., *Electrophoresis* **1999**, *20*, 189-197.

[14] Fillet, M., Chankvetadze, B., Crommen, J., Blaschke, G., *Electrophoresis* **1999**, *20*, 2691.

[15] Fillet, M., Hubert, P., Crommen, J., *J. Chromatogr. A* **2000**, *875*, 123-134.

[16] Gahm, H., Lee, J., Chang, L. W., Armstrong, D. W., *J. Chromatogr. A* **1998**, *793*,

135-143.

[17] Rawjee, Y. Y., Staerk, D. U., Vigh, G., *J. Chromatogr.* **1993**, *635*, 291-306.

[18] Fillet, M., Hubert, P., Crommen, J., *Electrophoresis* **1998**, *19*, 2834-2840.

[19] Roos, N., Ganzler, K., Szeman, J., Fanali, S., *J. Chromatogr. A* **1997**, *782*, 257-269.

[20] Liu, L., Nussbaum, M. A., *J. Pharm. Biomed. Anal.* **1999**, *19*, 679-694.

[21] Matthijs, N., Perrin, C., Maftouh, M., Massart, D. L., Heyden, Y. V., *J. Pharm. Biomed. Anal.* **2002**, *27*, 515-529.

[22] Gong, X., Yeung, E. S., *Anal. Chem.* **1999**, *71*, 4989-4996.

[23] Melanson, J. E., Baryla, N. E., Lucy, C. A., *Anal. Chem.* **2000**, *72*, 4110-4114.

[24] Williams, B. A., Vigh, G., *J. Chromatogr. A* **1997**, *777*, 295-309.

[25] Vespalec, R., Bocek, P., *Chem. Rev.* **2000**, *100*, 3715-3753.

[26] Eeckhaut, A. V., Boonkerd, S., Detaevernier, M. R., Michotte, Y., *J. Chromatogr. A* **2000**, *903*, 245-254.

[27] Otsuka, K., Terabe, S., *Trends Anal. Chem.* **1993**, *12*, 125.

[28] Fanali, S., Kilar, F., *J. Capillary Electrophor.* **1994**, *1*, 72.

[29] Issaq, H. J., Chan, K. C., *Electrophoresis* **1995**, *16*, 467.

[30] Lloyd, D. K., Li, S., Ryan, P., *J. Chromatogr. A* **1995**, *694*, 285.

[31] Wren, S. A. C., Rowe, R. C., *J. Chromatogr. A* **1992**, *603*, 235-241.

[32] Chankvetadze, B., *J. Chromatogr. A* **1997**, *792*, 269-295.

[33] Vincent, J. B., Sokolowski, A. D., Nguyen, T. V., Vigh, G., *Anal. Chem.* **1997**, *69*, 4226-4233.

[34] Fillet, M., Bechet, I., Schomburg, G., Hubert, P., Crommen, J., *J. High Resolut. Chromatogr.* **1996**, *19*, 669.

[35] Chankvetadze, B., Endresz, G., Blaschke, G., *J. Chromatogr. A* **1995**, *704*, 234-237.

[36] Mikkers, F. E. P., Everaerts, F. M., Verheggen, T. P. E. M., *J. Chromatogr.* **1979**, *169*, 1-10.

[37] Mikkers, F. E. P., Everaerts, F. M., Verheggen, T. P. E. M., *J. Chromatogr.* **1979**, *169*, 11-20.

[38] Cai, H., Nguyen, T. V., Vigh, G., *Anal. Chem.* **1998**, *70*, 580-589.

[39] Fanali, S., *J. Chromatogr. A* **1997**, *792*, 227-267.

[40] Chankvetadze, B., *Trends Anal. Chem.* **1999**, *18*, 485.

[41] Surpaneni, S., Ruterbories, K., Lindstrom, T., *J. Chromatogr. A* **1997**, *761*, 249.

[42] Lelievre, F., Gareil, P., Bajaddi, Y., Galons, H., *Anal. Chem.* **1997**, *69*, 393.

[43] Fillet, M., Fotsing, L., Crommen, J., *J. Chromatogr. A* **1998**, *817*, 113-119.

[44] Tanaka, Y., Terabe, S., *J. Chromatogr. A* **1997**, *781*, 151-160.

[45] www.beckmancoulter.com/resourcecenter/labresources/ce/pdf/chiral38-alpha.pdf

FIGURE CAPTION

Figure 1. Chemical structures of typical test compounds: (1) aminoindane, (2) oxyphenylclimine, (3) pindolol, (4) promethazine, (5) disopyramide (pKa = 10.2), (6) homatropine (pKa = 4.35), (7) chlorthalidone (pKa = 4.4), (8) econozole, (9) suprofen (pKa = 3.91), (10) phenylprophen oxide, (11) bupivacaine (pKa = 8.09), (12) isoproterenol (pKa = 8.64), (13) phenylpropanolamine (pKa = 9.44), (14) chlorpheniramine.

Figure 2. Separations by using S-β-CD alone. Conditions: 25 mM tris-phosphate buffer, pH 3.0, containing 1.5% S-β-CD; 0.0005% DDAB dynamic coating (75 μm ID, 60 cm effective length); –8 kV, 214 nm.

Figure 3. Optimization by tuning the pH value. Conditions: 25 mM phosphate buffer at pH 5.0, 7.0, and 9.0; 1.5% S-β-CD; +8 kV, 214 nm.

Figure 4. Separations by using a 1.2% S-β-CD with 10 mM HP-β-CD (dual system). Other conditions are the same as in Fig. 2.

Figure 5. Separations by using a 5 mM SBE-β-CD with 20 mM TM-β-CD (dual system). Other conditions are the same as in Fig. 2.

Figure 6. Separations by using a 5 mM SBE-β-CD with 5 mM DM-β-CD (dual system). Other conditions are the same as in Fig. 2.

Figure 7. Structures of the compounds not resolved in the four selected buffers. (1) Acebutolol, (2) 4-bromomandelic acid, (3) 2-phenyl-1-propanol, (4) menadione sodium bisulfite, (5) methyltetralone.

Figure 8. Structures of the 35 test compounds not shown in Figures 1 or 7.

Table I. Migration times (t) and apparent selectivity coefficients (α) of the 17 compounds separated by using 1.5%S-β-CD in 25 mM tris-phosphate buffer (pH 3.0). $\alpha = \mu_1^{app}/\mu_2^{app}$ (μ^{app} is the apparent mobility of the enantiomer).

Analyte	t_1 (min)	t_2 (min)	α
Alprenolol	33.14	37.97	1.145
Aminoindane	38.50	51.00	1.320
Atropine	29.65	37.22	1.255
Bupivacaine	42.35	47.85	1.130
Chloramphetamine	24.41	27.33	1.120
Disopyramide	29.86	35.01	1.172
Homatropine	37.41	53.15	1.420
Isoproterenol	21.23	24.62	1.160
Labetalol	21.62	22.13	1.024
Metaproterenol	13.30	13.95	1.049
Nefopam	14.04	17.32	1.233
Phenylpropanolamine	27.64	30.91	1.118
Pindolol	23.89	25.69	1.075
Sotalol	35.76	39.46	1.103
Tetrahydropaveroline	24.63	31.80	1.291
Tetrahydrozoline	33.12	38.29	1.156
Tryptophanamide	49.36	51.79	1.049

Table II. Migration times (t) and apparent selectivity coefficients (α) of the 9 compounds separated by using a mixture of 1.2% S-β-CD and 10 mM HP-β-CD in 25 mM tris-phosphate buffer (pH 3.0).

Analyte	t_1 (min)	t_2 (min)	α
Bupivacaine*	39.90	45.31	1.136
2,3-di(4-pyridyl)-2,3-butanediol	29.75	31.46	1.057
Cloperastine	21.26	23.89	1.124
Nafronyl	29.36	31.57	1.075
Pindolol*	17.98	19.33	1.075
Prilocaine	40.33	43.85	1.087
Trimiprazine	30.06	32.22	1.071
Trimipramine	29.42	34.13	1.160
Tropicamide	33.67	60.61	1.800

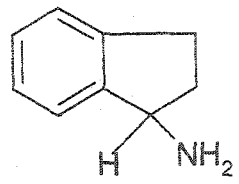
*Also resolved under the conditions in Table I.

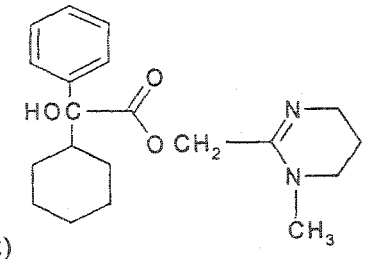
Table III. Migration times (t) and apparent selectivity coefficients (α) of the 5 compounds separated by using a mixture of 5 mM SBE-β-CD and 20 mM TM-β-CD in 25 mM tris-phosphate buffer (pH 3.0).

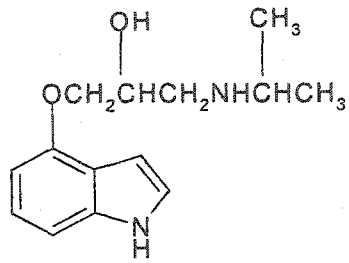
Analyte	t_1 (min)	t_2 (min)	α
Chrysanthemum	22.88	24.15	1.056
Doxylamine	30.65	32.26	1.052
Fenoprofen	31.00	34.89	1.125
Phenoxypropionic acid	33.82	34.92	1.033
Suprofen	24.37	25.51	1.046

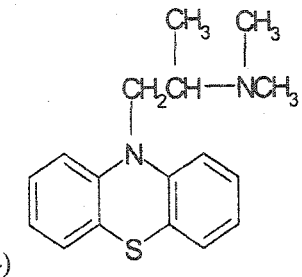
Table IV. Migration times (t) and apparent selectivity coefficients (α) of the 20 compounds separated by using a mixture of 5 mM SBE-β-CD and 5 mM DM-β-CD in 25 mM tris-phosphate buffer (pH 3.0).

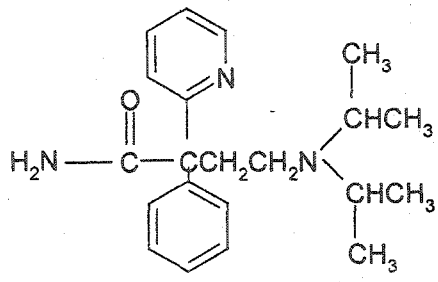
Analyte	t_1 (min)	t_2 (min)	α
Abscisic acid	37.82	40.08	1.060
Atenolol	55.44	60.35	1.089
1,1-binaphthyl-2,2-	17.29	19.29	1.116
Chlormezanone	35.49	37.48	1.056
Chlorthalidone	23.73	27.88	1.175
Chlorpheniramine	31.05	32.66	1.052
Dobutamine	14.34	14.91	1.039
Econozole	19.58	22.01	1.124
Flurbiprofen	25.30	26.37	1.042
Indapamide	31.55	34.82	1.104
Ketoprofen	18.57	19.05	1.026
Methylphenylsuccinimide	28.65	29.50	1.030
Oxyphenylclimine	24.35	28.33	1.163
Phenylbutyric acid	29.37	31.72	1.080
Phenylpropene oxide	53.32	57.56	1.080
Primaquine	21.63	24.80	1.146
Promethazine	33.06	36.18	1.094
Thioridazine	24.90	28.32	1.137
Tropic acid	26.26	29.11	1.109
Warfarin	23.70	25.56	1.078

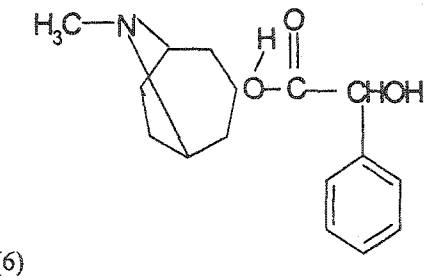
(1) 

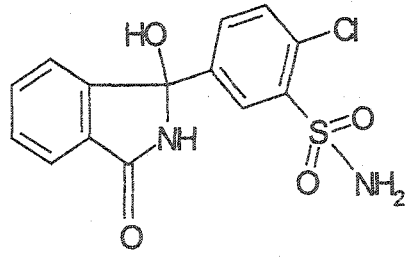
(2) 

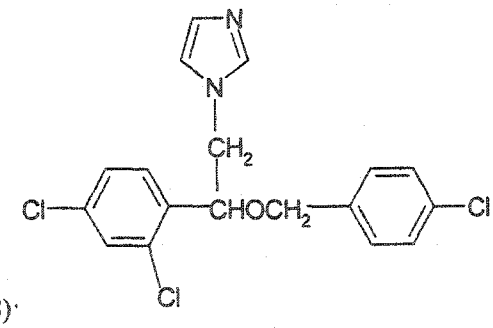
(3) 

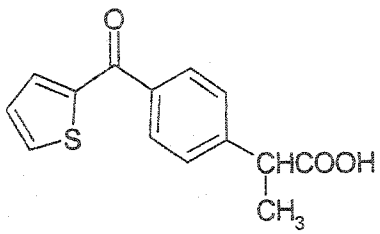
(4) 

(5) 

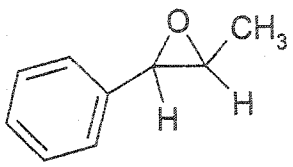
(6) 

(7) 

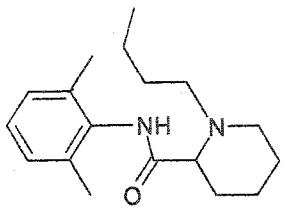
(8) 



