ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS USING MOLECULAR AFFINITY SEPARATION AND MASS SPECTROMETRY

by

ASHISH MISRA

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

written under the direction of

Dr. Sobin Kim

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2009

ABSTRACT OF THE DISSERTATION

Analysis of Single Nucleotide Polymorphisms Using Molecular Affinity Separation and Mass Spectrometry

by Ashish Misra

Dissertation Director:

Sobin Kim

Single nucleotide polymorphisms (SNP) constitute the most abundant human genetic variations and are important markers for studying interindividual variability with many different techniques having being developed for their study. The use of mass spectrometry (MS) is an attractive detection method for SNP genotyping due to label free detection based on molecular mass, especially for use in diagnostic applications that would require analysis of tens of SNPs in cohorts of individuals. The work in this thesis explores the use of a molecular affinity purification system for improving multiplexing levels of SNP genotyping using MS based detection and approaches for improving its throughput.

The strong molecular affinity between biotin and streptavidin has been employed for isolation of biotinylated oligonucleotides before analysis by MS using the previously developed solid phase capture-single base extension approach. We have been able to genotype up to 50 SNPs simultaneously using two genes from the cytochrome 450 family of genes as model system. These results have demonstrated the utility of the biotin-streptavidin affinity system for highly multiplexed SNP genotyping using MS. Following this, two

approaches have been employed to reduce processing time and improve throughput of the technique. First, we have used monomeric avidin coated microbeads to fabricate a device that leads to a substantial reduction in processing time for the isolation step to ~2 hours, and allows simultaneous processing of multiple samples for genotyping a limited number of SNPs. Additionally, the microbead device can be reused 5 times with a simple regeneration protocol thus acting as a low cost tool for enhancing sample cleanup prior to MS. A second approach involves the use of heat and water for breaking the biotin-streptavidin interaction that allows direct analysis of released fragments by MS. We have shown its utility for reducing processing time after the isolation step substantially, and used it for highly multiplexed SNP genotyping. In concert, these studies demonstrate the feasibility of using the molecular affinity interaction between biotin and (strept)avidin for high throughput genotyping of single nucleotide polymorphisms using MS.

ACKNOWLEDGEMENTS

I would like to begin by thanking my advisor Sobin for giving me the opportunity to work with her. She has provided direction to my research and helped me stay focused through a lot of testing times. I am grateful for her encouragement through my stay in the lab.

I would like to thank my committee members for their feedback that provided new avenues to explore during the course of this dissertation. I would have liked to pursue more of the ideas that were proposed but there was not enough time.

I have been fortunate to have many friends for suggestions and feedback with research. The frustrations in the lab have been less vexing due to discussions with many friends, especially Sumati, Aina, Salah, Anand, Bill, Ana, Nripen, Natesh, Larry, Alex, Eric, Norman, Dan, Carolyn, Lavanya, JP, and other acquaintances. I am grateful for their company hanging out in the lab after hours and over the weekends. I would like to thank the following people in particular for help and guidance with experiments - Sumati for her guidance and support through my graduate work, Madesh and Ravi for teaching me about PCR that I could not learn from any book, Salah for teaching about the importance of experimental protocol and organization and help with DNA extraction, Larry and the Zahn group for providing microfluidic devices and Anand for help with protein binding and running protein gels.

Additionally, I would like to thank the Zen group, especially sensei Kurt for his constant support and wisdom. I wish I could have gone to more sesshins and completely forgotten about research for a few days.

Above all, I would like to thank my parents and sisters, Archana and Alpana, for their support through my graduate school years. I have received constant encouragement from

their side which has always cheered my spirits, especially my mother who would offer me suggestions though she could never understand the exact details of my work.

TABLE OF CONTENTS

LISTS OF TABLES

CHAPTER 2

CHAPTER 3

LIST OF ILLUSTRATIONS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1

INTRODUCTION

1.1 Background

The human genome is composed of around 3.2 billion base pairs and majority of the sequence is conserved across the human population (1, 2). Genetic variation between any two individuals is constituted by less than 0.1 % of the bases in the genome and a large number of these variations have been found with the completion of the human genome project (1, 3-6). It is thought that many of the variations arose early in human evolution and have since become fixed in the population. Such variations that are commonly found in the population are referred to as polymorphisms and are conserved across generations (7, 8).

Genetic polymorphisms are responsible for the majority of interindividual differences and serve as important genetic markers for determining the genetic component of variability (9). The vast majority of these polymorphisms are silent and do not lead to changes in phenotype. However, some of the polymorphisms that affect the phenotype can lead to disease, susceptibility to disease or other imbalances (4, 10). The study of genetic polymorphisms for understanding their effects on phenotype requires the use of genetic markers that are easy to identify and the development of high throughput methods for reducing time and cost associated with their analyses (8).

1.2 Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are single base changes that occur at specific locations in the human genome and are by definition, found in more than 1 % of the population. They are the most commonly occurring polymorphisms and constitute ~90% of genetic variation. They are found throughout the genome with an average frequency of 1 in every 300-1000 bp and are estimated to be around 10 million in number (11). SNPs are not distributed uniformly across the genome and their frequency can vary several-fold between different regions of the genome (5, 12). Distribution of SNPs across the genome is important because of the effect SNP location has on phenotype. In general, SNPs occur much less frequently in coding regions of the genome than in noncoding regions, with most SNPs being located in noncoding regions (4, 10). SNPs in noncoding regions, although they do not alter encoded proteins, serve as important genetic or physical markers for comparative or evolutionary genomics studies. SNPs, when present in regulatory sites of a gene, can affect rates of transcription causing changes in the production of encoded protein. In the coding regions, SNPs can cause alterations in protein structure and hence function, leading to the development of disease or change in response to a drug or environmental toxin.

1.3 Importance of SNPs as Genetic Markers

SNPs are stable polymorphisms and preserved in populations over generations making them ideal markers for association studies (13). SNPs are usually biallelic making it easier to genotype them though in some instances, tri-allelic and tetra-alleic variants have also been found. They are commonly used in genetic association studies for correlating genotypes (genetic makeup) and phenotypes (observed trait) (14, 15). Such studies can be broadly classified into disease association studies and pharmacogenomic studies (16, 17). The goal of the former is to identify SNPs that are associated with a disease state or susceptibility to a

disease while the latter aim to uncover SNPs in genes that lead to differential drug metabolism in individuals (18, 19).

In disease association studies, if a phenotype is known to be associated with some protein(s), SNPs in corresponding gene(s) can be investigated to determine their effect on the phenotype. Such approaches for SNP characterization typically involve studying SNPs in a limited number of target genes and are hence termed candidate gene studies (16, 20). When there is no knowledge of the underlying genetic cause for a particular phenotype, studies are performed to identify the gene or genetic locus that is associated with the phenotype. These involve studying SNPs across the genome either using familial inheritance or performing a case-control study with a set of individuals that exhibit the phenotype (case) and another set that does not (control) (21). SNPs that are found to be significantly associated with the phenotype being investigated help in identifying role of the corresponding genomic region.

Another major application where SNPs are utilized as markers is in pharmacogenomic studies (19, 22-24). In such studies, the target is to elucidate effects of genetic polymorphisms on drug responses. Therefore, only patients who have been administered a specific drug can participate in pharmacogenomic studies, and clinical trials are the only source of samples, unlike population genetic approaches that can use homogeneous populations or familial inheritance. This presents additional challenges in the design of the study. However, despite the obstacles, pharmacogenomic application of SNP genotyping is becoming increasingly popular. One cause for this growing popularity is the strong demand for personalized medication, as adverse drug reactions (ADRs) are a major clinical problem that occur frequently and can sometimes be fatal. Only a handful of genes are involved in the

metabolism of drugs; hence pharmacogenomic studies are usually restricted to these sets of genes (25). Isoenzymes from the cytochrome P450 (CYP450) superfamily are especially important in such studies as they are involved in the phase 1-dependent metabolism of a large number of xenobiotic chemicals including 70-80 % of all therapeutically important drugs (26). The human CYP450 superfamily consists of 57 genes organized in several families and subfamilies based on protein homology, of which, drug-metabolizing genes are from families 1, 2 and 3 (27). Many of the characterized genes have been found to be highly polymorphic and many of the polymorphisms have been found to be responsible for differential activity of the corresponding enzyme (28, 29).

In both disease genetics and pharmacogenomic studies, a large number of SNPs need to be genotyped in equally large sample populations (30, 31). Hence, achieving a high level of throughput has been a critical factor in recently developed genotyping assays (32-34). A typical approach for improving the throughput of genotyping is to increase the number of SNPs that are analyzed simultaneously, in other words the multiplexing level of the assay (33-35). Multiplexing, therefore, is a major element in many current SNP genotyping technologies in addition to cost of genotyping and stringent requirements for accuracy.

1.4 SNP Genotyping Techniques

Genotyping typically involves generation of allele specific products for SNP of interest followed by their detection to determine the genotype. Several techniques have been developed to decrease cost and time and improve accuracy for genotyping SNPs (36, 37). All current genotyping technologies with only a few exceptions require the PCR amplification

step. In most techniques, PCR amplification of desired SNP containing region is performed initially to introduce specificity and increase number of molecules for detection following allelic discrimination (33). Following PCR, two steps are used for determining SNP genotype: generation of allele specific products and their detection. The first step discriminates between alleles at a SNP site by producing allele specific products while the second step is used for characterizing these products to reveal SNP identity. SNP genotyping techniques can be distinguished on the basis of strategy employed for generation of allele specific product, type of detection strategy and level of throughput achievable with the technique (38).

Allele discrimination is typically performed using allele-specific biochemical reactions, and there are four popular methods: primer extension (nucleotide incorporation), hybridization, ligation and enzymatic cleavage (Figure 1.1). Primer extension approaches involve allele-specific incorporation of nucleotides in primer extension reaction with DNA template, utilizing enzyme specificity to achieve allelic discrimination (39-41). Hybridization approaches use differences in thermal stability of double stranded DNA to distinguish between completely matched and mismatched target-probe pairs for achieving allelic discrimination (42-44). Ligation approaches employ specificity of ligase enzymes that join two oligonucleotides hybridized to single stranded template DNA to form a single oligonucleotide under certain specific conditions which can be used for allelic discrimination (45-47). Enzymatic cleavage for allele discrimination is based on the ability of certain classes of enzymes to cleave DNA by recognition of specific sequences and structures. Such enzymes can be used for discriminating between alleles when SNP sites are located in an

enzyme recognition sequence and allelic differences affect recognition (48, 49). Many techniques use a combination of these methods to improve discrimination specificity (36, 50).

After allelic discrimination, SNP genotype is determined by detection of allele specific products most often using fluorescence, chemiluminescence or mass spectrometry (32, 37). Monitoring fluorescence signals is widely used in current genotyping because of its simple implementation and rapid and sensitive detection. It is the most widely used detection methodology used in SNP genotyping techniques and a large number of detection platforms have been developed based on fluorescence (37, 51). It is most commonly used for direct sequencing (DS) using capillary array electrophoresis, which has been the workhorse for the Human Genome Project and most of the initial SNP identification studies (52). For very high throughput studies involving genotyping of large numbers of SNPs in multiple samples, array based fluorescence detection is preferred (53). Chemiluminescence has several advantages as a detection technique, such as high signal-to-noise ratio, rapid detection, and feasibility for automation and has been used successfully in Pyrosequencing™ (Biotage, Sweden) – a sequencing by synthesis approach (54). The approach has proven to be suitable for quantitative SNP genotyping in pooled DNA samples, but is somewhat limited in multiplexing for SNP analysis. Mass spectrometry determines the molecular mass of a molecule, which is an intrinsic property and does not require any additional labels for detection. As a result, it has less likelihood of error compared with most other detection methodologies that use indirect measurement and is hence an attractive format for the detection of allele specific products (50).

1.5 SNP Genotyping Using MS

Of the different techniques for MS detection, matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) has emerged as a high throughput detection platform for SNP genotyping due to short analysis time and capacity for quantitation (50, 55, 56). It is widely used for the analysis of various biomolecules such as DNA oligonucleotides, peptides, proteins, and oligosaccharides (57, 58). It involves use of a small organic molecule called matrix that absorbs energy from a laser source of a certain wavelength for ionization. When analytes are mixed and co-crystallized with matrix, they are ionized in the form of intact molecules due to transfer of energy from the matrix molecules. Resulting ions are accelerated by a voltage gradient through a field free region and separated by time of flight of each ion. The time of flight depends on the mass and charge of an ion and MALDI-TOF MS detects the mass to charge ratio of molecules. Predominantly, single charged analyte ions are produced during ionization with matrix, which makes spectral analysis easier. Using MALDI-TOF MS, a mixture of many oligonucleotides with differing masses can be rapidly separated and accurately analyzed without the need for specific detection labels.

Of the various allelic discrimination strategies, methods based on primer extension are most widely used with MALDI-TOF MS due to high accuracy for SNP genotyping and scope for analyzing multiple SNPs simultaneously (59). PinPoint assay (60), MassEXTEND™ (61, 62), SPC-SBE (63), and GOOD assay (64) are all primer extension based methods that use MALDI-TOF MS for the detection of discriminated alleles. In these methods, SNP specific primers are simultaneously extended with various nucleotides using PCR products as template to yield extension products of different masses. Each extension product corresponds to one of the alleles of each SNP and subsequent mass analysis reveals SNP genotype (Figure 1.2). The Pinpoint assay uses ddNTPs for single base extension (SBE) of

primer and is the simplest approach among the primer extension methods (60, 65). It does not involve any chemical modification of nucleotides or primers and has demonstrated the potential of MALDI-TOF based methods for multiplex SNP detection. Reduced accuracy due to low spectral peak resolution, especially in the detection of A–T heterozygous SNPs, has been the major limitation of this approach. This limitation has been overcome by using mass tagged ddNTPs to increase mass difference between alleles (66). The MassEXTEND™ method (Sequenom, CA) uses a mixture of regular nucleotides (dNTPs and ddNTPs) for primer extension (61, 62). This increases mass difference between extension products corresponding to both alleles of a SNP and improves resolution of mass spectra leading to enhanced accuracy in genotyping. MassEXTEND™ has been incorporated in an automated manner on the MassARRAY™ (Sequenom, CA) platform for high-throughput genotyping. The GOOD assay uses primers that are phosphorothioate-modified at their 3' ends for SBE in the presence α-S-ddNTPs (64). This is followed by digestion of unmodified part of the primer, neutralization of the phosphorothioates yielding charge-tagged DNA fragments. The charge-tagging procedure eliminates need for sample purification before MS analysis and use of shorter extension fragments improves spectral resolution. The assay has been recently modified to allow use of regular ddNTPs and PCR products without cleanup of excess dNTPs for SBE reaction (67)

A more recently developed approach utilizes the strong affinity of biotin for streptavidin for solid phase capture (SPC) of biotinylated oligonucleotides using streptavidin coated solid surface before release and analysis by MS (63). The SPC-SBE (solid phase capture-single base extension) approach provides removal of primers prior to mass spectrometry analysis by solid phase capture of extension products after SBE (Figure 1.3). Briefly, it involves PCR

amplification of desired genomic regions followed by inactivation of dNTPs and PCR primers. Primer extension reaction is performed with ddNTP terminators attached to biotin resulting in extension products carrying a biotin moiety at their 3' ends. Biotinylated extension products are isolated by affinity purification on a streptavidin-coated solid surface. They are then desalted and analyzed using MALDI-TOF MS to yield a spectrum containing peaks of extension products only. Genotypes are inferred by comparing masses of peaks from the spectrum with expected masses of extension products. Elimination of peaks corresponding to unextended SBE primers from resulting mass spectrum allows detection of larger numbers of extension products, leading to higher levels of multiplexing.

Of the current mass spectrometry based methods for genotyping single nucleotide variations including SNPs, SPC-SBE offers the highest level of multiplex genotyping. Previously, it has been used for highly multiplexed screening of mutations. We explore the use of SPC-SBE for highly multiplexed SNP genotyping using genes from the cytochrome P450 family as model system. Further, current implementation of the SPC-SBE protocol is slow making it unsuitable for high throughput, large-scale studies. Improvements in the technique to reduce processing time will increase throughput and in combination with high multiplexing levels, it should facilitate analysis of larger number of single nucleotide variations. In concert, these studies will establish feasibility of the use of molecular affinity based separation in combination with MS for large scale analysis of SNPs.