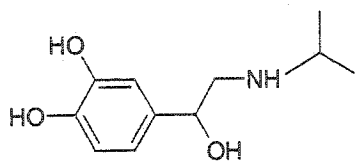
(9)



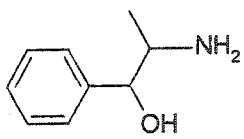
(10)



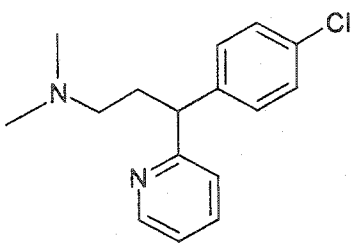
(11)



(12)



(13)



(14)

FIGURE 1

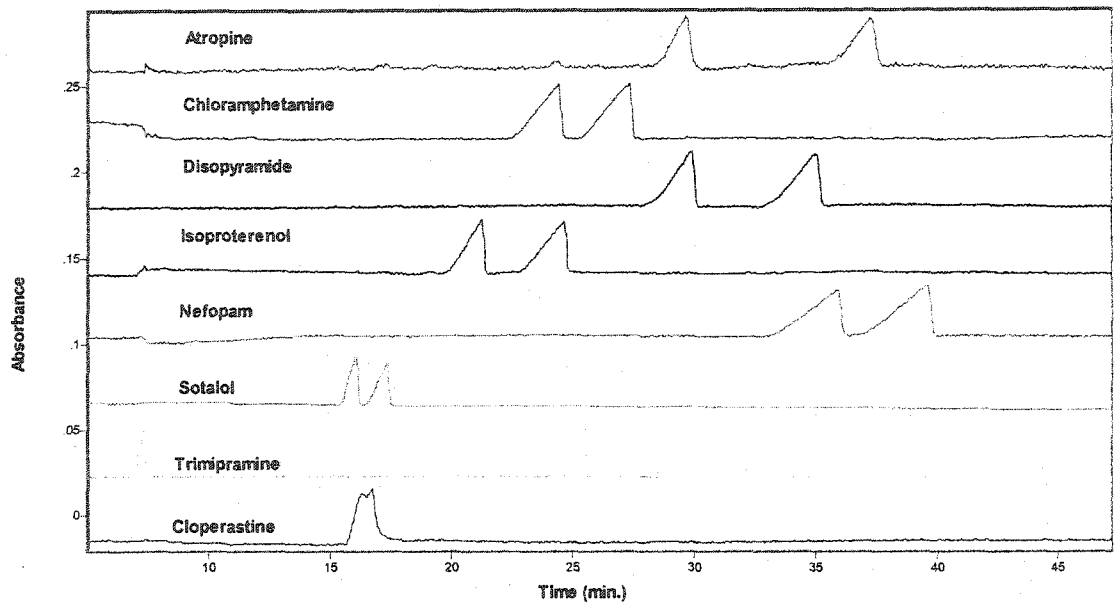

Atropine
Chloramphetamine
Disopyramide
Isoproterenol
Nefopam
Sotalol
Trimipramine
Cloperastine
Absorbance
.25
.2
.15
.1
.05
0
10
15
20
25
30
35
40
45
Time (min.)

FIGURE 2

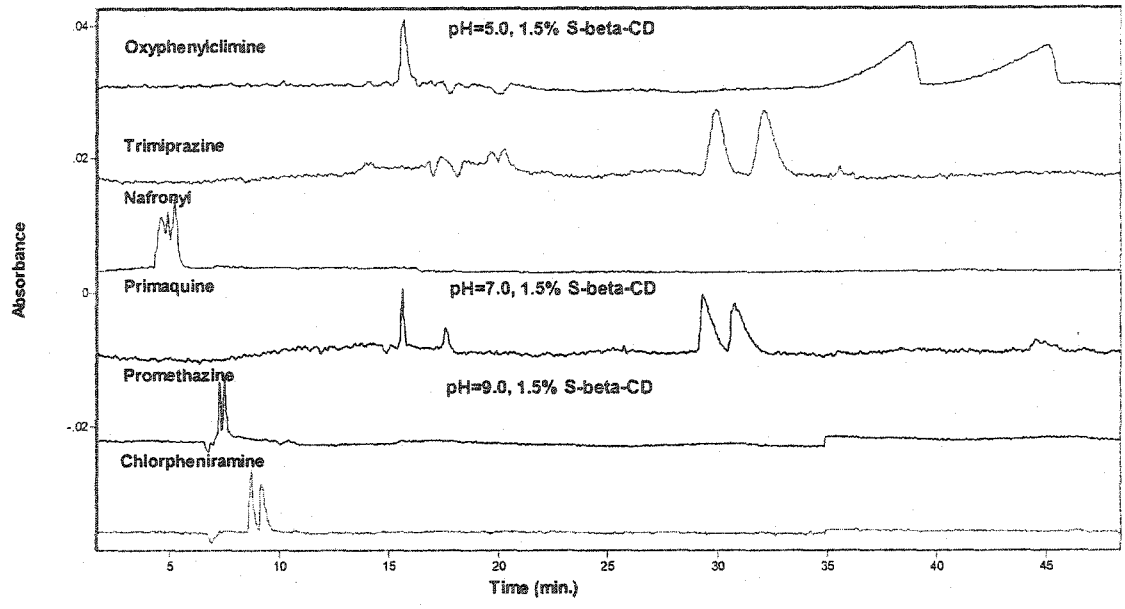

Absorbance
.04
.02
0
-.02
Oxyphenylclimine
pH=5.0, 1.5% S-beta-CD
Trimiprazine
Nafronyl
Primaquine
pH=7.0, 1.5% S-beta-CD
Promethazine
pH=9.0, 1.5% S-beta-CD
Chlorpheniramine
5
10
15
20
25
30
35
40
45
Time (min.)

FIGURE 3

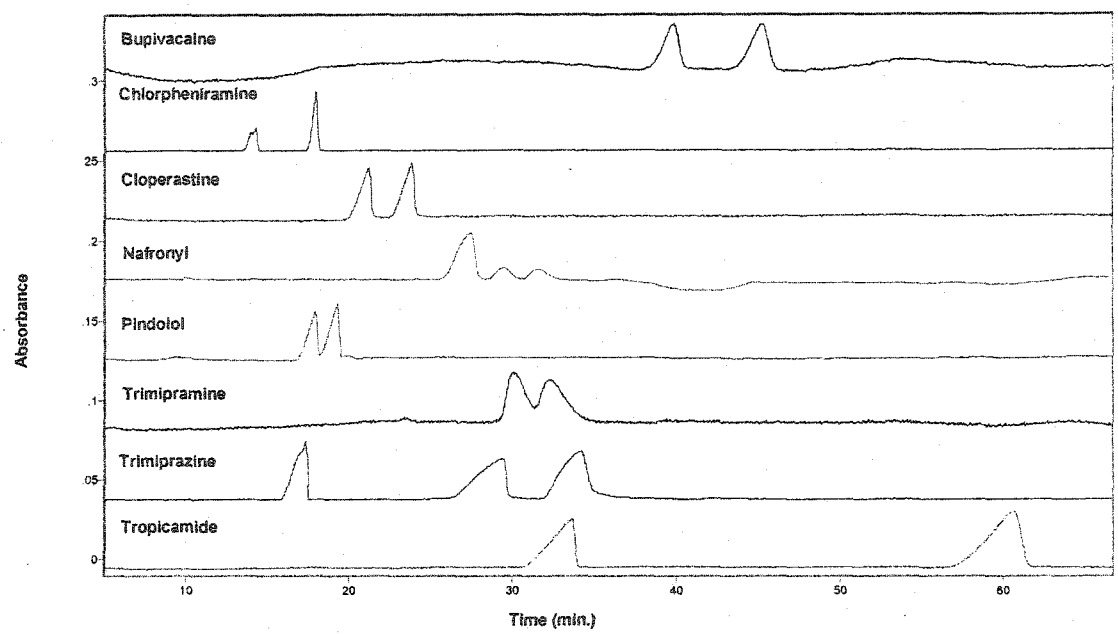

Bupivacaine
Chlorpheniramine
Cloperastine
Nafronyl
Pindolol
Trimipramine
Trimiprazine
Tropicamide
Absorbance
.3
.25
.2
.15
.1
.05
0
10
20
30
40
50
60
Time (min.)

FIGURE 4

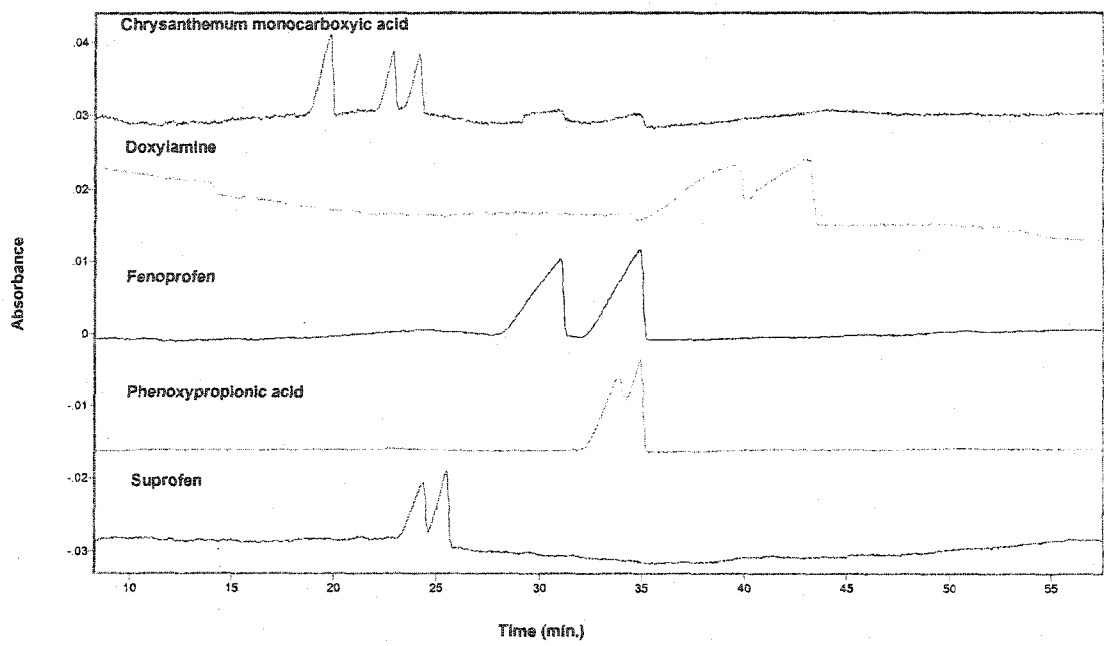

Chrysanthemum monocarboxyic acid
Doxylamine
Fenoprofen
Phenoxypropionic acid
Suprofen
Absorbance
.04
.03
.02
.01
0
-.01
-.02
-.03
10
15
20
25
30
35
40
45
50
55
Time (min.)

FIGURE 5

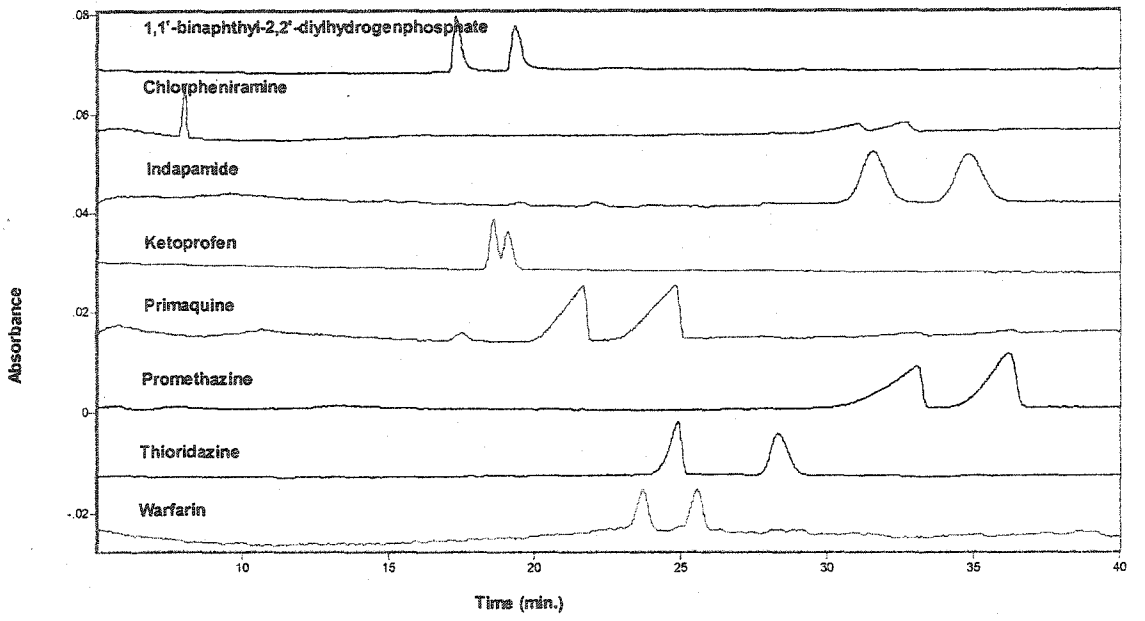
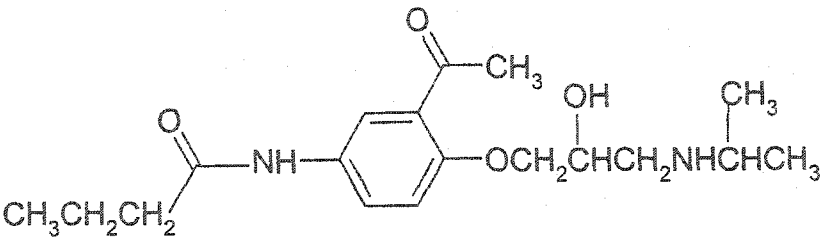
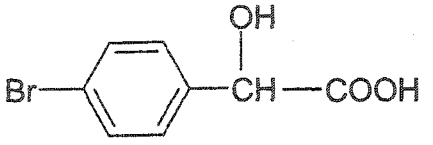

1,1'-binaphthyl-2,2'-diylhydrogenphosphate
Chlorpheniramine
Indapamide
Ketoprofen
Primaquine
Promethazine
Thioridazine
Warfarin
Absorbance
.08
.06
.04
.02
0
-.02
10
15
20
25
30
35
40
Time (min.)

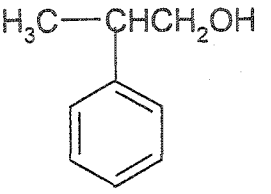
FIGURE 6



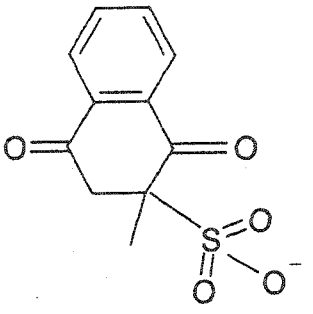
(1)



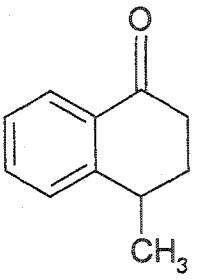
(2)



(3)



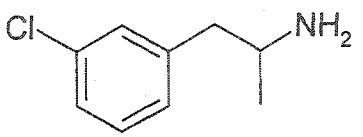
(4)



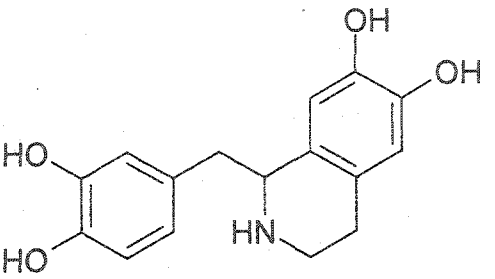
(5)