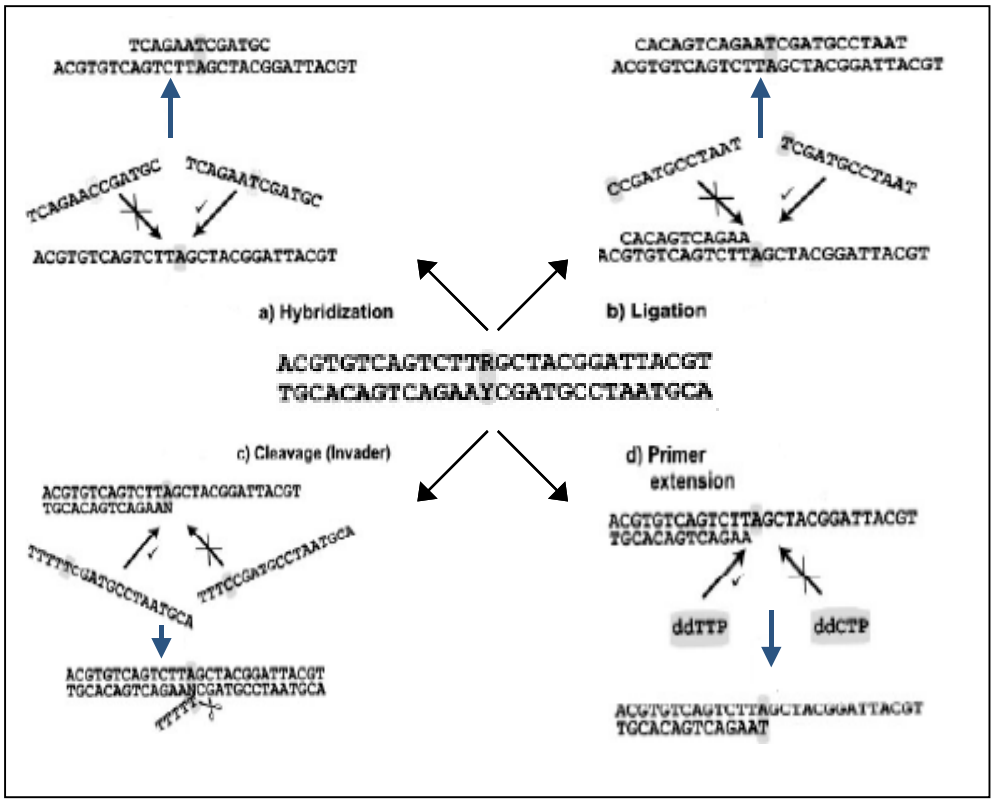

Figure 1.1. Approaches for generation of allele specific products for use in SNP genotyping. *(Illustration from Ref 50)*

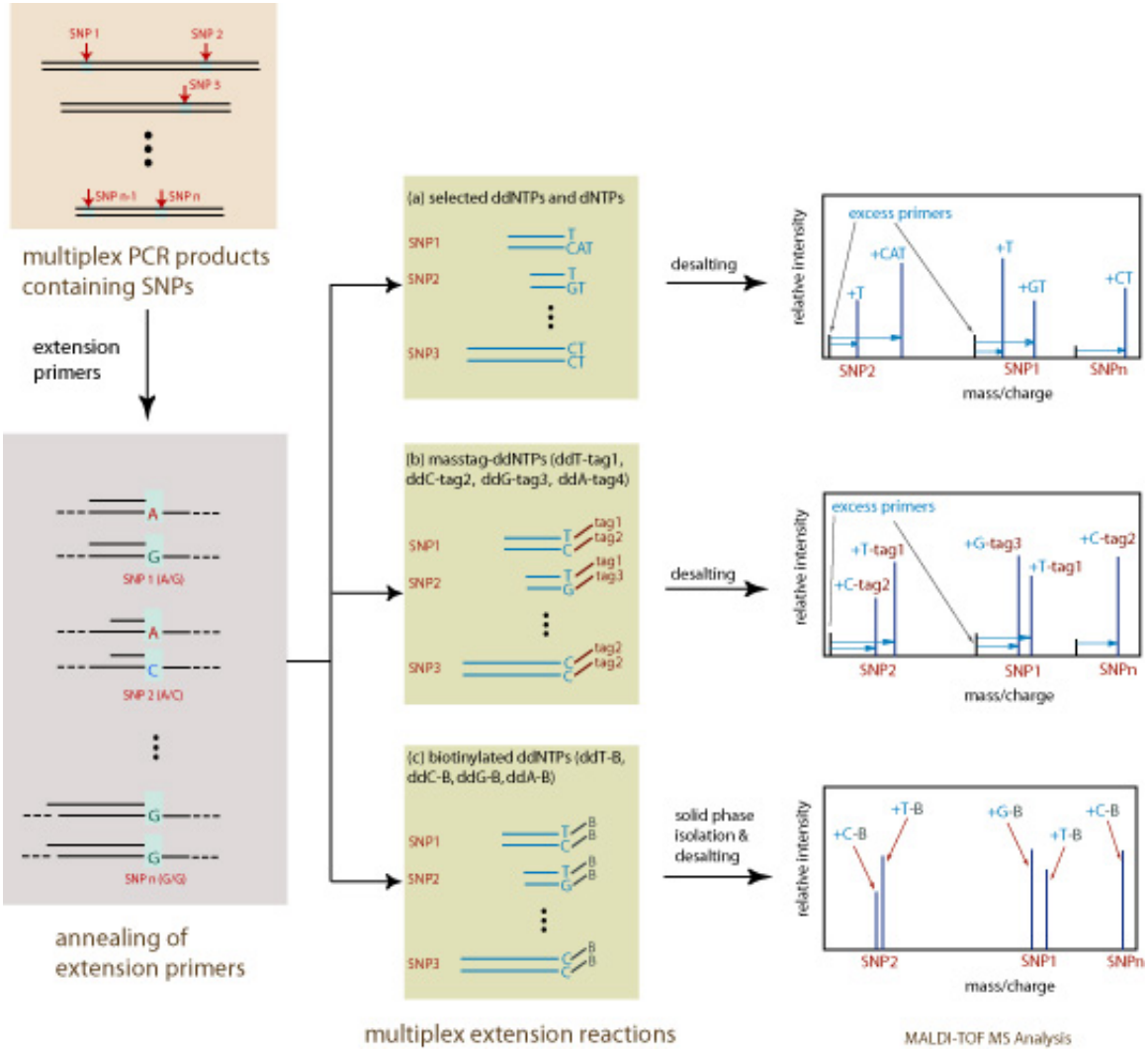

Figure 1.2. Multiplex SNP genotyping by primer extension and MALDI-TOF MS. (a) MassEXTEND™ approach in which primers are extended by one or more bases, specific to each allele. (b) Modified Pinpoint approach using mass-tagged ddNTPs for primer extension. This increases mass difference between extension products of alleles of a SNP, when compared with regular ddNTPs, enhancing accuracy of genotyping. (c) SPC-SBE approach involving generation of primer extension products with biotinylated ddNTPs for isolation on a solid surface.

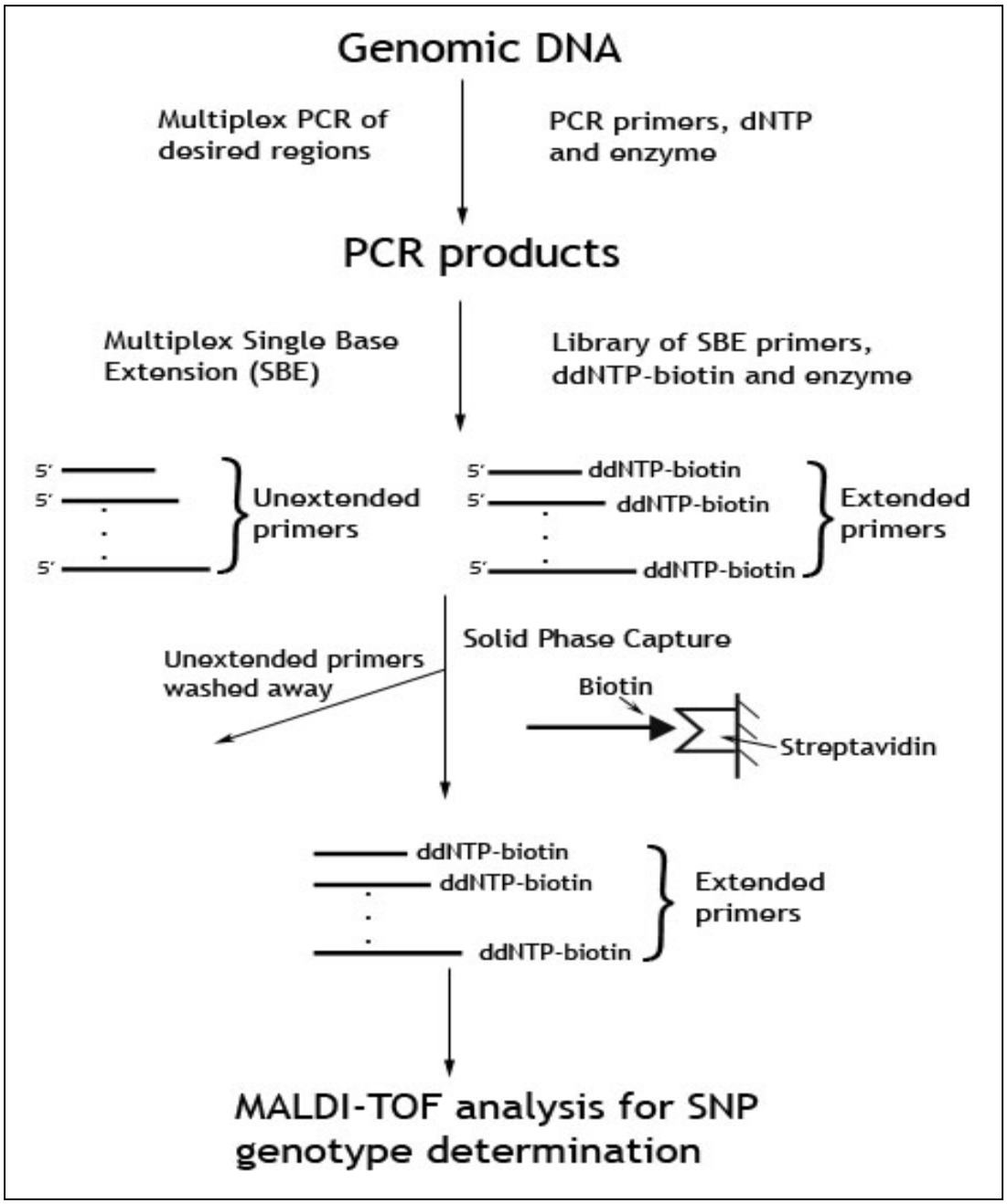

Figure 1.3. SPC-SBE approach for genotyping SNPs. SNP-containing genomic regions are amplified and used as template for extension of SBE primers using biotin-ddNTPs. Extension products are separated from unextended primers on a solid surface followed by MALDI-TOF MS analysis to determine SNP genotypes.

CHAPTER 2

INVESTIGATION OF HIGHLY MULTIPLEXED SNP GENOTYPING WITH MOLECULAR AFFINITY SEPARATION USING CYP450 SNPS

2.1 ABSTRACT

Here, we explore the solid phase capture-single base extension (SPC-SBE) approach for highly multiplexed SNP genotyping using two genes belonging to the CYP450 family (CYP2C9 and CYP2A13) as model system. Previously, high multiplexing capability of the SPC-SBE approach has been demonstrated by simultaneous screening of 30 frequent mutations in the p53 gene. In this study, assays for concurrent analysis of 40 SNPs of the CYP2C9 gene and 50 SNPs of the CYP2A13 gene were developed. Desired SNP containing regions for each gene were amplified in a single- step multiplex PCR reaction followed by genotyping of SNPs using SPC-SBE. 11 samples were analyzed for CYP2C9 and 14 samples for CYP2A13 with unambiguous detection of SNPs in all samples. Many samples showed high occurrence of heterozygotes for both genes with as many as 10 out of 50 SNPs appearing as heterozygotes in one sample genotyped for CYP2A13. These results strongly indicate that the system utilizing SPC-SBE provides an efficient means for genotyping SNPs involving CYP2C9 and CYP2A13. Furthermore, the approach presented here can be easily extended to other genotyping applications.

2.2 INTRODUCTION

Single-nucleotide polymorphisms (SNPs) are single-base changes that occur at specific locations in the genome at regular intervals of 300 to 1000 bases with a frequency of at least 1% in the population (68). SNPs comprise ~90% of genetic variations and are commonly used markers in association studies for correlating genotypes and phenotypes (14, 68). The majority of techniques for genotyping SNPs use two steps: generation of allele-specific products and their detection. (36-38, 69). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has emerged as a high-throughput detection platform for SNP genotyping because of its short analysis time and capacity for quantitation (50, 55, 70, 71). Different approaches have been coupled with MALDI-TOF MS for this purpose, including hybridization, cleavage, ligation, and primer extension (50). Of these, methods based on primer extension are most widely used because of their high accuracy for SNP genotyping and scope for analyzing multiple SNPs simultaneously (50, 56).

A new MALDI-TOF MS technique based on primer extension, solid-phase capture–single-base extension (SPC-SBE) can analyze larger numbers of nucleotide variations compared with conventional MALDI-TOF MS primer extension methods (72). The feasibility of multiplex SNP genotyping using SPC-SBE has been demonstrated by simultaneous screening of 30 point mutations in the *p53* gene (73). We report here SPC-SBE assays for multiplex genotyping of SNPs from 2 genes of the CYP450 family.

Cytochrome P450 (CYP450) isoenzymes are involved in the phase 1–dependent metabolism of 70%–80% of all therapeutically important drugs (26). *CYP2C9* is involved in the

metabolism of ~10%–15% of these drugs, including tolbutamide, S-warfarin, and many nonsteroidal antiinflammatory drugs (74, 75). The gene is highly polymorphic, with around 290 SNPs in the National Center for Biotechnology Information (NCBI) dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) and 30 alleles in the Human CYP allele database (http://www.cypalleles.ki.se) (76, 77). Many of these alleles result in altered catalytic activity of the corresponding protein for different substrates (75). *CYP2A13* is one of the less well characterized members of the CYP450 family and was initially thought to be nonfunctional. Only recently has its role in the activation of 2 important environmental carcinogens, tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and aflatoxin B1 been demonstrated (78, 79). Fewer than 70 SNPs of the gene have been reported in the NCBI dbSNP and Human CYP allele databases combined, and only a few of those have been functionally characterized. Since *CYP2A13* was characterized, there have been few studies involving SNPs of the gene using PCR single-strand conformation polymorphism or direct sequencing for SNP analysis (80). Because of wide applications in pharmacogenomics and environmental cancer studies, there is a need for genotyping SNPs of *CYP2C9* and *CYP2A13*. Here we demonstrate simultaneous genotyping of 40 SNPs of *CYP2C9* and 50 SNPs of *CYP2A13* in separate assays using the SPC-SBE approach.

2.3 MATERIALS AND METHODS

2.3.1 DNA Samples

For *CYP2C9* analysis, 11 DNA samples were used, of which 6 were human cell line based; the remaining 5 samples were from individual anonymous donors. For *CYP2A13* analysis, 14 samples were used, 5 human cell line–based samples and 9 individual anonymous donor samples. The procedure for use of DNA samples was approved by the Institutional Review Committees of University of Medicine and Dentistry of New Jersey and Rutgers University. The individuals who provided DNA samples gave written informed consent.

2.3.2 SNP Selection and SBE Primer Design

We selected SNPs for both genes from the NCBI dbSNP and Human CYP allele databases. We chose 40 exonic and intronic SNPs in the vicinity of exons 3, 6, and 7 for the *CYP2C9* gene and 50 SNPs spanning the entire gene including all exonic SNPs for *CYP2A13*. We designed SBE primers for each SNP with sufficient mass difference between successive primers to accurately perform multiplex genotyping and avoid overlap between double-charged ions of larger extension products and single-charged ions of smaller extension products using a PERL based script with the following algorithm (Tables 2.1 and 2.2). An initial set was generated containing all possible primers for each SNP having melting temperatures greater than 44°C and lengths 11-40. A library of SBE primers was selected from this set to include a primer for each SNP site ensuring no overlap between double charged ions of largest products and single charged ions of smaller products. Mass difference between successive primers was chosen to be greater than 104 Da to ensure good resolution

for all 4 possible extension products. For SNPs on adjacent bases, we selected a reverse primer to anneal to the complementary strand for detecting the latter SNP.

2.3.3 PCR Amplification

We designed PCR primers to amplify genomic regions containing the selected SNPs for both genes (Tables 2.3 and 2.4) using reference sequences from GenBank (accession no. AY341248 for *CYP2C9* and NG_000008 for *CYP2A13*) (81). We selected 3 pairs of primers for CYP2C9 to amplify desired SNP-containing regions and 7 sets of primers for *CYP2A13* to cover all selected SNPs and verified the uniqueness of primer sequences using NCBI BLAST to prevent nonspecific amplification of other members of the CYP450 family (82) . We obtained all PCR primers from Integrated DNA Technology and performed all PCR reactions on a PTC-200 thermocycler (MJ Research). For *CYP2C9*, we mixed primers (12 pmol F1 and R1; 15 pmol F2 and R2; and 12 pmol F3 and R3) with 500 ng genomic DNA, 2.5 mmol/L dNTPs, 4 units Ex Taq enzyme, and 1× Ex Taq buffer in a 65-µL reaction volume. PCR conditions were as follows: 94°C hot start for 2 min, followed by 40 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 120 s, with final extension at 72°C for 10 min. For *CYP2A13*, we mixed primers (5 pmol F1 and R1; 5 pmol F2 and R2; 7.5 pmol F3 and R3; 5 pmol F4 and R4; 5 pmol F5 and R5; 7.5 pmol F6 and R6; and 7.5 pmol F7 and R7) with 500 ng genomic DNA, 2.5 mmol/L dNTPs, 3 units Ex Taq enzyme, and 1× GC buffer II in a 35-µL reaction volume. PCR conditions were as follows: 94°C hot start for 2 min, followed by 40 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 70 s, with final extension at 72°C for 10 min. PCR products (3 µL) were run on a 1.2% agarose gel with 1× Tris-acetate-EDTA buffer for visualization. Next, PCR products (50 µL) were incubated with 5 units exonuclease I and 5 units shrimp alkaline phosphatase (USB Corp.) at 37°C for 90 min to inactivate

excess PCR primers and dNTPs, respectively, followed by enzyme deactivation at 94°C for 15 min. We performed PCR amplification of individual regions using the same conditions with region-specific forward and reverse primers only.

2.3.4 Single-Base Extension

We obtained all SBE primers from Midland Certified Reagent Co. For *CYP2C9*, we mixed 35 µL corresponding PCR cleanup products with SBE primers (Table 2.1), 210 pmol biotin-11-ddA, 500 pmol biotin-11-ddC, 260 pmol biotin-11-ddG (Perkin Elmer), 250 pmol biotin-16-ddU (Enzo Diagnostics), 10 units Thermo Sequenase enzyme (GE Healthcare), and 1× reaction buffer in 75-µL reaction volume. Cycling conditions for SBE were as follows: 94°C hot start for 2 min, followed by 40 cycles of 94°C for 30 s and 58°C for 30 s. For *CYP2A13*, we mixed 35 µL corresponding PCR cleanup products with SBE primers (Table 2.2), 300 pmol biotin-11-ddA, 550 pmol biotin-11-ddC, 450 pmol biotin-11-ddG, 500 pmol biotin-16-ddU, 12 units Thermo Sequenase enzyme, and 1× reaction buffer in 85-µL reaction volume. Cycling conditions for SBE were as follows: 94°C hot start for 2 min, followed by 10 cycles of 94°C for 30 s and 72°C for 30 s, followed by 35 cycles of 94°C for 30 s and 62°C for 30 s.

2.3.5 Solid-Phase Capture

We separated the extension products from unextended SBE primers by use of streptavidin-coated magnetic beads (Seradyn) as follows. The bead solution (80 µL) was prewashed twice with 90 µL binding-washing buffer (1 mol/L ammonium chloride, 2× Tris-HCl EDTA, pH 7.3) and resuspended in 90 µL binding-washing buffer. We mixed products from the SBE reaction with beads and incubated them for an hour with constant vortexing, then washed the

beads twice with 180 µL binding-washing buffer and once with 180 µL of 0.1 mol/L triethylammonium acetate solution and 180 µL deionized water. The beads were then resuspended in 12 µL formamide and incubated at 94°C for 7 min to denature the streptavidin-biotin complex. We magnetically separated formamide-containing extension products from the beads and mixed them with 120 µL ethanol for overnight incubation. The tube was centrifuged at 14000 rpm for 45 minutes, dried of formamide, and resuspended in 20 µL of 0.1 mol/L triethylammonium acetate solution for desalting using a reverse-phase ZipTip C_{18} column (Millipore).

2.3.6 MALDI-TOF MS Analysis

The desalted DNA was dried and resuspended in 1 µL deionized water and 1 µL matrix and spotted on a 100-well stainless steel sample plate. Matrix consisted of 35.6 mg 3-hydroxypicolinic acid and 6.2 mg ammonium citrate dissolved in 800 µL of 50% acetonitrile solution (Sigma-Aldrich). We performed MALDI-TOF MS analysis in linear positive mode of a Voyager DE Pro instrument (Applied Biosystems) using 25 kV accelerating voltage, 94% grid voltage, 0.07% guide wire voltage, and a delay time of 350 ns taking 75 shots for each spectrum, and accumulating 4 spectra for each sample. We performed external calibration with 5 unextended SBE primers in the 5000 to 11 000Da interval as mass calibrators and used Data Explorer software provided with the instrument for noise removal and smoothing on each accumulated spectrum.

2.3.7 Direct Sequencing

We performed bidirectional sequencing of individual regions with individual PCR products and corresponding forward or reverse PCR primer on an ABI 3730xl capillary gel sequencer using BigDye® Terminator v3.1 mix (Applied Biosystems). Briefly, 100 ng of PCR products were mixed with 4 pmol of primer, 0.6 µl of BigDye® Terminator v3.1 mix, 1 X reaction buffer and deionized water for 20 µl reaction volume in one well of a 96 well plate. Cycling conditions used for sequencing reaction were: 96°C hot start for 1 min, followed by 35 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 90 s. After the reaction, 50 µl of 100 % ethanol and 2 µl of 3 M sodium acetate were added to each well and the plate was centrifuged at 3500 rpm and 4°C for 60 min. The plate was then inverted and liquid from the tubes dispensed by shaking the plate vigorously. 100 µl of 70 % ethanol was added to each well and the plate was centrifuged at 3500 rpm and 4°C for 30 min. The plate was then inverted and liquid from the tubes dispensed by shaking the plate vigorously. 10 µl of Hi-Di formamide was then added to each well to resuspend sequencing products. The plate was heated to 96°C for 15 min and then cooled to 4°C before loading it in the sequencer.

2.4 RESULTS

2.4.1 PCR Amplification

For genotyping SNPs of *CYP2C9* and *CYP2A13*, we amplified SNP-containing regions of each gene in a multiplex PCR reaction. Gel electrophoresis of multiplex PCR products showed desired bands and absence of bands for any nonspecific products. Three bands were detected for *CYP2C9* and 7 bands for *CYP2A13*, with sizes that matched well with those of expected products (Figure 2.1). We confirmed the results by direct sequencing of individual PCR products using capillary gel electrophoresis. The sequences obtained by direct sequencing were found to match well with those of the individual regions from reference sequences of the genes.

2.4.2 SNP Genotype Determination by SPC-SBE

We used PCR products as templates for simultaneous extension of corresponding SBE primers, and isolated and analyzed extension products by MALDI-TOF MS to determine SNP genotypes. Representative mass spectra for analysis of 40 SNPs in *CYP2C9* are shown in Figure 2.2. Genotypes of all 40 SNPs were unambiguously determined by comparing masses of peaks from the spectrum with expected masses (Table 2.1). For example, in the case of Figure 2.2B, a peak is observed at a mass of 7199 Da. When compared with expected masses for *CYP2C9*, this corresponds to the A extension of primer for SNP #16, indicating that the SNP is homozygous for A. Similarly, 2 peaks are observed at 10134 and 10223 Da in the same spectrum. On comparing these masses with expected masses, both are seen to result from extension of primer for SNP #37, indicating that the SNP is a C/T heterozygote. For *CYP2A13*, 50 SNPs were analyzed simultaneously, and genotypes for all SNPs were inferred

by comparing masses of extension products with expected masses (Table 2.2). Representative spectrum are shown in Figure 2.3, with strong spectral peaks observed even in the higher mass range. SNP genotypes for all samples that were analyzed matched those that have been previously reported in the NCBI dbSNP and Human CYP allele databases.

2.4.3 Direct Sequencing for Validation of SNP Genotypes

For validating results obtained by SPC-SBE, we sequenced individual regions using capillary gel electrophoresis for 2 of the *CYP2C9* cell line–based samples, and 4 of the cell line–based samples and 1 anonymous donor sample for *CYP2A13*. A perfect match was seen between genotypes deduced from SPC-SBE and those obtained by sequencing.

2.5 DISCUSSION

We have demonstrated simultaneous genotyping of 40 *CYP2C9* SNPs and 50 *CYP2A13* SNPs using the SPC-SBE approach. The use of molecular affinity between biotin and streptavidin for solid-phase isolation of extension products confers many advantages on SPC-SBE. It provides removal of unextended SBE primers before MALDI-TOF MS analysis, thus eliminating peaks corresponding to primers and associated dimers from resulting mass spectra. As a result, primers do not compete for ion current in the detector, leading to better detection of extension products. This eases spectrum interpretation and increases the number of SNPs that can be genotyped simultaneously. A majority of SBE primers have masses that are very close to or overlapping expected masses of extension products (Tables 2.1 and 2.2). After SPC-SBE, only peaks for extension products are seen in the final spectrum, as shown in Figures 2.2 and 2.3, resulting in accurate genotyping of SNPs. Also, minimum mass difference between heterozygotes increases to 16 Da for A/G heterozygote in SPC-SBE because of the use of biotinylated ddNTPs, compared with 9 Da for A/T heterozygote with regular ddNTPs. This leads to better resolution for heterozygotes and ensures their unambiguous detection, as demonstrated by the spectra in Figures 2.2 and 2.3. Figure 2.2B shows the detection of an A/T heterozygote in the largest extension product with good resolution over 10 000 Da. Figure 2.3C clearly shows an A/G heterozygote for extension of primer 30 with good resolution even at 9500 Da. A mass difference of more than 105 Da between successive primers ensures minimum mass difference of 16 Da between all combinations of extension products. This allows detection of all 4 possible genotypes for each SNP, leading to identification of additional alleles that may not have been reported at a SNP site previously.

SPC-SBE has been used for simultaneous analysis of 30 point mutations in the *p53* gene (73). A mutation is a single-base variation that is usually found to occur at a very low frequency, and hence different samples are expected to have similar spectra with little variation when detecting mutations. A SNP, by definition, must have a less frequent allele that occurs in >1% of the population. Consequently, for the case of SNP genotyping, there is much higher likelihood of obtaining varying genotypes in different samples with high occurrence of heterozygotes, making analysis of SNPs more difficult than analysis of point mutations—for example, Figure 2.2B shows a spectrum for 40-SNP genotyping with 3 heterozygotes; Figure 2.3C shows a spectrum for 50-SNP genotyping with 10 heterozygotes. We were able to deduce unambiguous SNP genotypes for all samples based on spectra obtained from analysis of extension products for both *CYP2C9* and *CYP2A13*. Simultaneous genotyping of 50 SNPs is the highest level of multiplexing reported so far using MALDI-TOF MS.

Previous studies with SNPs in *CYP2A13* have focused on the discovery of SNPs in the gene, and to the best of our knowledge no assays with high levels of multiplexing have been applied for genotyping them. For *CYP2C9*, assays for multiplex genotyping of SNPs have been developed, focusing on genotyping exonic SNPs using hybridization with oligonucleotide array, pyrosequencing, and SNaPshot® approaches (83-85). Pyrosequencing is a novel approach that uses chemiluminescence detection but is currently limited in its capacity for genotyping multiple SNPs. The array-based hybridization approach for genotyping uses fluorescence detection but is better suited to studies that involve genotyping of large numbers of SNPs in a limited number of samples (38, 56). SNaPshot technology uses fluorescence detection and has been used for genotyping 3 SNPs of *CYP2C9* gene (85). The method uses SBE with fluorescently labeled ddNTPs followed by capillary gel

electrophoresis for SNP genotyping and has been used for analysis of up to 35 SNPs simultaneously (86). In principle, the first step for SNP genotyping in the SNaPshot approach, single-base extension for generation of allele-specific products, is the same as SPC-SBE. The main difference between the 2 techniques lies in the way the extension products are detected and used to infer SNP genotype. In SPC-SBE, SNP genotype is deduced by mass spectrometry using direct measurement of the mass-to-charge ratio, which is an intrinsic property for any molecule. Most other techniques like SNaPshot use indirect measures such as fluorescence for detection. Because mass spectrometry relies on an intrinsic property for detection, it has less likelihood of error compared with indirect measures. This is particularly true in the case of heterozygotes that may be relatively harder to detect using fluorescence owing to overlapping peaks in the electropherogram (87). In contrast, a mass spectrum leads to unambiguous detection of a heterozygous SNP owing to the presence of multiple extension peaks.

In SPC-SBE, mass accuracy and mass resolution of spectral peaks are equally important in assigning SNP genotypes. To achieve good mass accuracy, we used 5 mass calibrators in the 5000 to 11 000 Da interval for external calibration of our samples. This resulted in a difference of less than 3 Da between observed and expected masses for all peaks, allowing accurate calling of genotypes. Mass resolution is important in distinguishing closely spaced peaks and is especially important in identifying heterozygotes. With the SPC procedure, we were able to increase minimum mass difference between heterozygotes to 16 Da and enhance sample purity to reduce formation of adducts leading to accurate peak identification. For example, in the case of Figure 2.2B, 3 heterozygotes are observed with the following genotypes: A/G, C/T, and A/T. Of these, the C/T and A/T heterozygotes are easily

distinguishable by mass differences of 89 Da and 66 Da between corresponding extension products, respectively. The A/G heterozygote is comparatively harder to distinguish with its smaller mass difference, but peaks corresponding to both extension products are seen by zooming into the region.

In contrast to problems due to mass accuracy and resolution, errors inherent in the SBE-SPC procedure are impossible to distinguish except by comparative study with a different technique. False positives can arise because of the presence of an additional peak in the spectrum resulting from nonspecific annealing and extension of a primer. False negatives can arise due to insufficient capture of an extension product, resulting in the absence of a corresponding peak from the spectrum, which would lead to improper genotype calling. We analyzed 25 samples in this study with confirmation of genotypes in 7 samples by direct sequencing; however, a much larger sample set would be required to adequately address the rate of false calls with the technique.

Further improvements in multiplexing levels using SPC-SBE are limited by inherent problems with analysis of larger-mass oligonucleotides using MALDI-TOF MS. Larger oligonucleotides are more susceptible to fragmentation during detection and give rise to peaks that are not well resolved, with low signal-to-noise ratio. As a result, it is difficult to use them in genotyping applications that require measurement of small mass differences. Another drawback with the current implementation of SPC-SBE is the long SPC step resulting from the use of streptavidin and formamide for capture and release of biotinylated fragments, which reduces overall throughput of the technique and restricts its use in an automated setting. In an extension of the SPC-SBE approach, a modification has been made

to the SPC step by using monomeric avidin for capture of biotinylated fragments instead of streptavidin (88). This allows release of biotinylated products by a change in pH to substantially reduce time associated with the SPC step. Another problem arises due to uneven crystallization of the 3-hydroxypicolinic acid matrix used for analysis of oligonucleotides. This results in heterogeneity of the sample spot and increases time for analysis owing to the presence of ”hot spots” where the sample is more concentrated.

In the future, improvements in the detection sensitivity of mass spectrometers, coupled with novel techniques for sample preparation and discovery of better matrices for ionization of oligonucleotides, will result in even higher levels of SNP genotyping using SPC-SBE (89, 90). Once large-scale association studies are able to identify SNPs that are important in drug metabolism or as markers for certain diseases, assays for genotyping these SNPs in clinical settings will be needed. The SNPs are likely to be located in different regions of the genome, and efficient approaches for genotyping a limited number of SNPs in large cohorts of individuals will be required. We have demonstrated the applicability of SPC-SBE for use in a candidate gene approach and shown simultaneous genotyping of 50 SNPs that are located over a 9-kb region using a 7-plex PCR reaction. Further innovations, such as construction of a device to allow faster isolation of extension products, will increase throughput of the procedure. Coupled with automated data collection, the technique will become amenable for use in diagnostic settings.

Table 2.1: Extension primers for genotyping 40 SNPs of *CYP2C9*.

#	SNP ID [a]	SBE primer sequence	Primer Mass	Masses of SBE Products (Da) [b,c]			
				B-ddC	B-ddA	B-ddG	B-ddU
1	rs9332211	ccc tct ccc cac ttt t	4679	**5344**	5367	5383	**5433**
2	rs9332118	ccc caa gta gct agc a	4835	**5500**	5523	5539	**5589**
3	rs9332207	cac atg gca aga gag g	4964	5629	**5652**	**5668**	5718
4	rs5031019	cgt ttc tcc ctc atg ac	5072	5737	5760	**5776**	**5826**
5	rs9332208	aga ggt tta agc cac tc	5195	5860	**5883**	**5899**	5949
6	rs1799853	aga gga gca ttg agg ac	5309	**5974**	5997	6013	**6063**
7	rs9332173	gag ata tgc tct cct tct	5441	**6106**	**6129**	6145	6195
8	rs7900194	gtg ttc aag agg aag ccc	5549	6214	**6237**	**6253**	6303
9	rs9332165	gta aac cca ttc tct tcc t	5674	**6339**	6362	**6378**	6428
10	rs9332209	aga ggt tta agc cac tca t	5812	6477	6500	**6516**	**6566**
11	rs12414460	aag aaa tgg aag gag atc c	5919	6584	**6607**	**6623**	6673
12	rs9332212	cac ttt tct tct tgt gat ca	6024	**6689**	6712	6728	**6778**
13	rs1057910	tgc acg agg tcc aga gat ac	6151	**6816**	**6839**	6855	6905
14	rs9332170	caa gtg cct tcc ttt ttc cca	6283	**6948**	6971	6987	**7037**
15	rs9332174	tgt ttc tct tag aga agc ttc	6402	7067	**7090**	**7106**	7156
16	rs1057909	tgg tgc acg agg tcc aga gat	6511	7176	**7199**	**7215**	7265
17	rs9332201	ctt aat gtc tct tta tcg cat t	6641	**7306**	**7329**	7345	7395
18	rs9332166	tca act tta taa tga atc tgg g	6749	7414	**7437**	**7453**	7503
19	rs17110288	gtc tgg att tat ggc agg aag a	6855	7520	**7543**	**7559**	7609
20	rs9332119	taa ctt cgt ttg ctg tta tct ct	6962	**7627**	7650	**7666**	7716
21	rs9332116	tgc ggc ctc tat gtg ggc ttt tg	7069	7734	**7757**	**7773**	7823
22	rs1934968	tca tgg gct taa tgt ctc ttt atc	7300	7965	**7988**	**8004**	8054
23	rs28371686	gca cga ggt cca gag ata cat tga	7411	**8076**	8099	**8115**	8165
24	rs2253635	agg ggg ttt gaa tga gaa aag cta	7530	**8195**	8218	8234	**8284**
25	rs9332117	aga cag agt ctt act ctg tag ctc a	7641	8306	8329	**8345**	**8395**
26	rs9332110	gag agt aaa ggg ctg caa tta gta t	7794	8459	**8482**	**8498**	8548
27	rs9332115	agg tga ttt gta taa tcc tat agc ca	7984	**8649**	8672	8688	**8738**
28	rs1856908	ggt tcg aat gct gga gta ggg aaa ct	8115	8780	8803	**8819**	**8869**
29	rs1934967	ggt att agt ttg tca att tcc caa aac	8249	**8914**	8937	8953	**9003**
30	rs9332169	ccc caa ttt acg ttt tct att aat tta g	8479	9144	**9167**	**9183**	9233
31	rs9332213	tgt gat cac tta gtt gta tct ttg aaa t	8599	9264	9287	**9303**	**9353**
32	rs9332171	aaa gag cct gat gaa tgg aat ttt tag g	8716	**9381**	**9404**	9420	9470
33	rs9332205	tat aaa taa aag agt ttc att ttg gct ca	8914	**9579**	9602	9618	**9668**
34	rs9332120	gga gga tgg aaa aca gag act tac aga gc	9058	**9723**	9746	9762	**9812**
35	rs9332206	act cct agt aag gac ttc agg tag ctt tca	9181	9846	**9869**	**9885**	9935
36	rs9332111	tca aga gaa tga cgc tgt agt tat gaa gac	9303	9968	9991	**10007**	**10057**
37	rs9332199	aca cgt tga ggc ttt cca ttc ctg aat ata a	9469	**10134**	10157	10173	**10223**
38	rs9332198	ggc tag ttt gga aag aac tga atc ttg aca c	9583	10248	**10271**	**10287**	10337
39	rs9332113	agg gag aga agt gtt tga agg cct tgt gtt a	9710	**10375**	10398	**10414**	10464
40	rs12772884	aga caa aag gag ctt gtt atg att att tct ca	9862	10527	**10550**	10566	**10616**

[a] SNP IDs are taken from the dbSNP database

[b] B-ddN (N=A, C, G, U) indicates biotinylated dideoxynucleoside triphosphate.