FIGURE 7

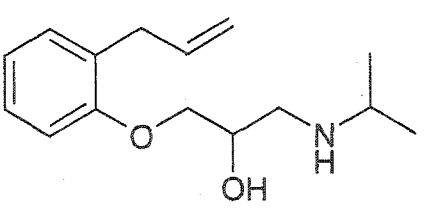
pH = 3, 1.5% S-β-CD



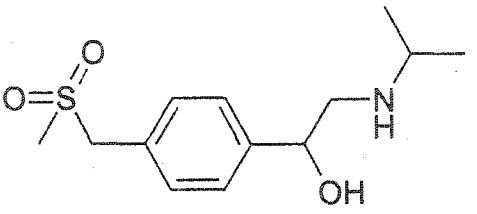
Chloramphetamine



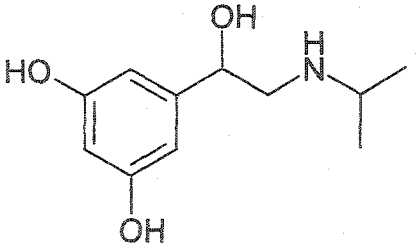
Tetrahydropaveroline



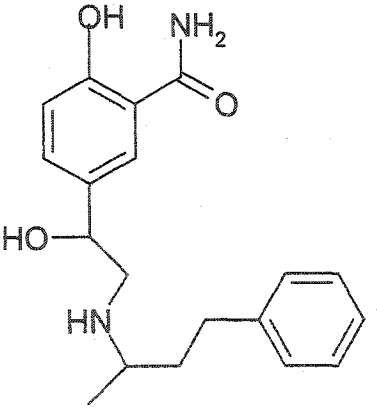
Alprenolol



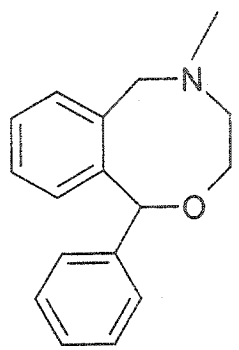
Sotalol



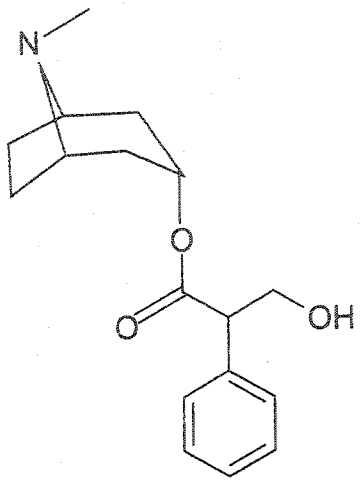
Metaproterenol



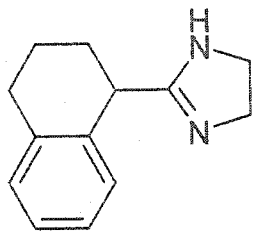
Labetalol



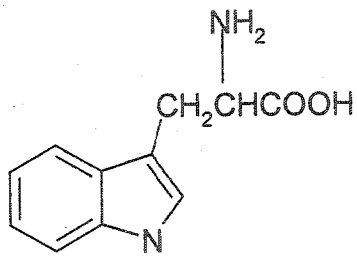
Nepofam



Atropine

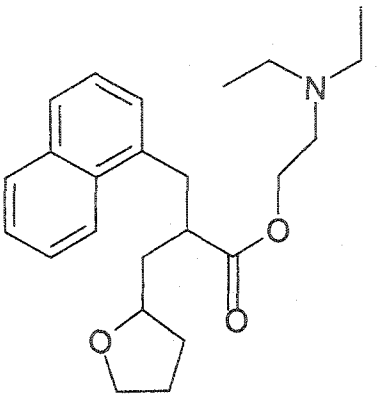


Tetrahydrozoline

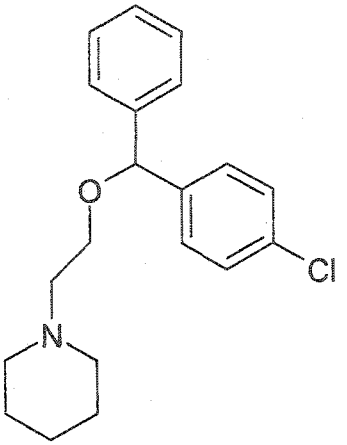


Tryptophanamide

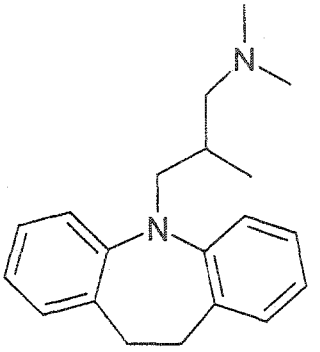
pH = 3, 1.2% S-β-CD, 10 mM HP-β-CD



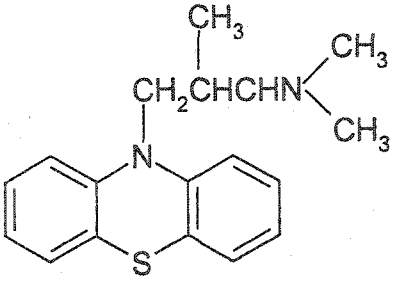
Nafronyl



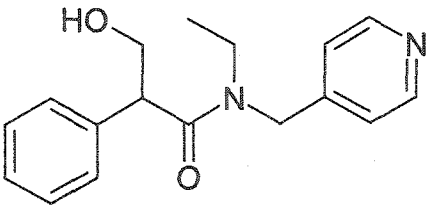
Cloperastine



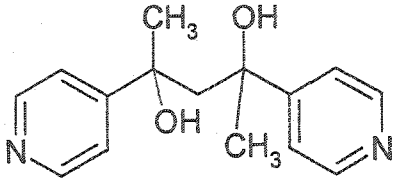
Trimipramine



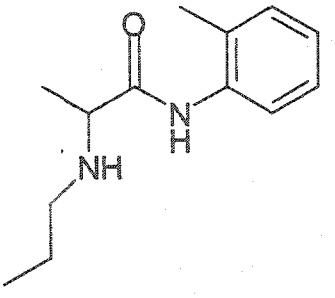
Trimiprazine



Tropicamide

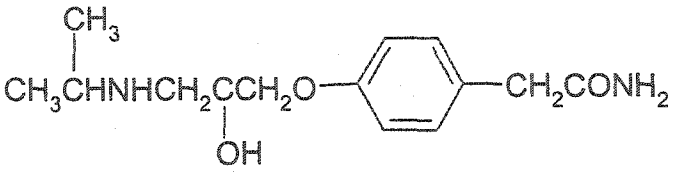


2,3-di(4-pyridyl)-2,3-butanediol

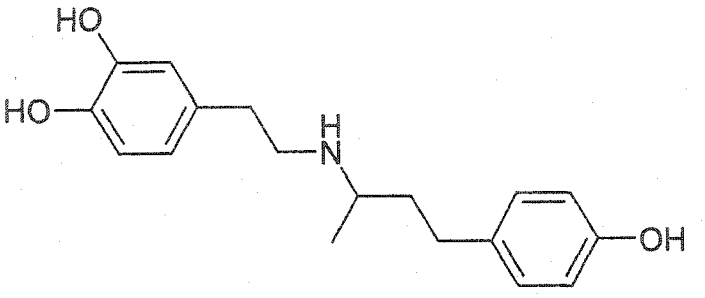


Prilocaine

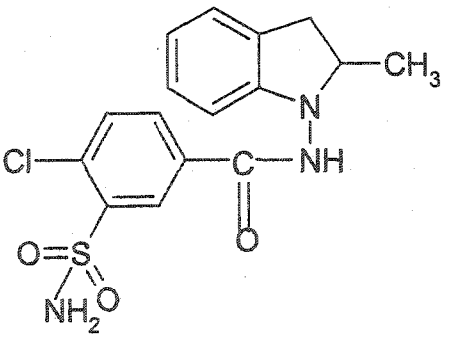
pH = 3, 5 mM SBE-β-CD, 5 mM DM-β-CD



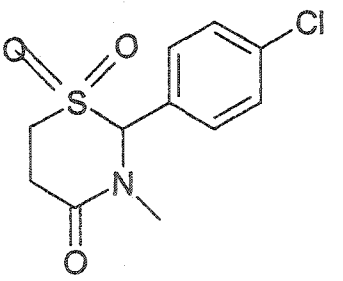
Atenolol



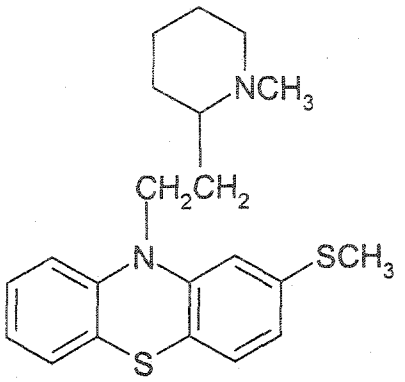
Dobutamine



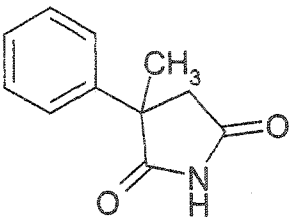
Indapamide



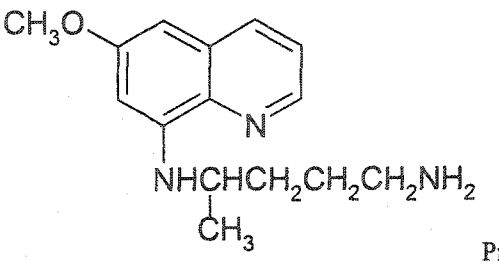
Chlormezanone



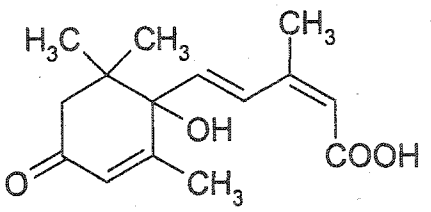
Thioridazine



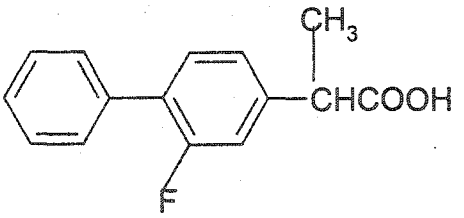
Methylphenylsuccinimide



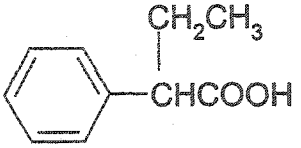
Primaquine



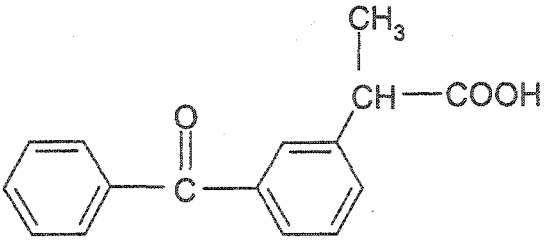
Abscisic acid



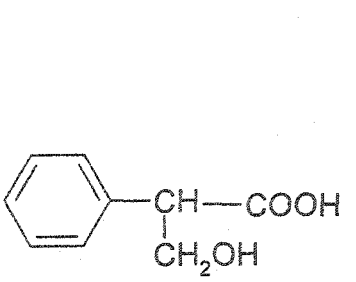
Flurbiprofen



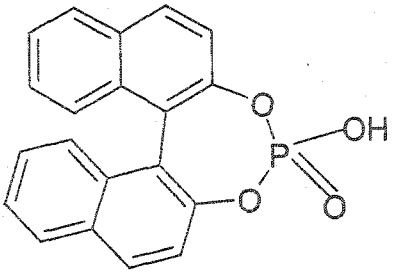
Phenylbutyric acid



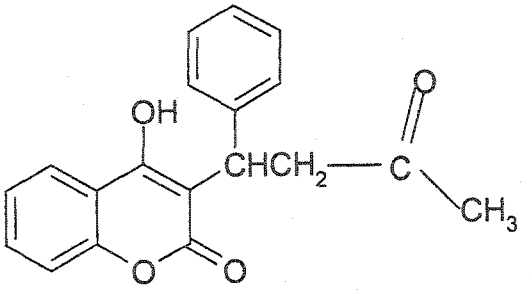
Ketoprofen



Tropic acid

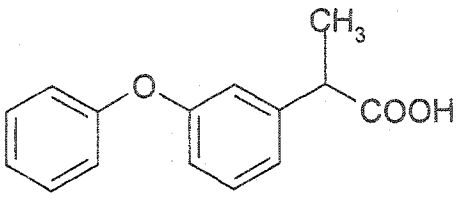


1,1'-binaphthyl-2,2'-diylhydrogenphosphate

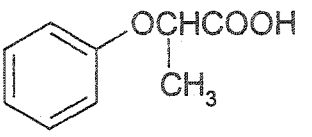


Warfarin

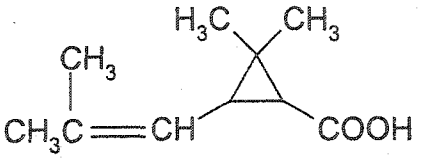
pH = 3, 5 mM SBE-β-CD, 20 mM TM-β-CD



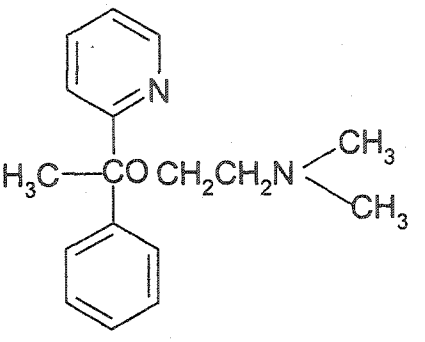
Fenoprofen



Phenoxypropionic acid



Chrysanthemum monocarboxylic acid



Doxylamine

FIGURE 8

CHAPTER 4. HIGH-THROUGHPUT ANALYSIS OF TOTAL RNA EXPRESSION PROFILES BY CAPILLARY GEL ELECTROPHORESIS

A paper published in Analytical Chemistry

Wenwan Zhong and Edward S. Yeung

ABSTRACT

We studied the expression profiles of total RNA in three types of human tissues. Reverse transcription (RT) was performed on total RNA samples extracted from kidney, normal breast, and breast tumor. Then, fluorophor-labeled cDNAs were synthesized by using random hexamers as primers for analysis by capillary gel electrophoresis (CGE). By overlaying the electropherograms and by performing multivariate analysis, we found that the RNA patterns were significantly different among the tissue samples. The protocol utilizes only submicroliter volumes of PCR products and is compatible with multiple capillary DNA sequencing instruments. These CE-based RNA patterns provide an alternative to RNA microarrays for expression profiling for understanding the molecular basis of disease.

---

* Reprinted with permission from Analytical Chemistry, in press.

INTRODUCTION

Gene expression profiling among different tissues has attracted a lot of interest in various areas of biomedical research. The comparison of healthy and disease tissues helps to identify therapeutic target genes, diagnostic markers, or pathways involved in cell regulation and tumor development.[1, 2] Furthermore, gene expression fingerprint in tumor tissues could lead to classification of cancer subtypes, discovery of novel molecular taxonomies, and prediction of clinical outcomes.[3-5] It may eventually bring about improved patient stratification and superior therapeutic strategies. Although the total number of genes in the human genome is around 30,000 to 35,000,[6] only ~15% are expressed in a distinct cell type to produce no more than 5000-6000 proteins, which orchestrate most of the cell functions. Therefore, the study of gene expression profiles among different tissues can provide direct clues to the biological functions of genes.

Many modern methodologies have been developed to achieve simplicity, high throughput, automation, and miniaturization in gene expression profiling.[7-9] Microarray is one of the most powerful techniques, because it allows the accumulation of large amounts of functional genomic information by visualizing the expression levels of thousands of genes in a parallel fashion. The key feature of microarrays is the DNA chip, a solid support that has thousands of known-sequence DNA fragments (probes) deposited on its surface at precise positions. The two oligonucleotide samples that are being compared are labeled with two different fluorescent tags, and then simultaneously hybridized to the chip. The fluorescence intensities are recorded and the relative intensities indicate the gene expression differences in the two samples.[8, 10, 11] Even though DNA chips offer many advantages, such as true paral-

lelism, miniaturization, multiplexing and automation, it is technically demanding and requires a large amount of sample (about 500 μL). DNA microarrays are also expensive to manufacture. Variations and errors cannot be completely avoided during the multi-step experiments, which results in noise in the microarray data. Data processing is complicated, prone to human errors, and liable to inconsistencies in probe hybridization.[12-16] Furthermore, it is not suitable for searching for novel genes and studying gene expression of new organisms because the construction of microarrays relies on the existence of a database of complementary DNA sequences of the organism.

Differential display (DD) is one gene expression profiling technique that does not require prior knowledge of the sequences of the genes. In DD, a set of arbitrarily designed primers is used to amplify mRNA sub-populations by polymerase chain reaction (PCR). After amplification, the products are separated by gel electrophoresis to derive the expression pattern. The genes responsible for the differences can subsequently be isolated and characterized.[9, 17-20] Compared to microarrays, DD is good for the discovery of novel genes, simpler and less expensive. Because a small sample volume is employed, it is more sensitive than microarrays in detecting low abundance mRNA species. However, there exist artifacts that give rise to false positives in expression comparison.[21] Also, in order to cover over 95% of the genes expressed in a cell, the total number of arbitrarily designed primers needed is 240. So, 240 PCR reactions and DNA separations must be performed on each sample, which limits its broad applicability.[18]

Instead of identifying expression level changes of specific genes, some groups have studied the overall changes in total RNA samples. Total RNA samples contain ribosomal

RNA (rRNA), transfer RNA (tRNA), and low molecular weight species besides messenger RNAs (mRNA). rRNAs are the major components of ribosome, where the proteins are synthesized. tRNAs serve as amino acid carriers in protein synthesis. So the expression of these RNAs is very important in the investigation of regulation of protein production. It has been found that a large amount of the cell's energy and resources during cell growth and cell division are used to produce ribosomes. Regulation of rRNA expression may provide an important mechanism for controlling these cell processes.[22] Han et al. found that the total amount of RNA in individual Chinese hamster ovary cells changed and the accumulation rate for specific RNA fractions was different at distinct stages of the cell proliferation cycle.[24] Low-molecular-mass RNA fingerprinting has been used for identification and classification of bacteria.[25] There, sensitivity was limited by the use of absorption detection. Also, irreproducibility in the separation implied that only qualitative differences were derived.

Here, we investigated the total RNA profiles of three types of tissues by using capillary gel electrophoresis (CGE) to separate the fluorophor-labeled cDNA fragments produced from reverse transcription (RT) and random hexamer priming. Only sub-microliter samples are required. Highly reproducible profiles were found for each tissue. Multivariate analysis was employed for pattern recognition based on chromatographic profiles. It has been widely used in many other fields such as the fingerprinting of flavor components in coffee,[26] characterization of mycobacteria species by HPLC,[27] classification of fungi,[28] and peptide mapping.[29] Several chemometric methods, such as principal component analysis (PCA) and cluster analysis, are used for comparing and identifying differences among samples.[26-31] Here, these methods are employed to reveal the differences in RNA expression among tissue

types.

EXPERIMENTAL SECTION

Chemicals and reagents

Human kidney total RNA, human breast total RNA, and human breast tumor total RNA were from Clontech (BD Biosciences, Palo Alto, CA). 10-bp DNA ladder, dNTPs, random hexamers, 5x first-strand buffer, DTT, DNase I, RNaseOut recombinant ribonuclease inhibitor and reverse transcriptase were purchased from Invitrogen (Carlsbad, CA). Alexa Fluor 488 dUTP for labeling cDNA was from Molecular Probes (Eugene, OR). TBE buffer dry powder was obtained from AMRESCO (Solon, OH). Poly(ethylene oxide) (PEO), molecular weights of 8,000,000 and 600,000, respectively, were from Sigma-Aldrich (St. Louis, MO).

cDNA synthesis and purification

Extreme care was taken while handling RNA. All reaction tubes and storage microcentrifuge tubes were RNase-free, and the water used was DEPC treated. The working area was cleaned with 70% ethanol before the experiments. Reverse transcription reactions were performed in a thermocycler (ABI 2400, Applied Biosystems, Foster City, CA). The protocol of fluorescence labeling RT-PCR was according to Molecular Probes. Four microgram of primers was annealed to 3 μg total RNA in a thin-wall PCR tube (Bio-Rad, Hercules, CA) by heating the mixture for 10 min at 98 °C at a total volume of 11 μL. cDNA synthesis was then initiated by mixing the annealed primer-RNA with 2 μL reverse transcriptase, 5 μL 5x first-

strand buffer, 2.5 μL 100 mM DTT, 2.3 μL d(GACT) (5 mM each of dGTP, dATP, dCTP, and 2.27 mM dTTP), 1.5 μL Alexo Fluor 488 dUTP, 1 μL DNase I and 1 μL Rnase OUT recombinant ribonuclease inhibitor in a final volume of 27 μL, and incubated for 10 min at 25 °C. The reaction was continued at 42 °C for 90 min. After the reaction, the mixture was purified by using a spin column (Princeton Separations, Adelphia, NJ), and then stored at –20 °C. Negative control reactions without total RNA templates were performed in parallel with the normal reverse transcription reactions. CGE experiments on the negative control showed no reverse transcription products other than the small amount of dye molecules left after being purified by the spin column.

CGE-LIF analysis of reverse transcription products

The CGE experiments were performed in a laboratory-built capillary electrophoresis laser-induced fluorescence (CE-LIF) system. 40 cm (30 cm effective length) fused-silica capillaries (75 μm i.d., 360 μm o.d.) from Polymicro Technologies (Phoenix, AZ) were used. A 488-nm argon ion laser (5 mW, JDS Uniphase Inc., Manteca, CA) was used for excitation. The laser beam was focused on the detection window by using a 1-cm focal length lens. Fluorescence was collected with a 10x microscope objective, passed through a 520-nm interference filter (Oriel, Stratford, CT) and focused onto a photomultiplier tube (PMT) (R928, Hamamatsu, NJ). The analog fluorescence signal collected by the PMT was transferred into a computer through a 22-bit A/D converter (Personal Daq/55, IO Tech, Cleveland, OH). A laboratory-developed Labview program recorded the signals at 2 Hz. The sieving matrix was a mixture of 1.5% (w/v) MW 8,000,000 PEO and 1.4% (w/v) MW 600,000 PEO in 1x TBE buffer with 7 M urea. The property of the gel was monitored every day using 10-bp ssDNA ladders in an ISCO

model 3140 electropherograph system (Lincoln, NE) to ensure the reproducibility of the separations. cDNA samples were denatured at 98 °C for 3 min before injection. Injections were performed electrokinetically at 75 V/cm for 40 s. The electric field applied for all the separations was 250 V/cm, and the temperature was kept at 25 °C. Multivariate analysis was performed by SIMPCA-P (Umetrix Co., Kinnelon, NJ) software.

RESULTS AND DISCUSSION

Reproducibility

In a typical human cell, there are about 10-30 pg of total RNA. Within the total RNA, 80-85% are rRNAs and 15-20% are tRNAs. The remaining 1-5% are mRNAs. So the total RNA from one cell may consist of 200,000 to 1,000,000 molecules.[33] Random priming further generates multiple products from each RNA species, with a bias for the GC-rich priming regions. So, the cDNA products may have thousands of fragments with different lengths that cannot be resolved even with existing high-resolution CGE instruments. However, it should be possible to perform pattern recognition and classification in order to extract expression profiles as long as reproducible electropherograms can be obtained.[25-29]

We chose the PEO matrix developed by our group for DNA sequencing as the sieving matrix.[33] Mixing high and low molecular weight PEO together helped to resolve both small and large DNA fragments. Under optimized conditions, this gel had been found to be suitable for sequencing DNA up to 1000 bases.[33] However, it would take a very long time to complete each run if those conditions were utilized. Furthermore, in comparing the electropherograms of cDNA products to the separation of the 10-bp ladder (data not shown), we estimate

that the longest cDNA products that have adequate S/N to be recognized contain about 600 bases. This is likely due to early termination of the dye-label reactions. Therefore, we used a higher electric field and a shorter capillary here to speed up the separation while preserving reasonable resolution. The resolving power of the PEO matrix was tested with 10-bp ssDNA ladder. The result is shown in Fig. 1. The spacing between the adjacent peaks can theoretically accommodate 2 more peaks with a resolution of R=0.5. This indicates that denatured cDNA fragments having 2-3 bp difference in lengths can be adequately resolved by using this sieving matrix. Because the goal was to obtain whole RNA profiles from different types of samples, this experimental condition was adequate for our purpose.