[c] Extension product masses for previously reported genotypes are bolded

Table 2.2: Extension primers for genotyping 50 SNPs of *CYP2A13*.

#	SNP ID [a]	SBE primer sequence	Primer Mass	Masses of SBE Products (Da) [b,c]			
				B-ddC	B-ddA	B-ddG	B-ddU
1	*3 (1970 A>C)*	ctt ctt cct gag ccg cac	5387	**6052**	**6075**	6091	6141
2	*3, 8 (1706 C>G)*	ggc ggg ctt cct cat cga	5492	**6157**	6180	**6196**	6246
3	*1j (523 C>T)*	gcg ggt cgt ggt gct gtg	5619	**6284**	6307	6323	**6373**
4	*5 (7343 T>A)*	atg gag ctc ttt ctc ttc t	5736	6401	**6424**	6440	**6490**
5	rs3745281	ctg ggc cca ttc aga gtg g	5845	6510	**6533**	**6549**	6599
6	*7 (578 C>T)*	ggc tga gga gtt cag cgg g	5950	**6615**	6638	6654	**6704**
7	rs3815711	gtt cac ctc ccc agg cgt gg	6070	**6735**	6758	6774	**6824**
8	rs3745279	tgt gga cca gag tct tag ga	6197	6862	**6885**	**6901**	6951
9	*6 (7465 C>T)*	tac acc atg agc ttc ctg ccc	6302	**6967**	6990	7006	**7056**
10	rs9916987	cct ccc agg gca ctg aag tgt	6407	7072	**7095**	7111	**7161**
11 [d]	*9 (5294 G>T)*	ccc ttt cca gtc tta ccc tcc a	6517	**7182**	**7205**	7221	7271
12	*1l (7365 T>C)*	tca cca cca tca tgc aga act t	6623	**7288**	7311	7327	**7377**
13	*2a, 2b (74 G > A)*	atg gtc ttg atg tca gtc tgg c	6772	7437	**7460**	**7476**	7526
14	*1k (1991 C>A)*	aca gtc tcc aat gtc atc agc tc	6944	**7609**	**7632**	7648	7698
15	rs3891220	tcc tgc tgc aac aat gcg aat gg	7049	**7714**	7737	**7753**	7803
16 [d]	rs17713907	ttc tca cct tct tta cct cac tga	7165	**7830**	7853	7869	**7919**
17	rs3885816	gtt tcc tgc tgc tca tga agc acc	7280	**7945**	7968	7984	**8034**
18	rs3891221	gtc ata ggt gga gct atg tca acc	7393	8058	**8081**	**8097**	8147
19	rs3745280	gaa gca tcc cag tac atg ata tct c	7610	8275	**8298**	**8314**	8364
20	*1f (6461T>A)*	tca cca atc ccc cca acc tgc ctc at	7716	8381	**8404**	8420	**8470**
21	rs3968432	ggg ccc tct act tca ccc act taa at	7826	8491	**8514**	**8530**	8580
22	rs16974961	tat gaa tgg tct acc tcc gtg tca ta	7936	8601	8624	**8640**	**8690**
23	*1b (- 12 C >A)*	acc acc cca gcc atc acc atc tat cat	8053	**8718**	**8741**	8757	8807
24	rs3968433	taa tca ttt cac atc tac aca aat gtc	8161	**8826**	**8849**	8865	8915
25	rs4061290	att agc cca ctg gaa tat ctg gcc aag	8268	**8933**	**8956**	**8972**	9022
26	*1h,2a,2b,3 (-*	tgc aca gcc aca ctt cca tct cac cct a	8374	9039	**9062**	**9078**	9128
27	*2a,2b (- 1429 A>G)*	caa aga ttc aac cag atc ata cac acc t	8479	9144	**9167**	**9183**	9233
28	*3 (1801 G>C)*	tgg gga tgg gga cta ggt ggg gaa agg	8591	**9256**	9279	**9295**	9345
29 [d]	rs1611827	tag tag aga cgg ggt ttc acc atg ttg g	8700	**9365**	9388	9404	**9454**
30 [d]	*1h,2b,3 (6424C>T)*	ggt tgg ggg gat tgg tga aag tac aca g	8814	9479	**9502**	**9518**	9568
31	*2a,2b (3375 C>T)*	tca tcg cca aga agg tgg agc aca acc ag	8930	**9595**	9618	9634	**9684**
32 [d]	rs3815703	tat tgg cgc ctg cag gtg tgg cgg gag aa	9054	**9719**	9742	**9758**	9808
33	rs1811583	tta aca cct gga caa gtg act gcg tca acc	9160	**9825**	9848	**9864**	9914
34 [d]	*3 (5578 A>C)*	gtt aaa cca ggt tgt gtc aga atc aca ggg	9295	9960	9983	**9999**	**10049**
35	*1h,2b, 3 (7233*	cta gag agt gca gcc ggg ggt cag tag ggg	9418	10083	10106	**10122**	**10172**
36	*1g,2b (7571 G>C)*	ggg agg ggc ggg gct aag aat ggg ggc agt	9523	**10188**	10211	**10227**	10277
37	*1h (2366 C>T)*	cat acc ctc agc tca gct ctc tca cct ggg ca	9641	**10306**	10329	10345	**10395**
38 [d]	rs3826712	gca gga gag tca ggg aga agg ctg ggg aca	9749	10414	10437	**10453**	**10503**
39 [d]	4 (579 G>A)	tga aga gcc agt cga agg tgg cct gct cgc ct	9866	**10531**	10554	10570	**10620**
40 [d]	*1h,2a,2b (-*	gat tcc tgc ttt gtc agg ggc tac tac tat tag	10126	**10791**	10814	10830	**10880**
41	rs3968434	ggc aca tgt tcc cat ccc caa ctt acc gta att t	10272	**10937**	10960	**10976**	11026
42	rs3815713	gct ccc gcc gcc ccc tga cct ctc tcc acc ccc	10383	**11048**	11071	**11087**	11137
43	*2a,2b,3 (2211 T>C)*	cca ccc ccc tcc aga cag tgt ccc ctc aaa	10511	**11176**	11199	11215	**11265**
44	*1e (1894 G>T)*	tcc tgg aat tct gac tct cct cag acc tct gag tt	10624	11289	11312	**11328**	**11378**
45	rs170054	cac tcg cgc tga atc cat ctc tct ccc acc cca	10733	**11398**	11421	11437	**11487**
46	rs11673470	atg cga atg gct gat gtc tgt tct gtt atg aat	10863	11528	11551	**11567**	**11617**
47	*3 (5530 T>C)*	att ggt cag tca ctc ctg tcc caa gcc cac tga	10972	**11637**	11660	11676	**11726**
48	rs3891222	taa gac ccc tag aca cct aaa cac att ccc	11087	**11752**	11775	**11791**	11841
49	rs1645692	cca tcc cca act tac cgt aat ttg taa cag gtg	11204	**11869**	11892	11908	**11958**
50	*1h,3 (7520 C>G)*	tgg tgc agg gct ggt ggg cgg ggc cag gga	11329	**11994**	12017	**12033**	12083

[a] SNP IDs are taken from the dbSNP database and the Human CYP allele database with entries from the latter italicized

[b] B-ddN (N=A, C, G, U) indicates biotinylated dideoxynucleoside triphosphate.

[c] Extension product masses for previously reported genotypes are bolded

[d] Indicates primer annealing to reverse strand

Table 2.3: PCR primers for *CYP2C9*.

#	Primer	Primer Sequence	Size (bp)
1	F1	tat gtg aac acc cct cgt tg	1965
	R1	cca gta agg tca gtg ata tgg ag	
2	F2	tcc tgc act act gca ccc ta	1777
	R2	ttt tct cag gca gat cac tac a	
3	F3	caa gac agg agc cac atg c	1840
	R3	aac ctc ctc tgc aat tct gc	

Table 2.4: PCR primers for *CYP2A13*.

#	Primer	Primer Sequence	Size (bp)
1	F1	gat cta ggc cag tca atg aag	1350
	R1	gca aca gat aaa ctg tgg aac tg	
2	F2	caa ccc tcc tga agt acc aga	1091
	R2	agt gag aga gaa gga gga gac	
3	F3	ttt gca gct ctc ctg ggc ac	1246
	R3	att tgg gtg tca ggg aat aat cta	
4	F4	ctg aaa tac cta aac acc tgg ac	914
	R4	tgt tat aag tgt tcc atg gcg g	
5	F5	cca act gac agc taa gtt gac a	1001
	R5	caa acc cat ggg gag cat g	
6	F6	ctc ttt gtg tca gga gaa tac ac	747
	R6	cca tta tct cct ttt cag gga g	
7	F7	tcc tcc cta gag agt gca g	666
	R7	ata gag aca act ttc tga gcc g	

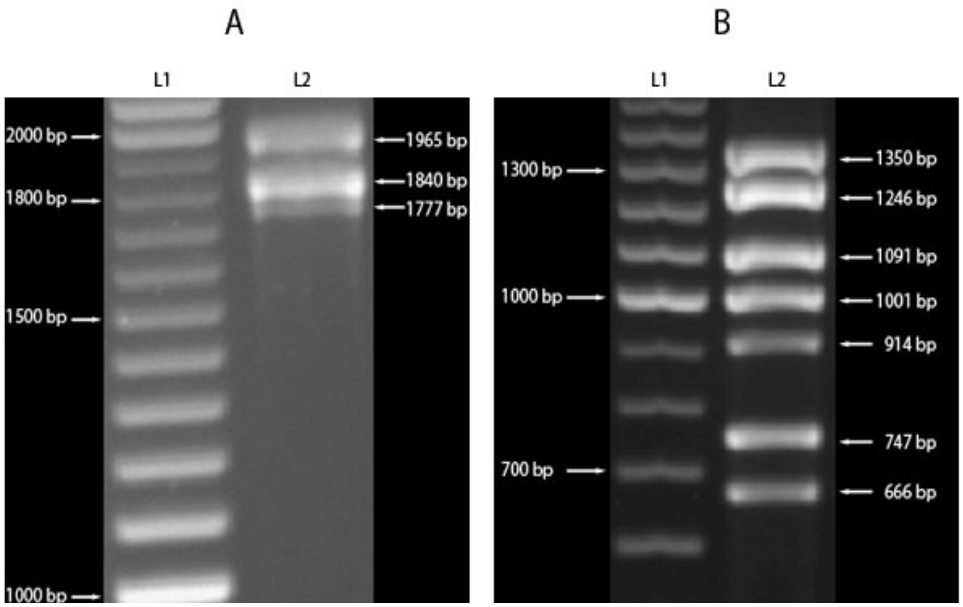

Figure 2.1. Analysis of PCR products of CYP2C9 and CYP2A13 by gel electrophoresis.
(a) Lane 1 is 100 bp ladder and lane 2 shows products obtained in a multiplex PCR reaction for 3 regions of CYP2C9. **(b)** Lane 1 is 100 bp ladder and lane 2 shows products obtained in a multiplex PCR reaction for 7 regions of CYP2A13.

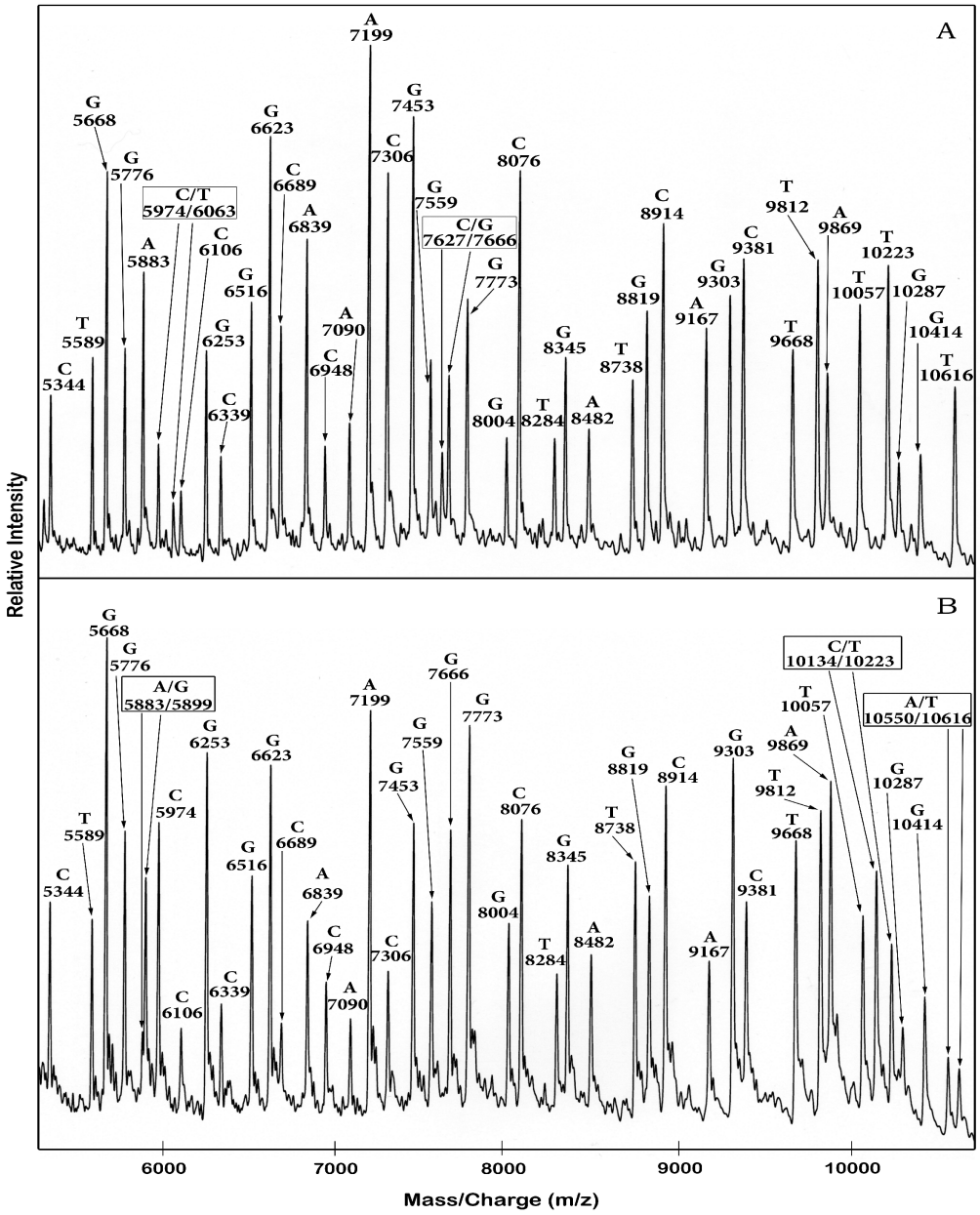

Figure 2.2. Multiplex genotyping of 40 SNPs of *CYP2C9* using SPC-SBE.
Spectra for 2 samples are shown with each peak in the spectrum corresponding to a unique extension product. Mass and corresponding genotype are indicated next to the peak.

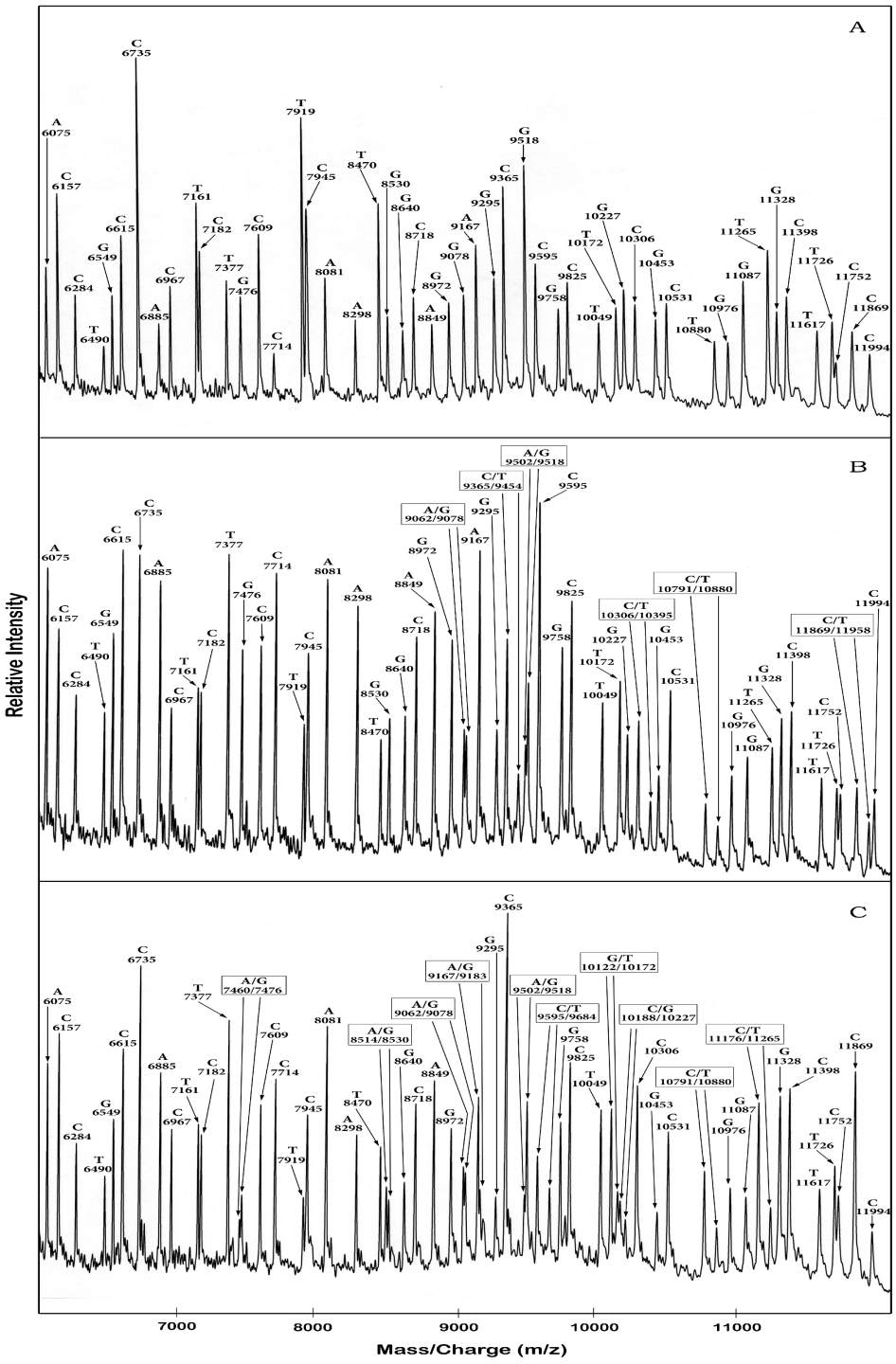

Figure 2.3. Multiplex genotyping of 50 SNPs of *CYP2A13* using SPC-SBE.
Spectra for 3 samples are shown with each peak in the spectrum corresponding to a unique extension product. Mass and corresponding genotype are indicated next to the peak.