RT reactions for total RNA from different types of tissues were performed in parallel to minimize operational variations caused by sample handling or experimental conditions. All other reaction reagents were pre-mixed in one 1.5-mL microcentrifuge tube. The prescribed volume of the cocktail solution and primers was added to each mixture of total RNA after annealing for the RT reaction. In this way, the only difference among these RT reactions is the total RNA templates. Three types of tissues were tested. They were normal kidney, normal breast and breast tumor samples. For each type of tissue, four RT reactions were performed to evaluate reaction reproducibility, and two of the reaction products were injected three times to ensure that the separations were also reproducible.

Fig. 2 shows one of the electropherograms obtained from the normal kidney sample. The baseline increased just prior to the appearance of the first peak, and decreased to the initial level after 25 min. The baseline change probably reflects the presence of many unresolved fragments. The first few major peaks can be attributed to the residual dye-dUTP and short cDNA products from low molecular weight RNA. These features were not reproducible

even among reactions from the same type of tissue or among different injections from the same reaction tube. Differences in the amount of dye residue were from irreproducibility in operating the spin columns. Also, the short DNA fragments (less than 50 bp) may not be well resolved under this polymer condition.[33] The concentration of the short-chain polymer in the separation matrix could be increased to provide higher resolution for the small DNA fragments.[34] Considering the above reasons, reproducible results for the first part of the electropherogram cannot be expected. However, we found that the later parts of the electropherograms (region indicated by the two arrows) were very reproducible among samples from the same type of tissue. This region should therefore provide useful information on the variations in the total RNA from different tissues.

It is known that variations in the surface chemistry of the capillaries and inhomogeneities in the gel matrix after being pushed into the capillaries can cause substantial changes in the migration times of the DNA fragments in CGE. However, these can be corrected for by normalization using internal standards.[35] Even though we did not co-inject any standards, we chose the first and the last peaks in the "reproducible" region as migration references to perform the normalization after background subtraction. Fig. 3 depicts the comparison of 6 electropherograms from two RT reactions of normal breast tissue. The results from triplicate injections of each set of reaction products were overlaid to evaluate the reproducibility between runs. We can see that the number of features (peaks) and their positions did not change from reaction to reaction. The relative peak areas changed slightly due to random variability of the RT reaction. Continuous degradation of template RNA in the sample may also contribute to the variation from run to run. Digestion of template RNA after RT reaction could improve the reproducibility. Another type of variation observed is the absolute intensity of

the fluorescence signal, which was caused by variations in alignment and focusing of the LIF detection system. Small variations from run to run should not affect the utility of the overall RNA profile for comparing different tissues. If desired, injection and signal intensity variations can be normalized by co-injecting internal standards in future experiments.[35] For the twenty or so major product peaks in ten replicates of the RT reactions from normal breast tissue, the normalized retention times and the signal magnitudes showed relative standard deviations (RSD) of less than 1% and 9%, respectively. Considering the complexity of the sample and the experimental operation, the reproducibility is excellent.

Comparison of the total RNA profiles from different tissues

Next, we compared the total RNA electropherograms derived from three types of human tissues: normal kidney, normal breast, and breast tumor. Fig. 4a superimposes one of the electropherograms (green) from kidney tissue on top of 10 electropherograms from breast tissue (black), Fig. 4b is the complement of Fig. 4a, Fig. 4c shows one electropherogram from breast tumor tissue (red) on top of electropherograms from kidney tissue (green), and Fig. 4d compares the patterns between healthy and disease breast tissues by overlaying the former (black) on top of the latter (red). Because of the existence of only small variations, averaging the electropherograms from the same type of tissue for a superimposed comparison is not useful. The electropherograms were normalized to adjust for the mobility shifts among runs. The process of data manipulation before normalization, background subtraction and baseline correction is illustrated in the figures included in Supporting Information.

Comparing the two normal tissues (Fig. 4a and b), several peaks show higher intensities in the kidney tissue, which means that the corresponding RNAs are expressed more in

kidney. For most of the other peaks, these two samples have a high degree of similarity. The total RNA pattern of breast tumor is easily distinguished from that of normal breast tissue (Fig. 4d). The most obvious features are the relative peak heights of the three largest peaks, a, b, and c. The average ratios of peak heights of a/b and c/b in the normal tissues were around 0.9 and 0.4 respectively, with 5% RSD. However, the average ratios in the breast tumor tissue were 0.4 for a/b and 0.05 for c/b. The genetic reasons causing such differences cannot be revealed using RT reaction only. It will be necessary to collect these fractions for further DNA sequencing on those fragments. Successful fraction collections from CGE followed by PCR and DNA sequencing on the recovered DNA fragments have already been accomplished recently.[36,37]

Since we used total RNA as the templates and random hexamers as primers to perform RT reactions, the labeled cDNA were produced not only from mRNA, which contains information on gene transcription from DNA, but also from rRNA and tRNA, which provide information on the translation process. Each RNA species can further produce multiple products (sizes) because of random priming (with some bias for priming the GC-rich regions). Our approach is easy to implement but increases the number of cDNA products, leading to complicated RNA profiles. If the starting material is $Poly^{+}(A)$ RNA or if Oligo (T) is used as the primer to produce labeled cDNAs from the most abundant mRNAs only, as in DD, fewer number of peaks and simpler patterns will be obtained from these experiments. The pattern recognition will be easier but it would sacrifice some information. In that case, the RT-PCR reaction protocol should be reoptimized to produce even larger amounts of cDNAs from mRNAs to facilitate detection.

Multivariate analysis

An objective of this study was to define differences among different tissues. Variations in migration times and baseline fluctuations prevent the electropherograms from the same tissue from being perfectly superimposable. It is very difficult to confidently associate an unknown electropherogram with one of the three tissue types through subjective judgment based on the electropherograms in Fig. 4. A statistical analysis method was therefore sought to perform objective pattern recognition on the data. Multivariate data analysis (MVDA) is a method that is useful in obtaining an overview of large data sets, identifying major trends within the data, and classifying unknown samples based on a model built on known standards. It is basically a data reduction technique that computes linear combinations (weighted sums) of the original variables to yield a set of new variables that summarizes the information in the data. MVDA has been employed in infrared spectral analysis, gene expression data interpretation, and bacteria, virus, oil and minerals classification.[38-40]

In our case, two methods, principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were used to group different tissues based on their CE-based RNA profiles. The commercial software employed was designed for spectral analysis, where arrays of y-data (intensities) are matched to discrete x-data (wavelengths). So, we must first transform our electropherograms into spectrum-like data sets. Each electropherogram was used as one object, noted as an observation in PCA and PLS-DA. Each data point of fluorescence intensity in an electropherogram is then a variable. All of the data points contained in the selected (reproducible) region in Fig. 2 were included in the analysis. As little human manipulation as possible was used in order to minimize bias in judgment. Since fluorescence intensities were recorded every 0.5 s, each electropherogram contained

over a thousand variables. Because in the original data the migration time changed from run to run, the peaks from the same length DNA fragments may not elute at the same time point. For the same reason, the total number of data points in each electropherogram were different. In order to correlate the signals originating from the same fragments in different electropherograms, the migration times were shifted and normalized,[35] and all the electropherograms were scaled to have the same number of points.

Other researchers have developed different methods to properly align the chromatographic profiles prior to performing statistical analysis on the entire chromatogram. Some representative ones are peak tracking methods by using the UV spectra of all peaks,[41] use of Bessel's inequality to determine the optimal retention time adjustment,[42] dynamic programming,[43] correlation optimized warping (COW),[44] and non-linear alignment.[45] Most of these methods require knowledge of the identities of the components in the sample, detection of significant peaks in the chromatographic profile, or subjective peak-picking. COW does not have these requirements, but is still rather complicated. Here, we developed a simple algorithm for data alignment.

It is well known that in CGE, the migration time increases monotonically with the size of the DNA fragment.[46] Our alignment method linearly shrinks the time frame to fit into that of the standard set by proportionally reducing the number of data points in the data sets that have more data points than the standard set. The normalization algorithm is illustrated in the following. Among the 26 data sets (10 electropherograms from normal breast, 8 from kidney, and 8 from breast tumor), the one having the smallest number of data points (1114 points) was designated as the standard. One of the data sets to be normalized had 1261

points. After putting the data into an Excel spreadsheet, each data point was represented by two numbers. One was the column number (n), corresponding to the retention time; the other was the row number (original or normalized) corresponding to the data set (sample) number. Since the data set to be normalized had 1261/1114 times more data points than the standard one, Excel was used to pick out 1114 of the 1261 data points based on the address calculated by the function:

Address (n, normalized) = Address (n * INT (1261/1114), original)

The INT function rounds the number down to the nearest integer. When n=1, it picks the first point of the original set as the first point of the normalized set; when n=1114, it picks the last point of the original set. This algorithm omits some data points while making up the normalized data set, but still maintains almost the same resolution (decreased 1261/1114 times) and profile. The graphs in Fig. 4 show an acceptable alignment using this method.

Subsequently, PCA and PLS-DA were performed on these normalized data sets. In this case, Simca-P, a MVDA package from Umetrics Inc., was utilized to perform both tests on the data. PCA was performed to build a mathematical model that shows an overview of the entire data set. It reduced over one thousand variables to 12 principal components, the summarizing variables, which were computed as linear combinations of all the original variables. The 12 principal components explained over 95% of the cumulative variance in the data, and the first two principal components accounted for about 88% variance of the data in our case. In cross validation, the probability of placing an unknown sample into the right class was 70%. This predictability is fair considering the small sampling numbers (8 to 10 times) we performed on each type of tissue. It could be increased by using a greater number of samples. Because the first two components summarized the major variances in the data, a

score plot with the first principal component (t1) plotted against the second principal component (t2) depicts the main trends in the sample set (Fig. 5). A Hotelling T^2 ellipse is included in the plot. The ellipse is based on a 95% confidence interval, and the observations found outside the ellipse are potential outliers. In this case, no outliers were seen among the objects. The position of the object on this plot reflects how far away the object is from the center of the PCA model. Objects having similar properties are located close to each other in the score plot. For enhanced visualization, the objects were colored according to tissue types in the plot. Three distinct groups of objects on the scatter plot can be seen. Samples from the same type of tissue were found to be located near each other. Most importantly, the three types of tissue can be separated from each other via the first two principal components alone. Both normal tissue sample groups were positioned on the right side of the plot ($t1 > 0$) while the tumor tissue samples were located on the left side ($t1 < 0$). So, the tumor tissue can be separated from the two normal tissues via the first principal component. All normal kidney tissue samples were located at the upper right corner ($t2 > 0$) while all normal breast tissue samples were located at the lower right corner ($t2 < 0$). So, the two normal tissue groups can be separated via the second principal component. Since the first principal component (t1) describes the major variations, this overview agrees with our observations in Fig. 4, that the difference between normal and tumor tissues is larger than the difference between the two normal tissues.

When PCA results in clustering of observations, it is often useful to further resolve such groupings by means of PLS discriminant analysis. The objective of PLS-DA is to find a model that separates classes of observations on the basis of their X-variables. In our case, they are the same variables employed in PCA. Three dummy matrices of the three sets of Y-

variables expressing the class identity of the tissue samples were constructed. Fig. 6 is the t1 vs. t2 score plot of the PLS-DA model. The three classes of tissue samples were clearly separated. The separation was slightly superior compared to that of PCA. The w*c1 vs. w*c2 loading plot (w^* are the PLS weights used to combine the X-variables to form the new variables, c are the weights for the Y-variables, 1 and 2 represent the weights used to calculate the first two principal components) is also informative. It identifies the contribution of each variable to the class separation. The variables that are close to the three sets of Y-variables contribute strongly to the separation of the classes. Here, the variables that belong to the largest peak in Fig. 4d are close to the Y variable that represents tumor tissues. This implies that it is the main contributor to the separation of tumor tissue from normal tissue. It would be interesting to conduct further studies on these fragments, such as DNA sequencing, to sort out the genetic origin behind this observation.

PCA and PLS-DA illustrate there is clear separation of the three classes among the samples we tested. However, there is no quantitative description of how different these groups are. For classification and prediction of unknown samples, a model of each type of tissue could be built after analyzing a large number of samples. Then, the predictability of the model could be tested by assigning the samples from different types of tissues to the model. Finally, the identity of the unknown sample could be revealed by placing it within each model and looking for the best fit.