CHAPTER 3

FABRICATION OF A MICROBEAD DEVICE FOR IMPROVING THROUGHPUT OF SPC-SBE USING MONOMERIC AVIDIN COATED MICROBEADS

3.1 ABSTRACT

We describe here a prototypical device for isolating biotinylated oligonucleotides for use in MS measurement. It consists of monomeric avidin coated microbeads trapped in a pipette tip and has been used for genotyping single nucleotide polymorphisms (SNPs) with the previously developed solid phase capture-single base extension (SPC-SBE) method. The device reduces processing time for genotyping by SPC-SBE and allows direct spotting of sample for rapid analysis by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). Additionally, it allows simultaneous processing of multiple samples and can be reused after regeneration of beads with no carryover effects. These results indicate that the microbead device is a low cost tool that enhances sample cleanup prior to MS for SNP genotyping.

3.2 INTRODUCTION

We present here the fabrication of a microbead device that facilitates isolation of biotinylated oligonucleotides and utilization of the device for genotyping single nucleotide polymorphisms (SNPs) using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). In the past few years, SNPs have emerged as important markers in pharmacogenomics and disease association studies (24, 91). Various techniques have been developed for genotyping SNPs using different approaches and detection platforms (37, 92). Of these, MS based detection forms an important subset that uses intrinsic property of molecular mass for determining SNP genotypes. This leads to improved accuracy and allows simultaneous detection of molecules of slightly differing masses (50). MALDI-TOF MS is the most commonly used MS detection platform for SNP genotyping due to ease of spectral acquisition and capacity for quantitation (93, 94).

One common requirement for MS-based genotyping is stringent sample purity. Analyte molecules must be free of alkali metal ions and other components of reaction buffers that reduce peak resolution and lead to inaccurate genotype determination (50, 95). To address this issue, the solid phase capture-single base extension (SPC-SBE) method has been developed that uses a set of SBE primers for extension at respective SNP sites with biotinylated dideoxynucleotide terminators (72, 87). Extended primers carry a biotin moiety at their 3’ end and are isolated with streptavidin-coated magnetic beads followed by MALDI-TOF MS analysis to determine SNP genotypes. Following this, SPC-SBE employs harsh conditions to denature streptavidin for the release of bound biotinylated fragments using formamide which cannot be analyzed directly by MALDI-TOF MS. This results in long processing times for analyzing biotinylated fragments before MS analysis. A more recent

report has demonstrated direct release of biotinylated fragments in water while avoiding denaturation of streptavidin (96). Despite this significant improvement, capture and release of biotinylated oligonucleotides for SPC-SBE by streptavidin coated magnetic beads still suffers from drawbacks such as difficulties with processing multiple samples due to the use of magnetic separation. Previously, monomeric avidin coated beads have been used instead of streptavidin coated magnetic beads for isolating biotinylated oligonucleotides to improve the SPC step (97). It was demonstrated that as monomeric avidin binds biotin less strongly than streptavidin, it allows release of bound biotinylated oligonucleotides by a change in pH and consequently improves processing time for SPC-SBE.

Here, we have developed a device consisting of monomeric avidin coated microbeads trapped in a pipette tip and shown that it allows efficient isolation of biotinylated oligonucleotides for genotyping SNPs of cytochrome P450 2A13 (*CYP2A13*) by MALDI-TOF MS. We were able to use the device for parallel analysis of multiple samples in conjunction with a multichannel pipette. Furthermore, the device can be reused a minimum of 5 times with a simple regeneration procedure that eliminates any carryover.

3.3 MATERIALS AND METHODS

3.3.1 Device Construction

Monomeric avidin coated beads of size 40-90 microns (Promega) and nylon mesh of pore size 20 microns (Small Parts) were used for device construction with a 20 µl pipette tip (USA Scientific). A small piece of nylon mesh was inserted from wider end of the pipette tip to act as a filter by completely blocking the dispensing end. Monomeric avidin coated beads (5 µl) were added from the wider end and trapped in the tip by the nylon mesh (Figure 3.1).

3.3.2 PCR Amplification

Two regions of the *CYP2A13* gene were amplified by PCR using primers shown in Table 2.1 (98). All forward and reverse primers were obtained from IDT and PCR reactions performed on a PTC-200 thermocycler. For each region, we mixed 12.5 pmol of forward/reverse primer with 200 ng genomic DNA, 2.5 mM dNTPs, 1 U Ex Taq enzyme and 1X GC buffer II in a 20 µl reaction volume. PCR conditions used were 94°C hot start for 2 minutes, followed by 40 cycles of 94°C for 30 sec, 62°C for 30 sec and 72°C for 70 sec with final extension at 72°C for 10 min. 1.5 µl of PCR products were run on a 1 % agarose gel with 1X TAE buffer for visualization. Then, PCR products were incubated with 1.5 U exonuclease I and 1.5 U shrimp alkaline phosphatase (USB Corporation) at 37°C for 90 min, to inactivate excess PCR primers and dNTPs respectively, followed by enzyme deactivation at 94°C for 15 min.

3.3.3 SBE Reaction

To determine binding capacity of the device, an oligonucleotide (5’ TTG TGG CCG TTT ACG TCG CC 3') with its 5’ end labeled with Cy5 was used in an SBE reaction. 200 pmol of

primer were mixed with 800 pmol biotin-11-ddGTP (Perkin Elmer), 25 pmol of complementary synthetic template, 8U Thermo Sequenase (GE Healthcare) and 1X reaction buffer in 60 µl reaction volume. Cycling conditions used were 94°C hot start for 2 min; followed by 20 cycles of 94°C for 30 sec and 50°C for 30 sec to obtain complete primer extension. To test device for SNP genotyping, four SNPs from region 1 and 5 SNPs from region 2 were selected and corresponding primers of differing molecular weights were used in SBE reaction as shown in Table 3.2 (98). For region 1, we mixed 6 µl of corresponding PCR product with 6 pmol of each primer, 45 pmol of biotin-11-ddATP, 45 pmol of biotin-11-ddGTP, 1.5 U Thermo Sequenase and 1X reaction buffer in 10 µl reaction volume. For region 2, we mixed 7.5 µl of corresponding PCR product with 6 pmol of each primer, 30 pmol of biotin-11-ddATP, 90 pmol of biotin-11-ddCTP, 30 pmol of biotin-11-ddGTP, 1.5 U Thermo Sequenase and 1X reaction buffer in 10 µl reaction volume. Cycling conditions used for both SBE reactions were 94°C hot start for 2 min; followed by 20 cycles of 94°C for 30 sec and 62°C for 30 sec; followed by 30 cycles of 94°C for 30 sec and 58°C for 30 sec. Extent of the reaction was tested by desalting with a ZipTip C_{18} tip using 1 µl of reaction product diluted in 7 µl of 0.1 M TEAA solution followed by MALDI-TOF MS analysis.

3.3.4 Isolation of Biotinylated Extended Primers

Pipette pressure was used to bring solutions in and out of the device for contacting them with beads in the device. Prior to initial use, the irreversible biotin binding sites on the beads were blocked using protocol mentioned in the Device Regeneration section. Then, the beads were prewashed two times with 10 µl binding buffer (100 mM ammonium chloride in 50 mM Tris-HCl, pH ~7.1). To determine binding capacity of device, differing amounts of products from SBE reaction with Cy5 labeled primer were mixed with 5 µl binding buffer and incubated

with the prewashed beads for 45 minutes with constant shaking. For SNP genotyping, 2 µl of products from SBE reaction were mixed with 5 µl binding buffer and incubated with the prewashed beads for 45 minutes with constant shaking. Beads were washed three times with binding buffer, once with 0.1 M TEAA and five times with water and incubated with 5 µl NH_4OH (pH ~12.5) for 10 minutes for elution of captured fragments. For fluorescence measurement, the solution was dispensed in a 96-well microplate, or for SNP genotyping, it was directly spotted on to a 96X2 well teflon coated MALDI sample plate and dried under vertical air flow.

3.3.5 Fluorescence Measurement for Determining Binding and Recovery from Device

We performed fluorescence measurements on a Fluoroskan Ascent FL fluorometer (Thermo Electron) with 646 nm excitation filter and 670 nm emission filter (for Cy5) using 1s integration time. All analyte samples were diluted to 200 µl with 1X TE buffer in a 96-well plate for measurement. Initially, a standard curve of fluorescence versus amount was generated by mixing known amounts of products from SBE reaction with Cy5-labelled primer. The recovery obtained after bead capture was determined based on fluorescence readings and the standard curve.

3.3.6 MALDI-TOF MS Analysis

After drying of sample, 0.6 µl matrix was spotted on the well and dried under air flow. Matrix consisted of 11 mg 3-hydroxy picolinic acid and 2 mg diammonium citrate dissolved in 800 µl of 80 % acetonitrile solution (Sigma-Aldrich). MALDI-TOF MS analysis was done in linear positive mode of a Voyager DE Pro instrument (Applied Biosystems) using 25 kV

accelerating voltage, 94 % grid voltage, 0.07 % guide wire voltage and a delay time of 350 nsec.

3.3.7 Device Regeneration

The beads were equilibrated with 100 mM Tris-HCl (pH~7.1), washed with 5 mM biotin in 100 mM Tris-HCl solution (pH 7.1) three times and incubated for ~20 minutes. Then, beads were washed with 10% acetic acid (pH 2) seven times to strip away all the biotin and washed with 100 mM Tris-HCl (pH ~7.1) seven times and incubated for ~45 minutes before use.

3.4 RESULTS AND DISCUSSION

3.4.1 Device Construction and Operation

We have used monomeric avidin coated beads to construct a prototypic device for efficient isolation of biotinylated oligonucleotides. The device consists of a pipette tip packed with 5 µl microbeads with a nylon mesh at its end (Figure 3.1a). The mesh acts as a filter and liquid is brought in and out of the pipette tip using pipette pressure. Thus the device allows rapid capture of biotinylated oligonucleotides from a solution and subsequent washing to remove any nonspecifically bound molecules. Monomeric avidin allows the release of biotinylated fragments bound to it solely by a change in pH. As a result, the captured fragments can be released and directly spotted on a sample plate and analyzed after drying. The choice of a pipette tip for device construction allows its use in combination with a multichannel pipette for processing multiple samples simultaneously (Figure 3.1b).

The principle for use of the device in SPC-SBE is shown in Figure 3.2. Briefly, products from the SBE reaction with biotinylated terminators are brought in contact with the beads in the device using a pipette and incubated with them to allow binding with monomeric avidin. Biotinylated extended primers bind with the beads and unextended primers and other reaction components are removed by a few washing steps. The biotinylated DNA fragments are then released at high pH and directly spotted on to a sample plate for analysis.

3.4.2 Binding and Recovery of Biotinylated Oligonucleotides from Device Based on Fluorescence Measurement

To determine recovery of biotinylated oligonucleotides captured with the device, we used a Cy5 labeled oligonucleotide primer in an SBE reaction with a complementary synthetic template and biotin-11-ddG. The SBE reaction was carried to completion to ensure all Cy5 labeled molecules carried a biotin moiety after the reaction. This was confirmed by MALDI-TOF MS analysis that showed a peak only for the extended primer (data not shown). A fourfold excess of biotin-11-ddG was used in the reaction to mimic conditions for isolating biotinylated extended primers for SNP genotyping using SPC-SBE.

Figure 3.3 shows plots for total amount of biotinylated oligonucleotide initially bound to beads in device and amount recovered after elution against total amount of biotinylated oligonucleotide that was loaded. For loading of smaller amounts of biotinylated products of up to 20 pmol, good binding and recovery of ~60% or greater was seen. As loading amounts were increased to 40 pmol, good binding was seen (~65%) but recovery declined to ~40%. For loading amounts of 60 pmol, the binding decreased to ~50% and recovery to ~25%. Using a recovery threshold of 60%, the device containing 5 µl microbeads is suitable for loading of up to ~20 pmol biotinylated product in the presence of threefold excess biotin ddNTPs. This would be suitable for analysis of ~20 SNPs since MALDI-TOF MS can routinely detect oligonucleotides in the high femtomolar range. In the SNP genotyping experiments reported here, total amount of biotinylated extended primers brought in contact with the device was lower than 5 pmol.

3.4.3 Use of Device for SPC-SBE Analysis by MALDI-TOF MS

We tested isolation of biotinylated oligonucleotides with the device using a set of 4 SBE primers (1-4 in Table 3.2). The primers were extended with biotin-ddNTPs at corresponding SNP sites to generate biotinylated extended primers. The reaction products were analyzed

after desalting with a ZipTip C_{18} tip to determine extent of reaction. Peaks for all 4 SBE primers and corresponding biotinylated extended primers were seen indicating partial completion of reaction (Figure 3.4a). To test device operation, we incubated products from the SBE reaction with beads in the device, followed by washing of beads to eliminate other reaction components. Then, bound fragments were released from the beads and analyzed by MALDI-TOF MS. The resulting mass spectrum showed peaks for all 4 extended primers with no peaks for any corresponding primers (Figure 3.4b). In addition, due to the washing steps, the salt adducts for the peaks were of small size without need for additional desalting.

3.4.4 Device Reusability

To test reusability, we selected a device that had previously been used for analysis of SBE reaction with primers 1-4. The device was then washed with biotin solution to competitively elute any residual biotinylated oligonucleotides from previous SPC and regenerated at low pH. A set of 5 SBE primers (5-9 in Table 3.2) was used for generating SBE products that were isolated with the device after regeneration. MALDI-TOF MS analysis of the isolated biotinylated molecules showed peaks only for extended primers from the second set with no peak for extended primers from the first set, confirming successful regeneration of the bead surface (Figure 3.4c). We were able to reuse each device a minimum of 5 times with no appreciable loss in mass spectral signal (Data not shown).

3.5 CONCLUSION

In this study, we have fabricated a novel device consisting of monomeric avidin coated beads trapped in a pipette tip. The device was used for the isolation of biotinylated oligonucleotides in SPC-SBE, a SNP genotyping approach that employs MALDI-TOF MS. SPC-SBE generates biotinylated extended primers in an SBE reaction and isolates them from primers and other reaction components prior to MALDI TOF MS analysis. Previously, we have used SPC-SBE for simultaneous genotyping of 50 SNPs from the *CYP2A13* gene (98). A library of 50 primers was simultaneously extended by biotinylated ddNTP terminators followed by isolation of biotinylated extended primers on streptavidin coated magnetic beads and simultaneous analysis of all extended primers after release. The use of streptavidin coated magnetic beads limits applicability of SPC-SBE for high throughput genotyping because it increases time required for isolation of biotinylated fragments. Harsh conditions using formamide and heat were employed in the original SPC-SBE approach to denature streptavidin for release of biotinylated fragments followed by ethanol precipitation before MALDI-TOF MS analysis (99, 100). More recently, a much simpler protocol using heating of streptavidin beads suspended in water to 70°C has been shown to break the streptavidin-biotin complex without denaturing streptavidin allowing reuse of beads (96). Even with this significant improvement, the use of streptavidin coated magnetic beads is not well suited for simultaneous processing of multiple samples due to magnetic separation.

The device, as introduced in this study, addresses these drawbacks and employs a simple protocol for isolating biotinylated oligonucleotides. The monomeric avidin coated beads in the device capture biotinylated fragments that are washed and directly released at high pH for

MALDI-TOF MS analysis. Thus, it reduces time for isolation of biotinylated oligonucleotides to less than 2 hours.

Furthermore, a major problem with analysis of oligonucleotides using MALDI-TOF MS is the presence of salt adducts which confound spectrum interpretation. As a result, most MALDI-TOF MS-based assays that use oligonucleotides for SNP genotyping use an additional purification step to get rid of salts and other impurities (95). The bead device allows efficient desalting prior to MALDI TOF MS analysis due to washing steps in the SPC procedure.

Other important issues in SNP genotyping include the requirement for analysis of multiple samples simultaneously and the cost per genotype (92). The bead device allows parallel processing of multiple samples in combination with a multichannel pipette. This arrangement can be further extended to an automated format with beads entrapped in microfluidic devices. Additionally, we have been able to reuse the device a minimum of 5 times with a simple regeneration protocol which reduces the cost of SNP genotyping.