CONCLUSIONS

In summary, we found that total RNA profiles from human tissues have CE-based expression patterns that are easily distinguishable from one another. Future experiments with

larger numbers of different tissues will be needed to make this protocol broadly applicable. Monitoring the profile changes of total RNA inside the tissues may thus be useful in disease diagnosis, cancer therapy development, and understanding biological functions. Also, RNA patterns of different types of cells or cells at different growth stages can be investigated to follow gene expression. Here, the patterns were obtained through labeling cDNA fragment in RT-PCR reaction, followed by CGE separation of the products. The process is simple and the results are reproducible. Since predetermined RNA probes are not required, the use of expensive RNA arrays is avoided. Additionally, our method should be more general for the study of new species. Also, high throughput can be achieved because highly multiplexed CE instruments are widely available. Finally, the amounts of sample required for injection in CE are in the sub-microliter level, resulting in substantial cost savings compared to the 500-μL volumes required for microarrays. Our method suffers from the same disadvantage as DD—the specific RNA fragments that are expressed differently must be further sequenced for positive identification. However, our method is much simpler, more sensitive, less expensive and faster in detecting differences than the existing methods. The ability to obtain expression profiles with such small amounts of sample makes the approach unique. It may act as a complementary technique to or as a prescreening step for other more specific gene expression profiling techniques, such as microarrays and DD.

We developed a simple algorithm to modify the electropherogram data sets for multivariate data analysis. The PCA and PLS-DA models both revealed strong clustering among the three types of tissue. This clustering effect not only confirms the reproducibility of the RNA profiles, but also points out the possibility of classification and differentiation of types of tissues based only on such electropherograms. Even from the limited amount of data

available in this study, significant differences among the total RNA profiles from different tissues were identified. If more data sets are collected for each type of sample, a statistic model could be built for potential clinical applications. Other classification methods (for example, the Fischers discrimination method) could also be employed to further confirm these results.

The experiments here can be modified to explore the expression profiles of mRNA simply by changing the type of template to Poly^{+}(A) RNA or by changing the primer to Oligo (T). The use of fluorescent dNTPs will substantially enhance the sensitivity compared to traditional DD. For more accurate data analysis and normalization, internal standards[35] could be employed as guides for the adjustment of migration time and intensities. It should be possible to provide detailed genetic information on the origin of the observed differences if further efforts are taken to collect selected bands from the capillary to perform DNA sequencing.[47-49]

ACKNOWLEDGMENT

The authors thank Anna Person from Umetrics Co. and Dr. Chanan Sluszny for assistance in the multivariate data analysis. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No.W-7405-Eng-82. This work was supported by the Director of Science, Office of Biological and Environmental Research, and by the National Institutes of Health.

REFERENCES

(1) Aburatani, H. *International Congress Series* **2002**, *1246*, 261-270.

(2) Ricci, M. S.; El-Deiry, W. S. *Current Opinion in Molecular Therapeutics* **2000**, *2*, 682-690.

(3) Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J. P.; Coller, H.; Loh, M. L.; Downing, J. R.; Caligiuri, M. A.; Bloomfield, C. D.; Lander, E. S. *Science* **1999**, *286*, 531-537.

(4) Sorlie, T.; Perou, C. M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Thorsen, T.; Quist, H.; Matese, J. C.; Brown, P. O.; Botstein, D.; Eystein Lonning, P.; Borresen-Dale, A. L. *Proc. Natl. Acad. Sci. (USA)* **2001**, *98*, 10869-10874.

(5) van 't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.M.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J.; Witteveen, A.T.; Schreiber, G.J.; Kerkhoven, R.M.; Roberts, C.; Linsley, P.S.; Bernards, R.; Friend, S.H. *Nature* **2002**, *415*, 530-536.

(6) "Human Genome Project Information", http://www.ornl.gov/TechResources/Human_Genome/project/info.html.

(7) Velculescu, V. E.; Zhang, L.; Vogelstein, B.; Kinzler, K. W. *Science* **1995**, *270*, 484-487.

(8) Derisi, J. L.; Iyer, V. R.; Brown, P. O. *Science* **1997**, *278*, 680-686.

(9) Liang, P.; Pardee, A. B. *Science* **1992**, *257*, 967-971.

(10) Schena, M.; Heller, R. A.; Theriault, T. P.; Konrad, K.; Lachenmeier, E.; Davis, R. W. *Trends in Biotechnology* **1998**, *16*, 301-306.

(11) Holloway, A.J.; van Laar, R.K.; Tothill, R.D.; Bowtell, D.D.L. *Nat. Genet.* **2002**, *32*, 481-489.

(12) Martin, K. J.; Graner, E.; Li, Y.; Price, L. M.; Kritzman, B. M.; Fournier, M. V.; Rhei, E.; Pardee, A. B. *Proc. Natl. Acad. Sci. (USA)* **2001**, *98*, 2646-2651.

(13) Wernish, L. *Comp. Funct. Genom.* **2002**, *3*, 372-374.

(14) Knight, J. *Nature* **2001**, *410*, 860-861.

(15) Brody, J.P.; Williams, B.A.; Wold, B.J.; Quake, S.R. *Proc. Natl. Acad. Sci. (USA)* **2002**, *99*.

(16) Chuaqui, R.F.; Bonner, R.F.; Best, C.J.M.; Gillespie, J.W.; Flaig, M.J.; Hewitt, S.M.; Phillips, J.L.; Krizman, D.B.; Tangrea, M.A.; Ahram, M.; Linehan, W.M.; Knezevic, V.; Emmert-Buck, M.R. *Nat. Genet.* **2002**, *32*, 509-514.

(17) Stein, J.; Liang, P. *Cell. Mol. Life Sci.* **2002**, *59*, 1235-1240.

(18) Liang, P. *Methods: A Companion to Methods in Enzymology* **1998**, *16*, 361-364.

(19) Matz, M. V.; Lukyanov, S. A. *Nucl. Acids Res.* **1998**, *26*, 5537-5543.

(20) Liang, P.; Pardee, A. B. *Curr. Opin. Immunol.* **1995**, *7*, 274-280.

(21) Lievens, S.; Goormachtig, S.; Holsters, M. *Nucl. Acids Res.* **2001**, *29*, 3459-3468.

(22) Polymenis, M.; Schmidt, E.V.; *Curr. Opin. Genet. Dev.* **1999**, *9*, 76-80.

(23) Neufeld, T.P.; Edgar, B. A. *Curr. Opin. Cell Biol.* **1998**, *10*, 784-790.

(24) Han, F.; Lillard, S. J. *Anal. Biochem.* **2002**, *302*, 136-143.

(25) Katsivela, E.; Hofle, M. G. *J. Chromatogr. A* **1995**, *717*, 91-103.

(26) Roberts, D. B.; Bertsch, W. *Journal of High Resolution Chromatography and Chromatography Communication* **1987**, *10*, 244-247.

(27) Ramos, L. S. *Journal of Chromatographic Science* **1994**, *32*, 219-227.

(28) White, R. L.; Wentzell, P. D.; Beasy, M. A.; Clark, D. S.; Grund, D. W. *Anal. Chim. Acta* **1993**, *277*, 333-346.

(29) Malmquist, G. *J. Chromatogr. A* **1994**, *687*, 89-100.

(30) Smit, H. C.; Heuvel, E. J. v. d. *Topics in Current Chemistry, Chemometrics and Species Identification*, Springer Verlag: Berlin, 1987, p 63.

(31) Forina, M.; Armanino, L. C. *Topics in Current Chemistry, Chemometrics and Species Identification*, Springer Verlag: Berlin, 1987, p 91.

(32) Garrett, R.; Grisham, C. *Biochemistry*, 2nd Edition, Harcourt Brace College Publishers, Florida, USA, 1999.

(33) Kim, Y.; Yeung, E. S. *J. Chromatogr. A* **1997**, *781*, 315-325.

(34) Barron, A.E.; Blanch, H.W.; Soane, D.S. *Electrophoresis* **1994**, *15*, 597-602.

(35) Xue, G.; Pang, H.-M.; Yeung, E. S. *Anal. Chem.* **1999**, *71*, 2642-2649.

(36) Berke, J.; Ruiz-Martinez, M.C.; Hammound, R.; Minarik, M.; Foret, F.; Sosic, Z.; Kleparnik, K.; Karger, B. *Electrophoresis* **2003**, *24*, 639-647.

(37) Shi, L.; Khandurina, J.; Ronai, Z; Li, B.; Kwan, W.K.; Wang, X.; Guttman, A. *Electrophoresis* **2003**, *24*, 86-92.

(38) Aske, N.; Kallevik, H.; Sjoblom, A. S. *Energy & Fuel* **2001**, *15*, 1304-1312.

(39) Arinaga, M.; Noguchi, T.; Takeno, S.; Chujo, M.; Miura, T.; Uchida, Y. *Cancer* **2003**, *97*, 457-464.

(40) Mittermayr, C. R.; Tan, H. W.; Brown, S. D. *Appl. Spectrosc.* **2001**, *55*, 827-833.

(41) Round, A. J.; Aguilar, M. I.; Hearn, M. T. W. *J. Chromatogr. A* **1994**, *661*, 61-73.

(42) Grung, B.; Kvalheim, O. M. *Anal. Chim. Acta* **1994**, *304*, 57-66.

(43) Wang, C. P.; Isenhour, T. L. *Anal. Chem.* **1987**, *59*, 649-654.

(44) Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr. A* **1998**, *805*, 17-35.

(45) Vogels, J. T. W. E.; Tas, A. C.; van den Berg, F.; van der Greef, J. *Chemometrics and Intel. Lab. Sys.* **1993**, *21*, 249-258.

(46) Belenky, A.; Smisek, D. L.; Cohen, A. S. *J. Chromatogr. A* **1995**, *700*, 137-149.

(47) Pang, H.-M.; Pavski, V.; Yeung, E. S. *J. Biochem. Biophys. Meth.* **1999**, *41*, 121-132.

(48) Pang, H.-M.; Yeung, E. S. *Nucl. Acids Res.* **2000**, *28*, e73, i-viii.

(49) Hashimoto, M.; He, Y.; Yeung, E. S. *Nucl. Acids Res.* **2003**, *31*, No. 8, e41.

FIGURE CAPTIONS

Figure 1. 10-bp ladder separation using 1.5% MW 8,000,000 and 1.4% MW 600,000 PEO solution in 1□ TBE, 7 M urea. E = 250 V/cm. Injection was done at 75 V/cm for 40 s.

Figure 2. Original electropherogram from human kidney total RNA. Separation conditions were the same as in Fig. 1. Arrows span the region where reproducible results can be obtained.

Figure 3. Aligned electropherograms of triplicate injections from each of two RT reactions (offset top and bottom) of human breast total RNA were overlaid.

Figure 4. (a) Overlay of one electropherogram from kidney sample (green) on top of 10 electropherograms from breast samples (black); (b) overlay of one electropherogram from breast sample (black) on top of 8 electropherograms from kidney samples (green); (c) overlay of one electropherogram from breast tumor (red) on the top of 8 kidney samples (green); (d) overlay of one electropherogram from breast (black) on top of 8 breast tumor samples (red).

Figure 5. Score plot from PCA. Spots from breast tumor are red; spots from kidney are green; and spots from normal breast are black.

Figure 6. Score plot from PLS-DA. Spots from breast tumor are red; spots from kidney are green; and spots from normal breast are black.

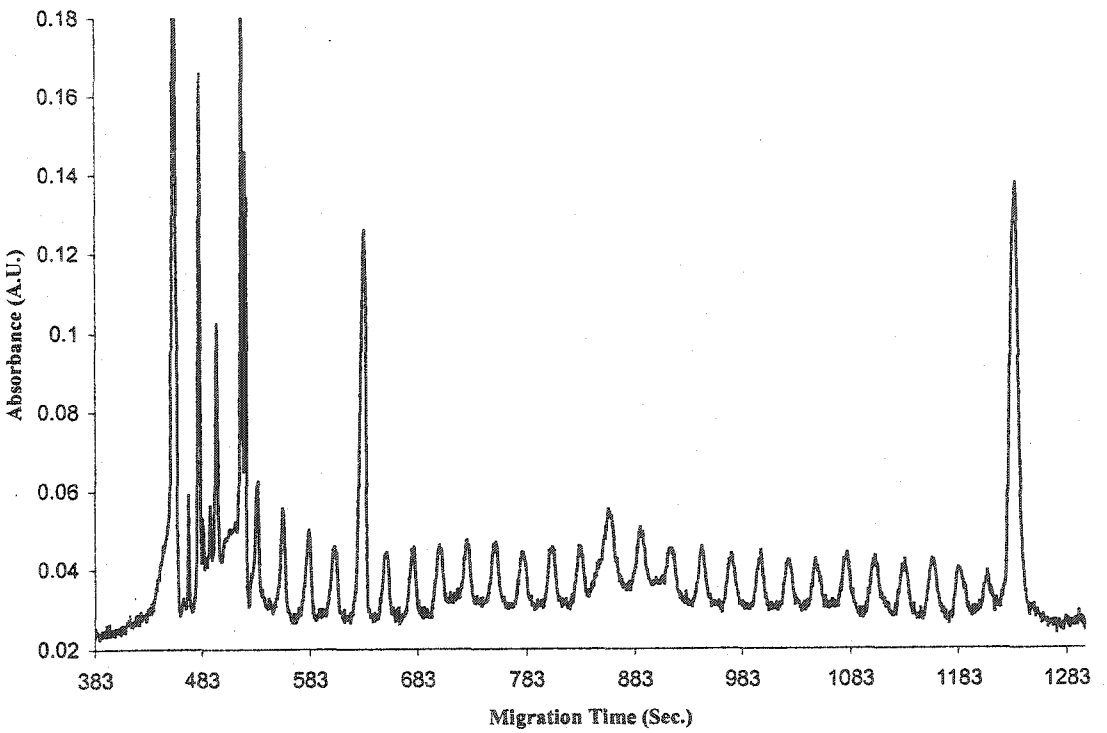

0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
Absorbance (A.U.)
383
483
583
683
783
883
983
1083
1183
1283
Migration Time (Sec.)

FIGURE 1

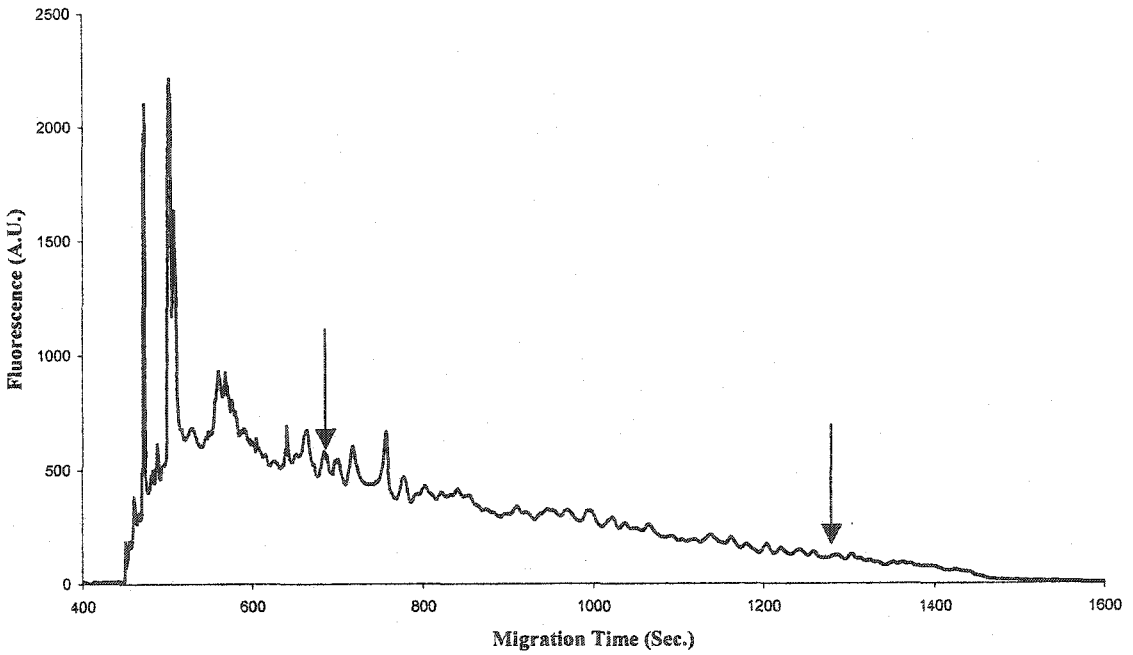

2500
2000
1500
1000
500
0
Fluorescence (A.U.)
400
600
800
1000
1200
1400
1600
Migration Time (Sec.)

FIGURE 2

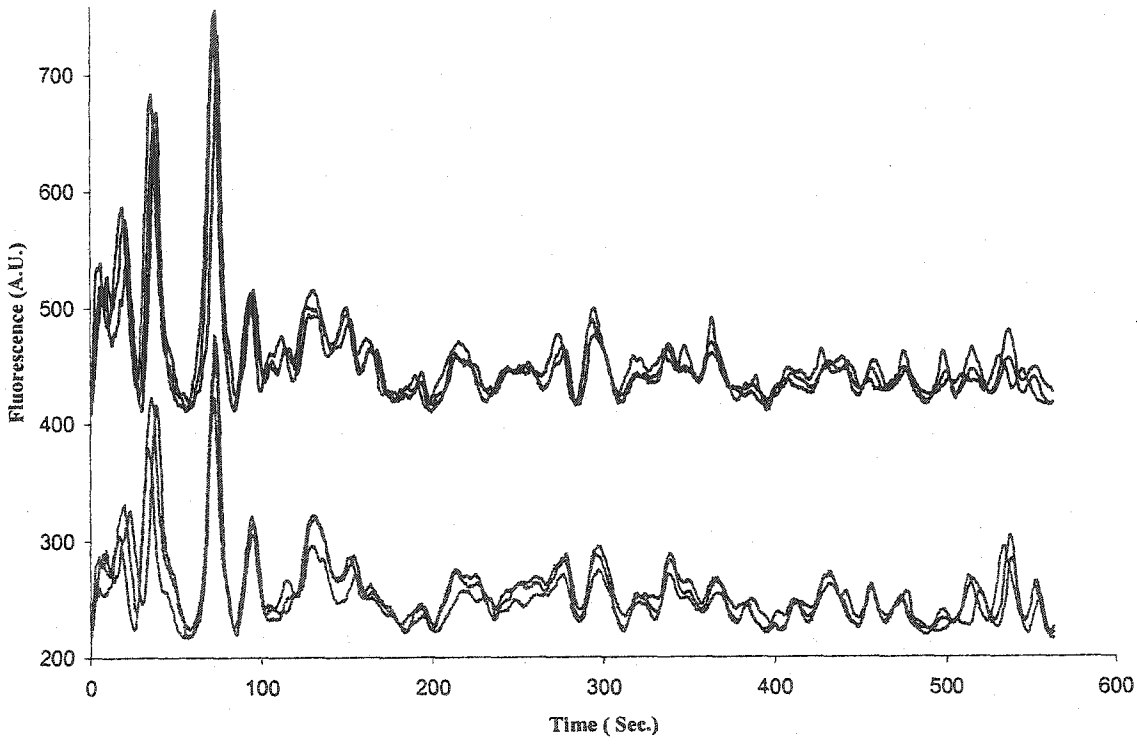

700
600
500
400
300
200
Fluorescence (A.U.)
0
100
200
300
400
500
600
Time (Sec.)

FIGURE 3

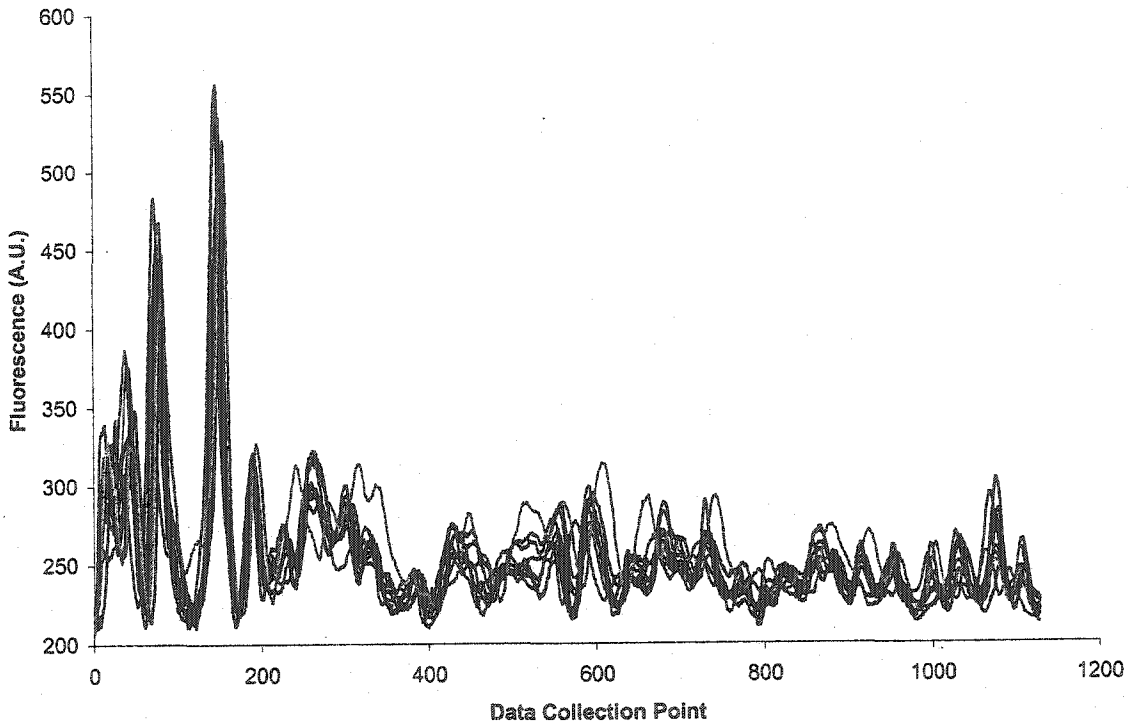

600
550
500
450
400
350
300
250
200
Fluorescence (A.U.)
0
200
400
600
800
1000
1200
Data Collection Point

FIGURE 4a

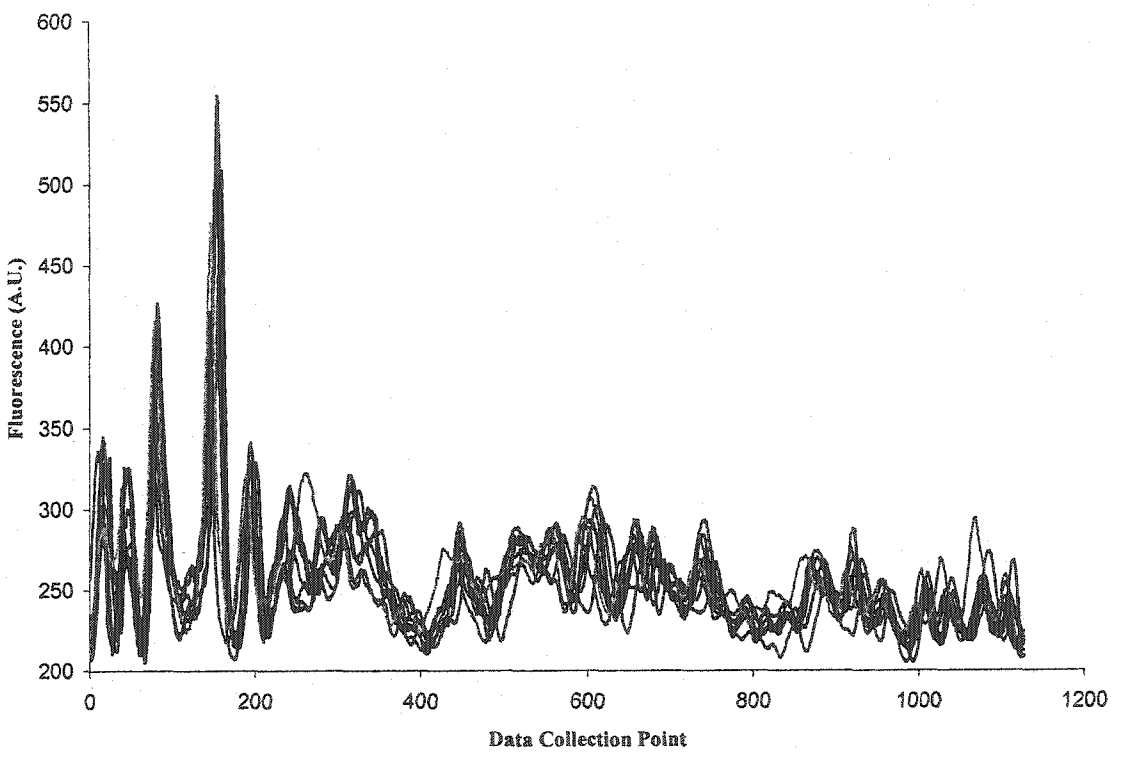

600
550
500
450
400
350
300
250
200
Fluorescence (A.U.)
0
200
400
600
800
1000
1200
Data Collection Point

FIGURE 4b

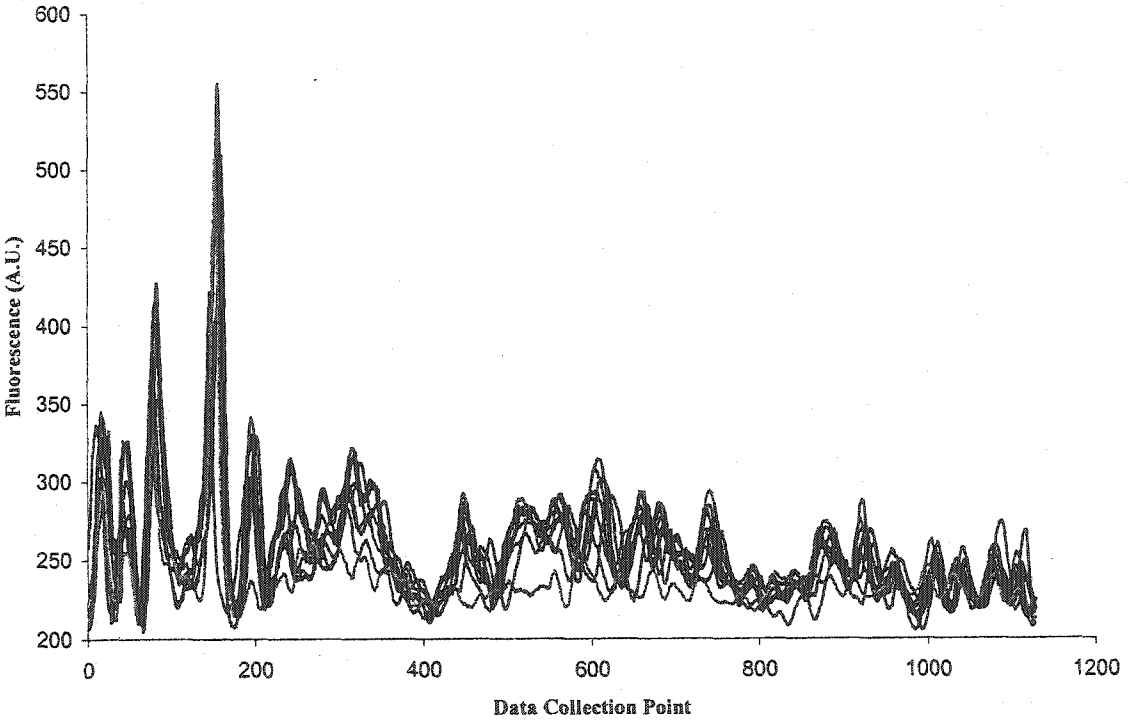

600
550
500
450
400
350
300
250
200
Fluorescence (A.U.)
0
200
400
600
800
1000
1200
Data Collection Point