In summary, the ease of capture and recovery of biotinylated oligonucleotides, the aptness for automation and reusability demonstrate the feasibility of the device for use in SNP genotyping by MALDI-TOF MS.

Table 3.1. PCR primers and expected sizes of products.

Region	Forward Primer	Reverse Primer	Size (bp)
1	caa ccc tcc tga agt acc aga	agt gag aga gaa gga gga gac	1091
2	caa acc cat ggg gag cat g	cca act gac agc taa gtt gac a	1001

Table 3.2. SBE primers for SNP genotyping. Expected product masses are shown in bold.

#	SBE Primer Sequence	Region	Primer	Masses of SBE Products (Da)			
				B-ddC	B-ddA	B-ddG	B-ddT
1	ctg ggc cca ttc aga gtg g	1	5845	6510	**6533**	**6549**	6599
2	tgt gga cca gag tct tag ga	1	6197	6862	**6885**	**6901**	6951
3	atg gtc ttg atg tca gtc tgg c	1	6772	7437	**7460**	**7476**	7526
4	gaa gca tcc cag tac atg ata tct c	1	7610	8275	**8298**	**8314**	8364
5	ccc ttt cca gtc tta ccc tcc a	2	6517	**7182**	**7205**	7221	7271
6	tcc tgc tgc aac aat gcg aat gg	2	7049	**7714**	7737	**7753**	7803
7	gtt tcc tgc tgc tca tga agc acc	2	7280	**7945**	7968	7984	**8034**
8	gtc ata ggt gga gct atg tca acc	2	7393	8058	**8081**	**8097**	8147
9	tat gaa tgg tct acc tcc gtg tca ta	2	7936	8601	8624	**8640**	**8690**

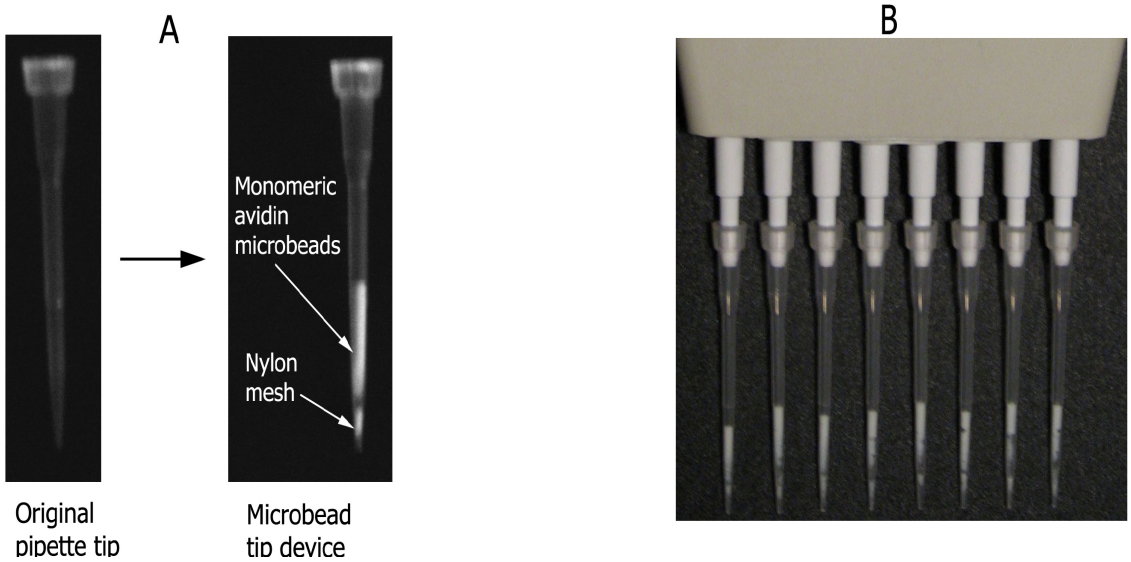

Figure 3.1. Microbead pipette tip device.
(a) Device consists of monomeric avidin coated microbeads trapped in a pipette tip with nylon mesh. **(b)** Use of device in combination with a multichannel pipette.

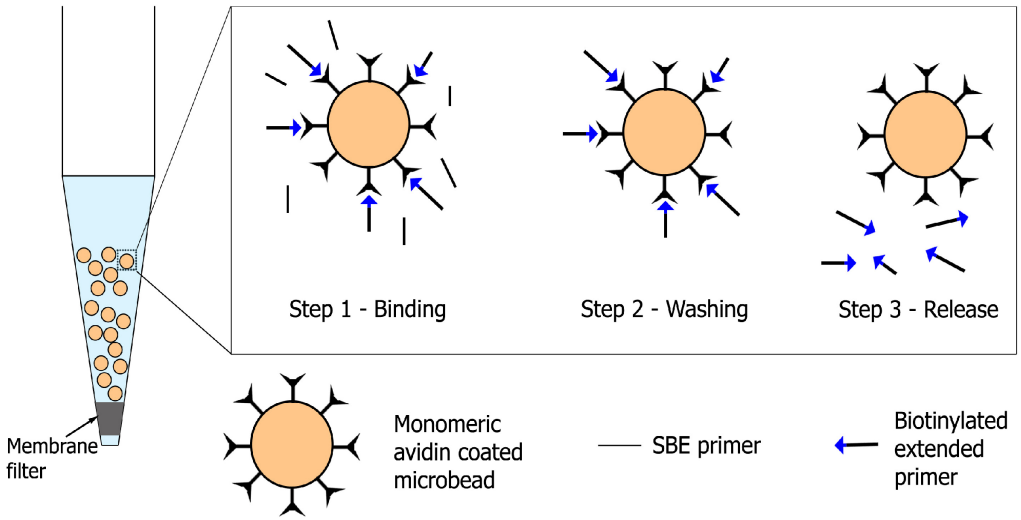

Figure 3.2. Schematic showing principle of microbead device operation.
Products from SBE reaction with biotinylated ddNTPs are incubated with beads to allow binding of biotinylated extended primers. The beads are washed to remove primers and other reaction components followed by release of bound biotinylated fragments by a change in pH.

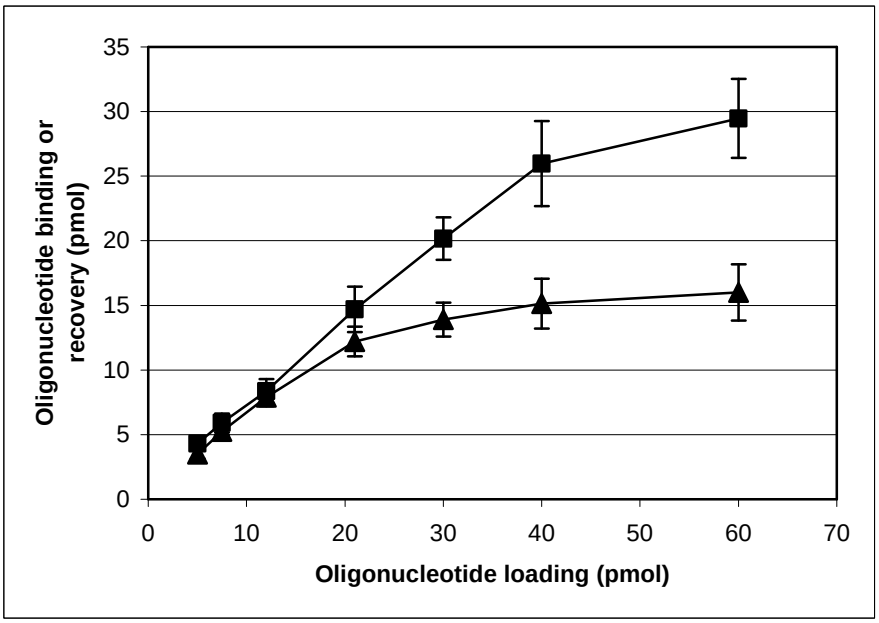

Figure 3.3. Binding capacity and recovery from device.
Plots of total amount of biotinylated oligonucleotide initially bound to beads in device (filled squares) and amount recovered after elution (filled triangles) against total amount of biotinylated oligonucleotide that was loaded. Error bars indicate standard deviation for an average of 3 measurements.

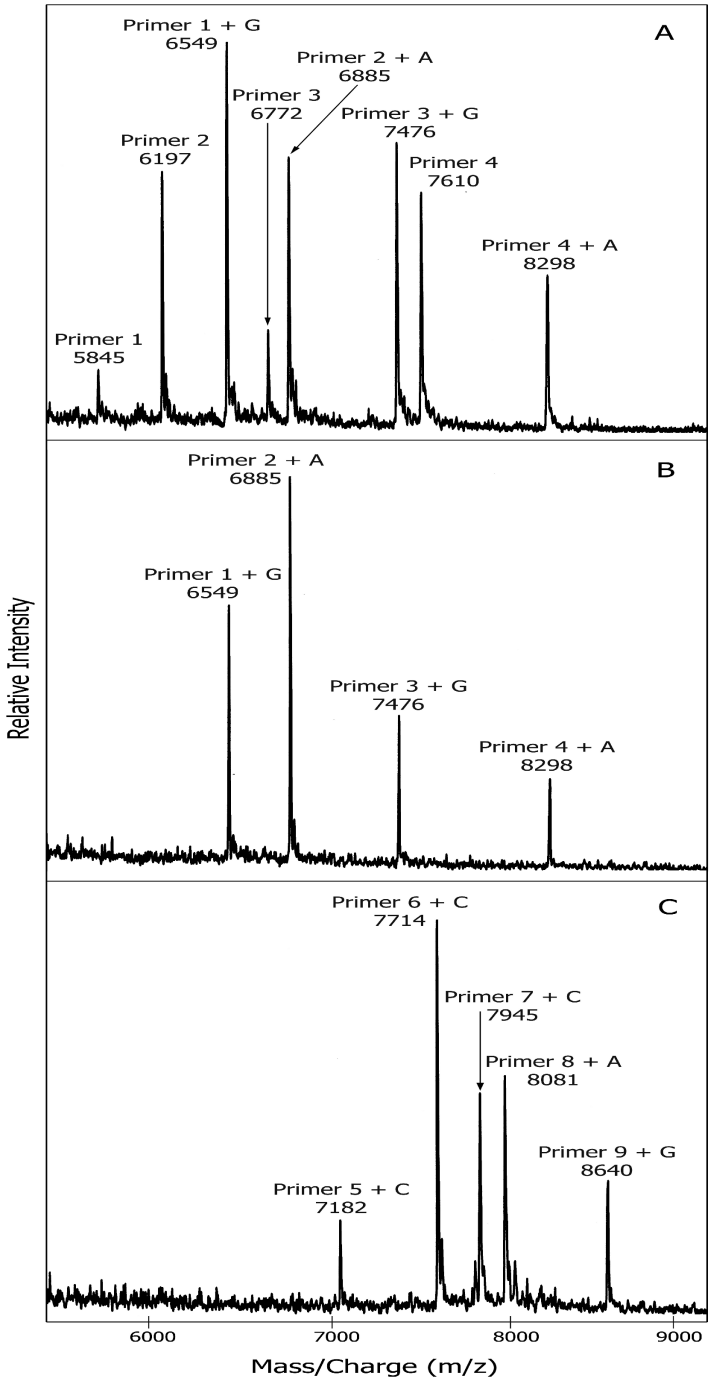

Figure 3.4. Analysis of biotinylated oligonucleotides generated after SPC-SBE using MALDI-TOF MS for SNP genotyping.

a) Analysis of SBE reaction with primers 1-4 after desalting with a ZipTip C18 tip.

b) SPC for SBE reaction with primers 1-4 using microbead device.

c) SPC for SBE reaction with primers 5-9 with previously used device after device regeneration.

CHAPTER 4

AN APPROACH FOR THE RELEASE OF BIOTINYLATED FRAGMENTS FROM STREPTAVIDIN FOR DIRECT MS ANALYSIS

4.1 ABSTRACT

We describe here the use of a previously developed approach for the release of biotinylated molecules bound to streptavidin using water for rapid analysis by MALDI-TOF mass spectrometry (MS). We have applied this approach for rapid genotyping of single nucleotide polymorphisms (SNPs) based on the solid phase capture-single base extension (SPC-SBE) approach that involves isolation of biotinylated oligonucleotides prior to MS analysis. We have demonstrated direct MS analysis of biotinylated oligonucleotides isolated using streptavidin-coated magnetic beads. This leads to substantial reduction in processing time for SNP genotyping with the SPC procedure. We have been able to reuse the beads 5 times without appreciable change in mass spectral signal and also used this protocol for highly multiplexed SNP genotyping using a previously developed assay.

4.2 INTRODUCTION

MALDI-TOF mass spectrometry (MS) is a commonly employed technique for the detection of biomolecules that uses direct measurement of intrinsic property of molecular mass allowing label free detection (101-103). However, it requires stringent sample purity involving removal of salts and other contaminants that interfere with MS analysis (95, 104). Consequently, analytes, especially those from complex mixtures, require purification and desalting prior to MS analysis and affinity separation is a common approach to this end (105-107). Affinity separation based on the interaction between biotin and streptavidin has been widely used for purification of different biomolecules including proteins, oligosaccharides and oligonucleotides prior to MS analysis due to its high specificity and affinity (108-110).

A MALDI-TOF MS based approach, solid phase capture-single base extension (SPC-SBE), uses the affinity between biotin and streptavidin to improve the analysis of single nucleotide polymorphisms (SNPs) (72, 98). SNPs serve as important genetic markers, and SNP genotyping is one of the most common applications of oligonucleotide analysis by MALDI-TOF MS (20, 92, 111, 112). In SPC-SBE, the SPC step facilitates the isolation and desalting of biotinylated oligonucleotides and improves multiplexing levels for SNP genotyping. However, the use of formamide for release of bound biotinylated fragments necessitates additional lengthy steps for its removal before MS analysis. Consequently, processing time for the elution protocol increases substantially thus offsetting the improvement in multiplexing capacity due to the SPC step.

Recently, we have demonstrated utility of a microbead device containing monomeric avidin coated microbeads for rapid isolation of biotinylated oligonucleotides and its use for MS based SNP genotyping with SPC-SBE (113). Monomeric avidin allows release of bound biotinylated fragments by a change in pH due to ~10^8 fold lower affinity for biotin compared to streptavidin (97). The microbead device works well for capture and release of small quantities of biotinylated fragments. However, a decrease in recovery for larger amounts of bound fragments limits its applicability for highly multiplexed SNP genotyping. Additionally, the lower binding affinity for biotin can increase the probability for non-specific binding compared to streptavidin.

4.3 MATERIALS AND METHODS

4.3.1 PCR Amplification

PCR amplification for *CYP2A13* was performed using appropriate set of primers as previously described (98). After the reaction, PCR products were isolated from unextended PCR primers and unused dNTPs using a Qiagen PCR product purification kit following manufacturer's protocol.

4.3.2 SBE Reaction

To characterize elution from beads, a Cy5 labeled primer was extended with biotin-11-ddG in an SBE reaction with 100 % yield as previously described (113). For SBE with *CYP2A13*, we used a set of previously designed primers as previously described (98). For region 2, 3 pmol of primers 3, 5, 6, 8, 19, 23; 6 pmol of primer 13; 9 pmol of primer 39 and 10.5 pmol of primer 38 were mixed with 35 pmol of biotin-11-ddA, 90 pmol of biotin-11-ddC, 70 pmol biotin-11-ddG (Perkin Elmer, Waltham, MA), 2.5 U Thermo Sequenase enzyme (GE Healthcare, Piscataway, NJ), 6 µl of region 2 PCR purified product and 1X buffer in a 20µl reaction volume. For region 5, 3 pmol of primers 11, 15, 17, 18; 7.5 pmol of primers 16, 23; 9 pmol of primers 48 and 12 pmol of primer 34, 24; 24 pmol of primers 46, 47 were mixed with 20 pmol of biotin-11-ddA, 90 pmol of biotin-11-ddC, 20 pmol biotin-11-ddG, 90 pmol biotin-16-ddU (Enzo Diagnostics, Farmingdale, NY), 2.5 U Thermo Sequenase enzyme, 8 µl of region 5 PCR purified product and 1X buffer in a 20µl reaction volume. For 7plex SBE, 2 pmol of primers 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 14, 15, 17, 18, 19, 23, 27; 3 pmol of primers 12, 20, 21, 25, 26; 4 pmol of primers 13, 24, 33; 5 pmol of primers 16, 23, 37; 7 pmol of primers 29, 30, 25, 38; 8 pmol of primers 28, 32, 34; 10 pmol of primers 36, 50; 12 pmol of primers 42, 45, 49; 16 pmol of primers 40, 41, 43, 44, 46, 47 were mixed with 230 pmol of biotin-11-

ddA, 410 pmol of biotin-11-ddC, 300 pmol biotin-11-ddG, 380 pmol biotin-16-ddU, 10 U Thermo Sequenase enzyme, 15 µl of 7plex PCR purified product and 1X buffer in a 30µl reaction volume. The same set of cycling conditions were used for all SBE reactions for *CYP2A13* primers: 94°C hot start for 2 min; followed by 45 cycles of 94°C for 30 s and 62°C for 30 s.

4.3.3 SPC and Release of Biotinylated Oligonucleotides

We used streptavidin coated magnetic beads (Seradyn, Indianapolis, IN) for capturing biotinylated extension products – 10 µl beads for testing release from beads, 15 µl beads for genotyping SNPs from regions 2 and 5 and 70 µl beads for 50plex SNP genotyping. All prewashing and washing steps were done with volumes 1.5 times the original bead volume. Initially, beads were prewashed twice with binding buffer (200 mM ammonium chloride in 20 mM Tris-HCl EDTA, pH ~7.2) and then resuspended in same amount of binding buffer. Then products from SBE reaction were mixed with beads and contacted with them by vortexing the solution for 45 minutes. Beads were then washed twice with 20 mM TE buffer and once each with 0.1 M TEAA solution (Sigma-Aldrich, St. Louis, MO) and deionized water before resuspending them in 7 µl deionized water. The solution was then heated at 0.5°C/s to a set temperature and incubated at that temperature for a certain time before cooling to 35°C and magnetic separation of solution from beads. For testing release from beads, incubation temperatures of 75°C and 80°C were used for differing times. For all SNP genotyping experiments, an incubation time of 75 s at 80°C was used. After product separation, beads were washed once with 20 mM TE buffer and then resuspended in it. Beads were regenerated prior to reuse using only the elution procedure without any product loading.

4.3.4 Fluorescence Measurement to Determine Release from Beads

Fluorescence measurements were done on a Fluoroskan Ascent FL fluorometer (Thermo Electron, Waltham, MA) as previously described (113). Briefly, all analyte samples were diluted to 200 µl with 1x TE buffer in a 96 well plate using a 646-nm excitation and 670-nm emission filter using an integration time of 1s for measurement. Initially, a standard curve of fluorescence versus amount of product was generated by mixing known amounts of products from Cy5 SBE reaction. The recovery obtained after bead capture was determined based on fluorescence readings from analyte samples and the standard curve.

4.3.5 MALDI-TOF MS Analysis

The solution containing eluted products was spotted on a 96x2 teflon coated plate and dried under vacuum followed by the addition of 0.5 µl matrix for MALDI-TOF MS analysis. Matrix consisted of 16 mg of 3-hydroxy picolinic acid and 2.8 mg of diammonium citrate dissolved in 800 µl of 70 % acetonitrile solution (Sigma-Aldrich, St. Louis, MO). Analysis was done in linear positive mode of a Voyager-DE Pro instrument (Applied Biosystems, Foster City, CA).

4.4 RESULTS AND DISCUSSION

4.4.1 Recovery of Biotinylated Fragments from Beads

We used a Cy5 labeled oligonucleotide primer to characterize recovery of biotinylated oligonucleotides from streptavidin coated beads. The primer was extended with biotin-11-ddG in an SBE reaction with complementary template so all Cy5 labeled molecules carried a biotin moiety at their 3' end. Reaction products were captured on streptavidin coated beads followed by multiple washing steps. Holmberg et al reported complete release of biotinylated oligonucleotides bound to streptavidin coated beads without denaturing streptavidin by heating the desalted beads in water at a rate of 0.5°C/s to 70°C or above with incubation at the final temperature for more than 1s (96). We tested release of biotinylated Cy5 labeled oligonucleotides by heating the bead solution, after washing, at 0.4°C/s to 75°C with incubation at 75°C for 5 s and obtained a recovery of ~50% (Figure 4.1). An increase in incubation time to 150 s at 75°C led to recovery of ~70% while incubation at 80°C for 75 s yielded recovery of ~80 % with a small increase in recovery seen by increasing incubation time at 80°C to150 s.

We were not able to obtain complete recovery of biotinylated oligonucleotides from the beads as seen in the report by Holmberg et al even after substantial increase in incubation time compared with their experiments. This could be related to one or a combination of things that were different in our experimental protocol – manual handling for all steps, much higher loading amounts of biotinylated molecules and use of beads smaller than 1µm. We incubated beads at 80°C for 75 s for the release of biotinylated fragments for all SNP genotyping experiments.

4.4.2 Use of Modified Elution Protocol for SNP Genotyping

We performed SBE reaction with primers for SNPs from region 2 of *CYP2A13* for initial testing of modified elution step for genotyping SNPs. After SBE, reaction products were incubated with streptavidin-coated beads followed by washing and release of captured products in water. Products were analyzed by MALDI-TOF MS and showed strong peaks for extended primers for all SNPs (Figure 4.2a). To test bead reusability, we regenerated the beads by use of elution protocol to strip the beads of residual bound products. Following this, SPC was done with products from SBE reaction with region 5 primers. After SPC step and elution, MS analysis showed peaks corresponding to extended primers for all 10 SNPs of region 5 with no peaks for extended primers from region 2 indicating successful reuse of beads (Figure 4.2b). We were able to reuse beads 5 times for SPC without appreciable loss in mass spectral signal with the modified elution protocol using region 2 and region 5 SBE products alternately.

Following this, we performed analysis of 50 SNPs of *CYP2A13* with the modified elution protocol using a previously genotyped sample (98). This highly multiplexed assay requires isolation and detection of many biotinylated oligonucleotides in the 6-12 kDa range. In general, longer oligonucleotides yield much lower peak intensity in a mass spectrum and hence require larger amounts for accurate detection. So far the 50-plex assay has only been performed using streptavidin-coated beads which provides efficient capture due to high binding affinity, followed by heating in formamide to denature streptavidin and ensure good recovery (98). In this study, we were able to determine genotypes for all 50 SNPs accurately as confirmed by direct sequencing using the modified elution protocol (Figure 4.3). This

suggests efficient recovery of bound biotinylated fragments with the water-based elution protocol that allows highly multiplexed SNP genotyping.

4.5 CONCLUSION

Here, we have demonstrated that a recently developed elution method for the release of biotinylated oligonucleotides from a streptavidin-coated surface can be effectively employed for MALDI-TOF MS analysis of SNPs. These results indicate that the modified elution protocol allows purification of biotinylated molecules without introducing organic solvents which must be removed prior to MS measurement thus reducing processing time substantially. Furthermore, we were able to perform 50-plex SNP genotyping using the new approach with accurate genotype determination of all SNPs. The combination of reduction in processing time and high multiplexing capacity, which result from the modified elution method, provides a robust means for higher throughput SNP genotyping. The approach can also be extended to the analyses of other biomolecules such as proteins and oligosaccharides. Furthermore, this elution approach can be implemented in an automated format using streptavidin-coated surfaces to allow rapid binding and release of biotinylated fragments for MS analysis (114, 115).

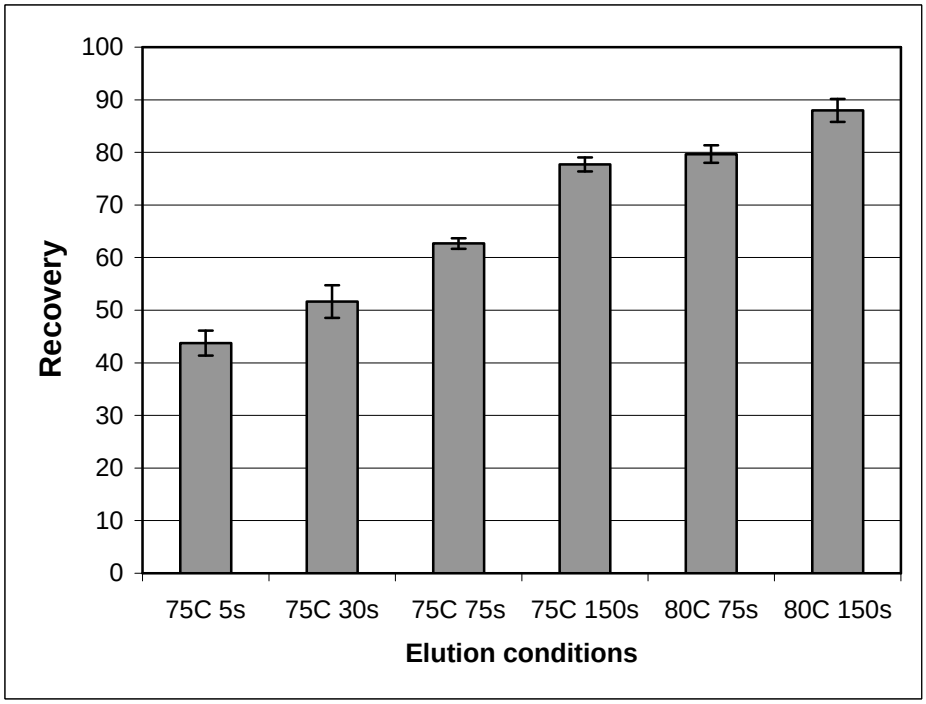

Figure 4.1: Percentage recovery of biotinylated Cy5 labeled oligonucleotide from streptavidin coated beads at different elution conditions.

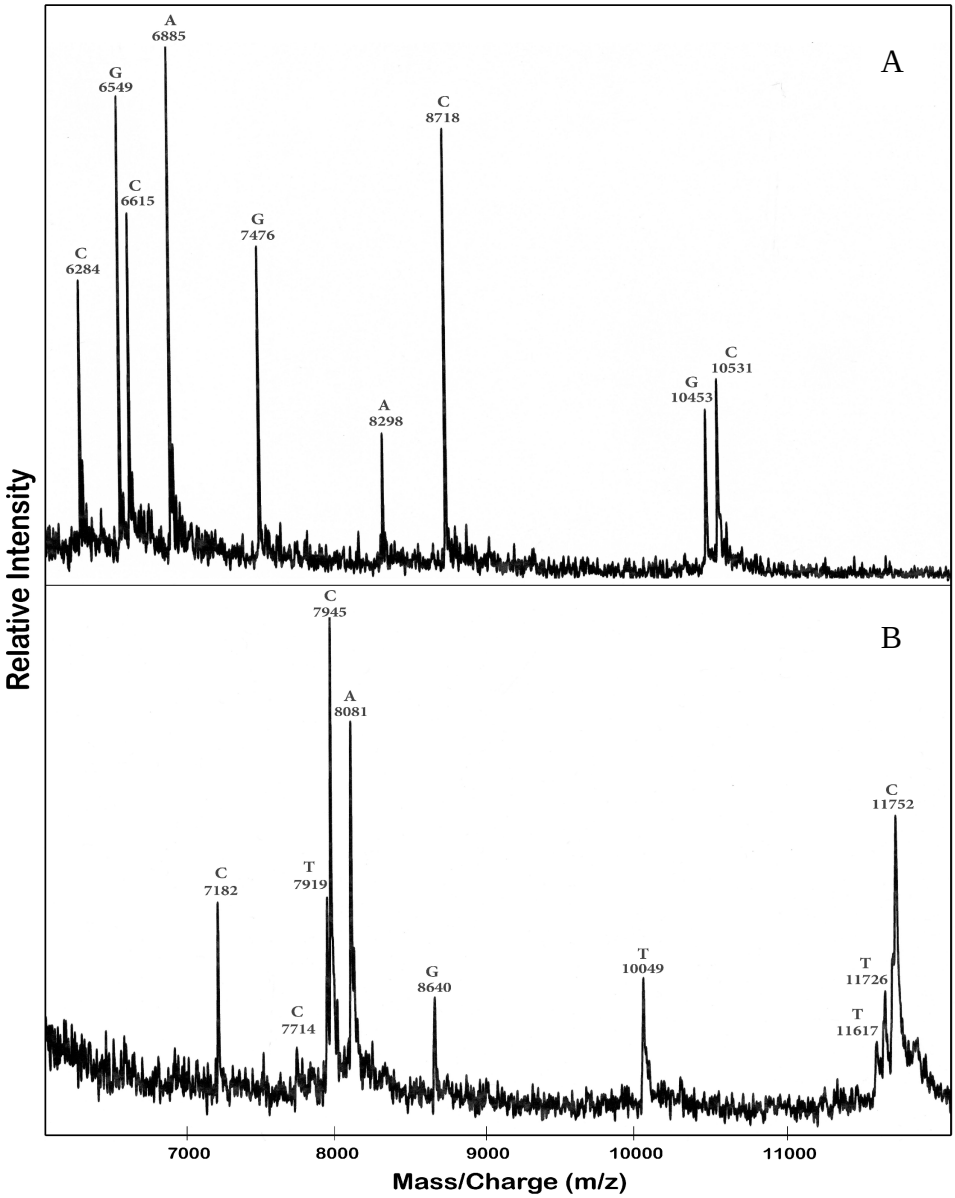

Figure 4.2: Use of modified elution procedure for SNP genotyping by SPC-SBE. Top figure shows analysis of SBE products from region 1 of CYP2A13. Bottom figure shows reuse of beads after regeneration for analyzing SBE products from region 4 of CYP2A13.

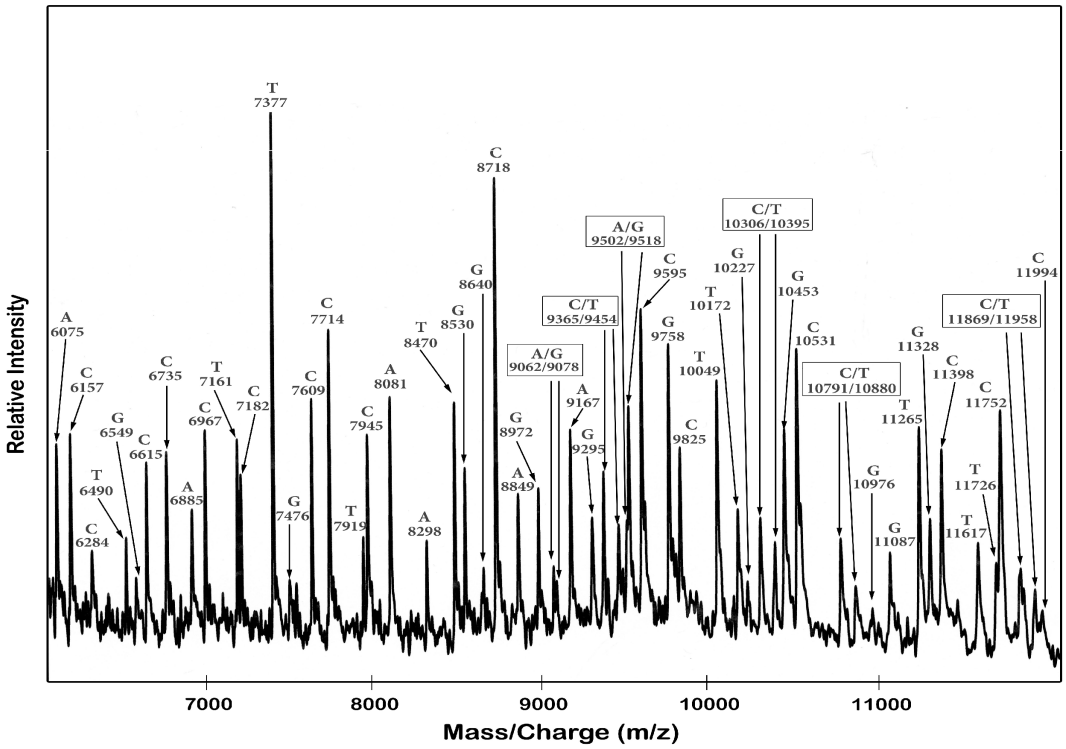

Figure 4.3: SNP genotyping of 50 SNPs from the CYP2A13 gene with SPC-SBE by use of modified elution procedure. Initially, products from SBE reaction with 50 primers were captured on streptavidin coated magnetic beads followed by washing of beads to eliminate undesired reaction components and salts. Bound fragments were released by heating beads in water and direct use of solution for mass spectrometry analysis to determine SNP genotypes.

CHAPTER 5

RESEARCH SUMMARY AND FUTURE DIRECTIONS

Single nucleotide polymorphisms (SNPs) are one of the most important markers for studying the genetic bases of interindividual variability and can be studied using a number of techniques. The focus of this work was to improve the multiplexing levels and throughput of SNP genotyping by mass spectrometry (MS) by using the molecular affinity between biotin and (strept)avidin for purification prior to MS analysis.

We explored the previously developed solid phase capture-single base extension (SPC-SBE) for improving multiplexing levels of SNP genotyping with MS detection. The approach involves the extension of a set of primers at SNP sites using biotinylated dideoxynucleotide terminators. The extended primers carry a biotin moiety at their 3'end and are isolated using streptavidin coated beads with simultaneous desalting using washing steps prior to MALDI-TOF MS analysis. We analyzed up to 50 SNPs with this approach using two genes from the cytochrome P450 family of genes as model system. This is the highest levels of multiplex SNP genotyping with MS detection which results from the use of the isolation step before MS analysis. However, this approach uses harsh conditions employing formamide to denature streptavidin for releasing bound biotinylated fragments resulting in long processing times before MALDI-TOF MS analysis. We used two approaches to address this issue by reducing processing time after isolation to improve throughput of the technique. First, we fabricated a microbead device using monomeric avidin coated beads for rapid processing of multiple samples. Monomeric avidin binds biotin with much lower affinity compared to

streptavidin and allows reversible dissociation of the interaction by a change in pH. The use of the monomeric avidin coated beads trapped in a pipette tip allowed rapid capture and release of biotinylated oligonucleotides and reduced time for isolation of fragments to less than 2 hours. Additionally, the reversible release of biotinylated fragments from monomeric avidin allowed reuse of the device 5 times thus reducing the cost of SNP genotyping. However, the microbead device allows simultaneous genotyping of a small number of SNPs thus limiting its applicability for multiplex SNP analysis. The second approach was based on a recent report that demonstrated direct release of biotinylated oligonucleotides from streptavidin using mild heat in water in the absence of any salts. We used this protocol for direct release of biotinylated oligonucleotides in water followed by rapid analysis of released products by MS analysis for highly multiplexed SNP genotyping. In concert, these studies attest to the feasibility of use of the affinity purification system for high throughput genotyping with MS detection.

However, further improvements are required in the implementation of the technique to eliminate variability and improve its throughput. Additionally, the technique needs to be validated by analyzing different sets of SNPs across multiple samples to determine its error rate.

Currently, all the steps of the technique are manual which inevitably leads to variability for the analysis of multiple samples. This is unfavorable for accurate SNP genotype determination and reduces the throughput of technique. The reproducibility and throughput of the process would improve substantially by automating the various steps of the process (34). The procedure for SNP genotyping using the technique can be broken down into four steps:

PCR and SBE reactions, SPC isolation, sample and matrix spotting and MS analysis. Automation of the steps could be done as follows: use of an automated handler for dispensing and mixing reactants for the reaction steps; use of approaches described in this thesis in a microfluidic device format for the SPC step; use of an automated spotter in temperature and humidity controlled conditions for spotting; and use of programs from the MS software that allow automated data collection and interpretation. Of these, the sample-spotting step is most critical for obtaining good mass spectral signal and influences reproducibility to the greatest extent especially for highly multiplexed genotyping. Hence, automation of sample spotting would be an important improvement, and in combination with the shorter SPC and release step allow further validation of the technique. The other steps have lower variability and their automation would help more with improving throughput.