FIGURE 4c

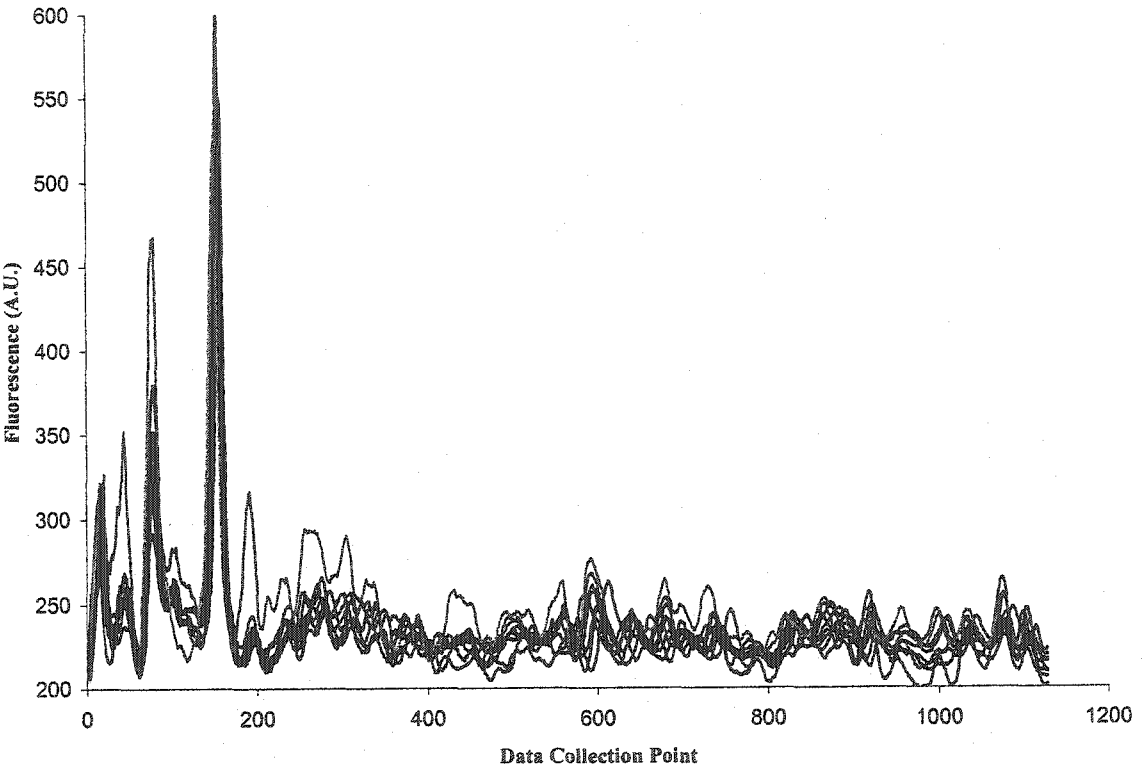

600
550
500
450
400
350
300
250
200
Fluorescence (A.U.)
0
200
400
600
800
1000
1200
Data Collection Point

FIGURE 4d

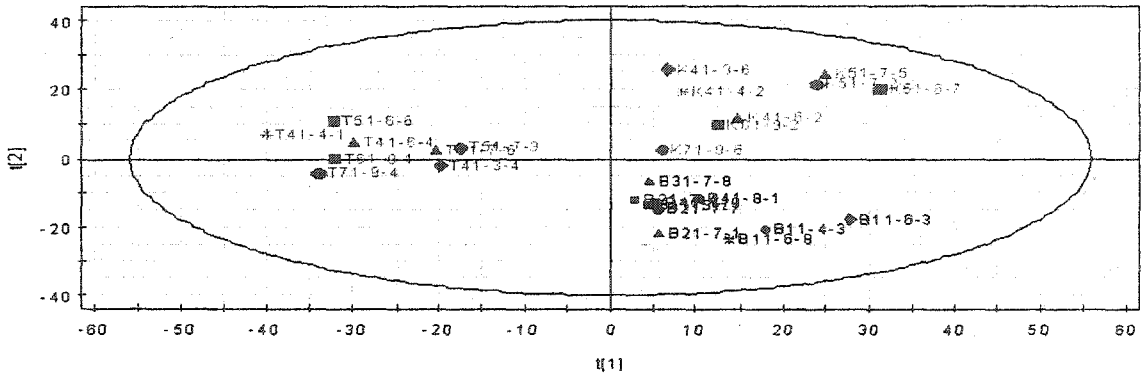

40
20
0
-20
-40
t[2]
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
t[1]
T51-6-6
T41-4-1
T41-6-4
T41-3-4
T71-9-4
B31-7-8
B11-6-3
B11-4-3
B21-7-1

FIGURE 5

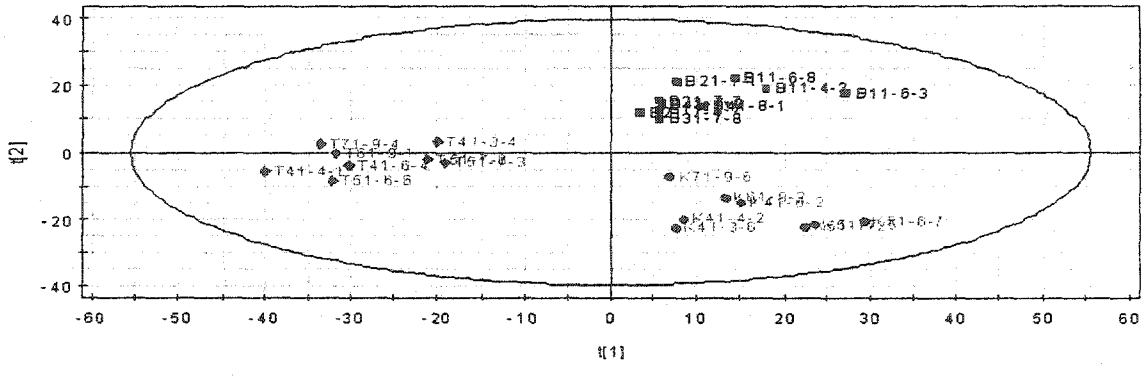

40
20
0
-20
-40
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
t[2]
t[1]
B11-6-8
B11-4-2
B11-6-3
T41-3-4
T41-6-2
T51-6-6
K41-3-6

FIGURE 6

CHAPTER 5. GENERAL CONCLUSIONS

High efficiency and simple approaches with low cost are the ultimate goals in developing new analytical methods to solve practical problems. The enormous research effort devoted to CE was evolved in more and more capabilities in diverse fields. CE can be performed with low operational and instrumental cost, has very high separation speed, efficiency and resolution, requires very small sample and buffer consumption, and can be run in a parallel manner easily. It is an ideal technique to be employed in developing cost effective, efficient and simple analytical methods. Here we demonstrated three new approaches of applying CE in DNA sequencing, chiral separations, and gene expression profiling.

Firstly, we showed that DNA sequencing could be done with lower cost using UV absorption detection in a capillary array than with using LIF detection. Since DNA has strong UV absorption itself, no dye is needed in our scheme to label the DNA fragments. Up to 100 bp DNA sequence was read after separating the products from four termination reactions in four individual capillaries and adjusting the migration time variation caused by the slight differences in gel geometry, running conditions, and surface chemistry of capillary column among different capillaries. The absorption detection scheme is more straightforward and the operational and instrumental cost is much lower than using lasers. This technique can be employed in gene mutation detection, antisense drug analysis, and other fields that require detection of the sequence of short DNA fragments. Longer fragment sequencing can also be done in this way if the reaction protocol is modified to produce enough DNA for detection, or the absorption background of the polymer solution is significantly lowered.

Secondly, a universal chiral screening system was designed using only four conditions to separate 49 out of 54 selected compounds. The chosen compounds have diverse structures and varied pK values, so that the system should be suitable for the separations of a large population of common enantiomers, and it is very useful in finding the right separation condition of the unknown compounds. Also, it can help to elucidate the chiral recognition mechanism based on the results. Even if the compounds may not be base-line resolved in our system, some knowledge about the more suitable chiral selector and separation condition should be gained. Further refinement of the system can be done by including CDs with the same substitution groups but different cavity sizes, and different CD concentrations. For some specific chiral compounds, other types of selectors can be employed using the same strategy in developing corresponding screening conditions. Further more, with the help of the capillary array system, theoretical study of the chiral separation mechanism becomes easier by employing gradually changing conditions, such as pH values or chiral selector concentrations, in one single run. The change of mobility of the analytes can be used to calculate the equilibrium constant of the complexation and the optimum chiral selector concentrations.

The third project was focused on applying capillary electrophoresis to perform RNA expression profiling. The electropherograms obtained from the total RNA of three types of human tissues contained significant profile difference, which revealed the overall changes of RNA expression among tissues. PCA and PLA-DA were employed for objective pattern recognition. This technique can serve as a complementary method to the existing ones in gene expression profiling, and is a simpler, more cost effective, and faster prescreening method for recognizing gene expression changes under different circumstances. If more

replicate experiments are performed for each type of tissue, a statistical model can be built for that specific tissue. Based on the models, we can perform prediction analysis on unknown samples. This would be very useful in differentiating diseased tissues from healthy tissues. Also multivariate data analysis can help to identify the variables, in our case the fluorescent signals collected at each data points, which contribute more to the profile differences. Fraction collection can be done subsequently and further DNA sequencing on the collected fragments should be able to disclose the genetic reasons of the changes. The part of the project we completed here started with total RNA sample. The fluorescent signals are mostly from tRNA, rRNA and most abundant mRNAs. In order to achieve more useful information of most of the genes expressed in different levels, further study should be devoted to the RNA only. Another advantage of our method is the low sample consumption. It can be further reduced if we decrease the RT reaction volumes, so that less than 1 μg RNA sample is enough for one experiment.

ACKNOWLEGEMENTS

First of all, I would like to express my sincere appreciation to my supervisor, Dr, Edward S. Yeung, who intrigues my continuously growing interest in science, teaches me to think boldly but act cautiously while conducting research, enlightens me with his dazzling ideas and creative thoughts, and encourages me never to be stopped by the obstacles on the way to success. Being an inspiring and patient mentor, he is also a brilliant scientist and a very decent person who I can always look up to. The experience I gained and the things I learnt during the last five years working under his guidance are priceless treasures for me, and will benefit my whole life.

Also I am very grateful to both my past and present committee members, Dr. Houk, Dr. Porter, Dr. Beitz, Dr. Woo, Dr. Small, and Dr. Ng for their precious time and advice.

The third party I own a great deal to is the whole Dr. Yeung's group. They have been so nice to me and so helpful all the time. I would like to thank Dr. Jinjian Zheng, Dr. Wei Wei, Dr. Yan He, Dr. Channan Sluvichi, Dr. Weihua Huang, Dr. Yinfa Ma, Dr. Park, Dr. Kang, Dr. Michael Shortreed, and Dr. Homing Pang for their assistance to my research, thank Jason Gruenhagen, Frank Lee, Michael Christodoulou, Hui Su, Xiaoyi Gong, Yan He, Hungwing Lee, and Guoxin Lu for the intriguing discussions, endless help, great friendship, and the wonderful time I spent with them. I am especially thankful to Hanlin, Gang Xue, and Lianjia Ma for all the useful advice, continuously encouragement, and friendly support to my work and in my personal life. I am very lucky to be a member of such a great family, I will never forget any of our group members, and I really wish we could be friends forever.

Finally, I would like to dedicate this dissertation to my families. I am very proud of my parents. It is my dear mother who always encourages me to pursue higher goal in my life, continuously uses her words and real action to impact me to be a better person, and always stands behind my back to support me with her strong hands. It is her expectation that pushes me to achieve all I have today, and I am so glad that I did not give up on half way to my weak mind and laziness because of her encouragement. It is my dear father who gives me strength and wisdom, tolerates all of my bad behaviors, and teaches me the importance of being honest, respectful and selfless. Together with my dear sister and brother, my parents give me a very healthy and happy family, provide me with endless love, and create a cozy home to go to whenever I feel tired and stressful. At the end, I would like to thank my dearest friends, who love me and support me without any condition, and whom I see as part of my family. I am so fortunate to have them in my life and millions thanks are not enough to show my appreciation. It is my families and dear friends that make my life fulfilled.

This work was performed at Ames Laboratory under Contract No. W-7405-Eng-82 with the U.S. Department of Energy. It was supported by the Director of Science, Office of Biological and Environmental Research, and by the National Institutes of Health.