Following improvements in the processing steps, the error rate of the technique needs to be estimated. We have used the technique for highly multiplexed SNP genotyping with 14 samples for CYP2A13 and 11 samples for CYP2C9 and validated the technique by confirmation of genotypes in a total of 7 samples by direct sequencing. The number of samples in the study is not large enough to address the rates of false positive and false negative calls for the technique. Additionally, it has been used for studying 2 genes only. The analysis of different sets of SNPs analyzed for a larger number of samples would be required to adequately address the accuracy of the technique by comparing them with results obtained from a different technique. A better approach would be to use SNPs in DNA samples that have been previously validated by a number of different approaches and pedigree analysis (eg. CEPH family collection) (50).

The work in this thesis has focused on application of the molecular affinity based MS technique for SNP genotyping considering all 4 possibilities at a SNP site. The levels of multiplexing levels can be further improved further by only considering biallelic possibilities at SNPs. Previously, MALDI-TOF MS based detection has been used for DNA and RNA quantitation with high accuracy (116). The SPC-SBE technique would be ideally suited for use in such quantitation since it performs desalting as part of the SPC step. As a result, it can be used for other applications such as allele frequency determination in pooled DNA samples and RNA quantitation using a real competitive PCR approach (55). Additionally, the SPC-SBE technique can also be used to study DNA methylation patterns using the bisulfite conversion method (50). The improvements in implementation of the technique and its validation and use in other applications will make it a good candidate for use in research and diagnostic settings.

APPENDIX A: PROGRAM FOR THE SELECTION OF SBE PRIMERS

The selection of primers used in the study was done using a program written in Perl (listed below) to provide sufficient mass difference between successive primers for simultaneous detection during mass spectrometry analysis.

```
# *** PROGRAM FOR SELECTION OF SBE PRIMERS FOR SNP GENOTYPING***

#Program requires a file containing fasta sequences of SNP containing regions with SNP
sites marked in capital letters and surrounding sequences shown in small letters (eg.
gagCgatcgacgtgacgatTgatgcgatgc - Here C & T indicate the site corresponding to SNPs and
the surrounding region is shown in small letters.

#Input file containing sequence and SNP sites being read from file 2a13.txt

$input = "/empyrean/home2/u26/misra/perl/sbe/primdesign/2a13.txt";

open IN, $input or die "Cannot open $stuff for read :$!";

while (<IN>) {
   @temp=split //,  ; # initializes array @temp
   push (@init,@temp); # all elements gathered in array @init
}


# Calculates number of SNPs and generates array of exonic and intronic SNPs separately.
Also generates an array containing last 6 letters of possible primers.

$count = 100;
foreach (@init) {
   $no=$no+1;
   if (/[agct]/i) {
         $seqchar=join '', $seqchar,$_; # string with entire sequence
         push (@seq, $_);
         $x = $x + 1;
         if (/[AGCT]/) {
            $count = $count + 1;
            push (@snpos, $x-1); # Array @snpos with SNP posn in @seq
            push (@count, $count); # Array keeping count on SNP #
            $curr = join '', @seq[$x-7..$x-2]; # primer 3' ends
            push (@primend, $curr); # @primend contains 3' for all primer
            if ($init[$no] eq E) {
                  push (@snpex, $count);
```

```
                $snpex{$count}=$_;    # Array @snp containing only Ex SNPs
            } else {
                push (@snpin, $_);    # Array @snp containing only SNPs
            }
        }
    }
}


# Only SNPs that have last 6 letters of 3'end repeated less than 5 times throughout the entire
PCR sequence considered for primer design.

for ($x=0; $x <= $#snpos; $x++) {  # for loop over all SNPs to generate
    # list of all possible primers
    $primend=$primend[$x];
    $rep =0;
    while ($seqchar =~/$primend/gi) {  # finds # times 3' end of primers is
        $rep=$rep+1;                # repeated in the seq
    }

    if ($rep <10){    # SNPs with 3' end repeated less than 5 times used

        foreach (@snpex){
            if ($_ == $count[$x]) {
                $exon = 'yes';
            } else {
                $exon = 'no';
            }
            last if ($exon eq yes);
        }


## **Mass of primer ($mpr) and melting temperature ($tem) calculation**

# Initially masses of first 10 bases adjacent to each SNP calculated without melting
temperature calculation since primer length always more than 10 bases.

        push (@rep,$rep);
        $adj = $snpos[$x];       # posn just before SNP
        $A=0,$C=0,$G=0,$T=0;         # Counters for nucleotides set to 0

        for ($y=1; $y <=10; $y++) {  #finding nucleotide identity for first 10
            $diff1 = $adj-$y;    #nucleotides adjacent to SNP
            $_=$seq[$diff1];
            if (/a/i) {$A=$A+1}
            if (/c/i) {$C=$C+1}
            if (/g/i) {$G=$G+1}
            if (/t/i) {$T=$T+1}
        }
```

```
        $mpr =($A*313.21)+($C*289.18)+($G*329.21)+($T*304.2)-61.96; #Mass of primer
```

#Melting temp and mass determined as primer sequence for each SNP extended by one base at a time for each SNP. If primer meets set criteria, then primer accepted and added to the hasharray #unsorted for the SNP to sort by MW thus generating a set of primer for the SNP. Then process is repeated for next SNP.

#The criteria used for selection are melting temperature and molecular weight cutoff. Melting temperature maintained above 45 C to prevent non-specific annealing. Lowest molecular weight cutoff used that includes desired/largest number of SNPs.

```
        for ($y=11; $y <=13; $y++) {
           $diff1 = $adj-$y;
           $_=$seq[$diff1];
           if (/a/i) {$A=$A+1, $mpr=$mpr+313.21}
           if (/c/i) {$C=$C+1, $mpr=$mpr+289.18}
           if (/g/i) {$G=$G+1, $mpr=$mpr+329.21}
           if (/t/i) {$T=$T+1, $mpr=$mpr+304.2}
           $tem = ($A+$T)*2 +($C+$G)*4;              #$tem is melting temp
           if (($tem > 45)and($mpr > 5490)) {
                 if ($exon eq yes) {
                    $primex {"$count[$x] $y"} = "$mpr";
                 } else {
                    $unsorted {"$count[$x] $y"} = "$mpr";
                 }
           }
        }

        for ($y=14; $y <=41; $y++) {
           $diff1 = $adj-$y;
           $_=$seq[$diff1];
           if (/a/i) {$A=$A+1, $mpr=$mpr+313.21}
           if (/c/i) {$C=$C+1, $mpr=$mpr+289.18}
           if (/g/i) {$G=$G+1, $mpr=$mpr+329.21}
           if (/t/i) {$T=$T+1, $mpr=$mpr+304.2}
           $tem = 64.9 + (41*($C+$G-16.4)/($A+$T+$C+$G));
           if (($tem > 45)and($mpr > 5490)) {
                 if ($exon eq yes) {
                    $primex {"$count[$x] $y"} = "$mpr";
                 } else {
                    $unsorted {"$count[$x] $y"} = "$mpr";
                 }
           }
        }
   }
}
```

```
## **Selection of single primer for each SNP with difference in mass of adjacent primers such that all 4 extensions can be detected during simultaneous MS analysis knowing masses of B-ddA, B-ddC, B-ddG & B-ddU**



foreach $ele (sort {$primex{$a} <=> $primex{$b}} keys %primex) {
   #sorts the elements of %unsorted by mass.

   $rem=-1;                          # primers with specified mass diff
   foreach $char (@count) {
         $rem=$rem+1;
         if ($ele =~/^$char/){      # checks to see if SNP has primer designed
            if (($primex{$ele}-$curr) >105){ # sifts thru %unsorted and picks
                  splice (@count,$rem,1);   # removes SNP from @count if primer selected
                  $curr = $primex{$ele};
                  $final {"$ele"} = "$curr"; #primer mass with SNP to hasharray final
                  push (@final,$curr);
            }
         }
   }
}

foreach $ele (sort {$unsorted{$a} <=> $unsorted{$b}} keys %unsorted) {
   #sorts the elements of %unsorted by mass.

   $rem=-1;                          # primers with specified mass diff
   foreach $char (@count) {
         $rem=$rem+1;
         if ($ele =~/^$char/){      # checks to see if SNP has primer designed

            $cond = 'no';
            $curr=$unsorted{$ele};
            @sortfin= sort {$a <=> $b} values %final;
           $last=$sortfin[$#sortfin];
            $first=$sortfin[0];

            if (($first -$curr)>105) {
                  $cond = 'yes';
            } elsif (($curr-$last)>105) {
                  $cond = 'yes';
            } else { for ($x=0; $x <= $#sortfin-1; $x++) {
                        if (($curr-$sortfin[$x])>105) {
                           if (($sortfin[$x+1]-$curr)>105) {
                                 $cond = 'yes';
                           }
                        }
                        last if ($cond eq yes);
```

```
                    }
                }

            if ($cond eq yes) {
                splice (@count,$rem,1);   # removes SNP from @count if primer selected
                $curr = $unsorted{$ele};
                $final {"$ele"} = "$curr"; #primer mass with SNP to hasharray final
                push (@final,$curr);
            }
        }
    }
}
```

Selected primers for each SNP are arranged in ascending order of molecular weight and their length and SNP listed alongside.

```
print " 100+SNP #  length   Primer sequence       Primer Molecular weight\n";
foreach (sort {$final{$a} <=> $final{$b}} keys %final) {
                @curr=split /\s+/, $_;
                $list{@curr[0]}=@curr[1];
                $start=$snpos[@curr[0]-101]-@curr[1];
                $end=$snpos[@curr[0]-101]-1;
                $priseq=join '', @seq[$start..$end];
                $list{$_}=$priseq;
                print "       $_       $list{$_}       $final{$_}\n";
}
```

REFERENCES

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304-51
2. Collins FS, Morgan M, Patrinos A. 2003. The Human Genome Project: lessons from large-scale biology. *Science* 300:286-90
3. Kruglyak L, Nickerson DA. 2001. Variation is the spice of life. *Nat. Genet.* 27:234-6
4. Li WH, Sadler LA. 1991. Low nucleotide diversity in man. *Genetics* 129:513-23
5. Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-82
6. Collins FS, Brooks LD, Chakravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8:1229-31
7. Thorisson G, Smith A, Krishnan L, Stein L. 2005. The International HapMap Project Web site. *Genome Research*
8. Rocha D, Gut I, Jeffreys AJ, Kwok PY, Brookes AJ, Chanock SJ. 2006. Seventh international meeting on single nucleotide polymorphism and complex genome analysis: 'ever bigger scans and an increasingly variable genome'. *Hum Genet* 119:451-6
9. Altshuler D, Brooks L, Chakravarti A, Collins F, Daly M, Donnelly P. 2005. A haplotype map of the human genome. *Nature* 437:1299-320
10. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, et al. 1998. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 19:233-40
11. McLeod HL. 2005. Pharmacogenetic analysis of clinically relevant genetic polymorphisms. *Clin Infect Dis* 41 Suppl 7:S449-52
12. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-33
13. Olivier M. 2003. A haplotype map of the human genome. *Physiol Genomics* 13:3-9
14. Suh Y, Vijg J. 2005. SNP discovery in associating genetic variation with human disease phenotypes. *Mutat. Res.* 573:41-53
15. Cantor CR. 2005. The use of genetic SNPs as new diagnostic markers in preventive medicine. *Annals of the New York Academy of Sciences* 1055:48-57
16. Kwok PY, Gu Z. 1999. Single nucleotide polymorphism libraries: why and how are we building them? *Mol Med Today* 5:538-43
17. Schork NJ, Fallin D, Lanchbury JS. 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* 58:250-64
18. Roses AD. 2002. SNPs--where's the beef? *Pharmacogenomics J* 2:277-83
19. Guzey C, Spigset O. 2004. Genotyping as a tool to predict adverse drug reactions. *Curr Top Med Chem* 4:1411-21
20. Tost J, Gut IG. 2005. Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin. Biochem.* 38:335-50
21. Daiger SP. 2005. Genetics. Was the Human Genome Project worth the effort? *Science* 308:362-4
22. Pirmohamed M, Park BK. 2001. Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 22:298-305
23. Marsh S, Kwok P, McLeod HL. 2002. SNP databases and pharmacogenetics: great start, but a long way to go. *Hum Mutat* 20:174-9
24. McCarthy JJ, Hilfiker R. 2000. The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat Biotechnol* 18:505-08

25. Wormhoudt LW, Commandeur JN, Vermeulen NP. 1999. Genetic polymorphisms of human N-acetyltransferase, cytochrome P450, glutathione-S-transferase, and epoxide hydrolase enzymes: relevance to xenobiotic metabolism and toxicity. *Crit Rev Toxicol* 29:59-124
26. Ingelman-Sundberg M. 2004. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends in Pharmacological Sciences* 25:193-200
27. Guengerich FP, Wu ZL, Bartleson CJ. 2005. Function of human cytochrome P450s: characterization of the orphans. *Biochem Biophys Res Commun* 338:465-9
28. Rendic S. 2002. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* 34:83-448
29. Ingelman-Sundberg M, Oscarson M, Daly AK, Garte S, Nebert DW. 2001. Human cytochrome P-450 (CYP) genes: a web page for the nomenclature of alleles. *Cancer Epidemiol Biomarkers Prev* 10:1307-8
30. Gunderson KL, Kuhn KM, Steemers FJ, Ng P, Murray SS, Shen R. 2006. Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics* 7:641-8
31. Judson R, Salisbury B, Schneider J, Windemuth A, Stephens JC. 2002. How many SNPs does a genome-wide haplotype map require? *Pharmacogenomics* 3:379-91
32. Kim S, Misra A. 2007. SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng. Mar* 28:289-320
33. Syvänen A. 2005. Toward genome-wide SNP genotyping. *Nat Genet* 37:S5-10
34. Gut IG. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 17:475-92
35. Pompanon F, Bonin A, Bellemain E, Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 6:847-59
36. Syvanen AC. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics* 2:930-42
37. Kwok PY. 2001. Methods for genotyping single nucleotide polymorphisms. *Annual review of Genomics and Human Genetics* 2:235-58
38. Chen X, Sullivan PF. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J.* 3:77-96
39. Sokolov BP. 1990. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res* 18:3671
40. Ross P, Hall,L., Smirnov,I.P. and Haff,L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnology* 16:1347-51
41. Medintz I, Wong WW, Berti L, Shiow L, Tom J, et al. 2001. High-performance multiplex SNP analysis of three hemochromatosis-related mutations with capillary array electrophoresis microplates. *Genome Res* 11:413-21
42. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, et al. 2003. Large-scale genotyping of complex DNA. *Nat Biotechnol* 21:1233-7
43. Matsuzaki H, Dong S, Loi H, Di X, Liu G, et al. 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 1:109-11
44. Holland PM, Abramson RD, Watson R, Gelfand DH. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* 88:7276-80
45. Landegren U, Kaiser R, Sanders J, Hood L. 1988. A ligase-mediated gene detection technique. *Science* 241:1077-80
46. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 265:2085-8

47. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, et al. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol* 21:673-8
48. Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-31
49. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, et al. 1999. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* 17:292-6
50. Tost J, Gut IG. 2002. Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.* 21:388-418
51. Kwok PY. 2002. SNP genotyping with fluorescence polarization detection. *Hum Mutat* 19:315-23
52. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-7
53. Ferguson JA, Steemers FJ, Walt DR. 2000. High-density fiber-optic DNA random microsphere array. *Anal Chem* 72:5618-24
54. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242:84-9
55. Ding C, Cantor CR. 2003. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proceedings of the National Academy of Sciences* 100:3059-64
56. Sauer S. 2006. Typing of single nucleotide polymorphisms by MALDI mass spectrometry: Principles and diagnostic applications. *Clin. Chim. Acta* 363:95-105
57. Karas M, Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299-301
58. Wu KJ, Steding A, Becker CH. 1993. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun Mass Spectrom* 7:142-6
59. Tost J, and Gut,I.,G. 2005. Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clinical Biochemistry* 38:335-50
60. Haff LA, Smirnov IP. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* 7:378-88
61. Braun A, Little DP, Koster H. 1997. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* 43:1151-8
62. Cashman J, Zhang J, Leushner J, Braun A. 2001. Population distribution of human flavin-containing monooxygenase form 3: Gene polymorphisms. *Drug Metabolism and Disposition* 29:1629-37
63. Kim S, Edwards JR, Deng L, Chung W, Ju J. 2002. Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Res* 30:e85
64. Sauer S, Gelfand DH, Boussicault F, Bauer K, Reichert F, Gut IG. 2002. Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Res* 30:e22
65. Ross P, Hall L, Smirnov I, Haff L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* 16:1347-51
66. Fei Z, Ono T, Smith LM. 1998. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res* 26:2827-8
67. Sauer S, Gut IG. 2003. Extension of the GOOD assay for genotyping single nucleotide polymorphisms by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 17:1265-72

68. Brookes AJ. 1999. The essence of SNPs. *Gene* 234:177-86
69. Shi MM. 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin Chem* 47:164-72
70. Hellard SL, Ballereau,S.J., Visscher,P.M., Torrance,H.S., Pinson,J., Morris,S.W., Thomson,M.L., Semple,C.A.M., Muir,W.J., Blackwood,D.H.R. et al. 2002. SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res.* 30:e74
71. Meyer K, Fredriksen A, Ueland PM. 2004. High-level multiplex genotyping of polymorphisms involved in folate or homocysteine metabolism by matrix-assisted laser desorption/ionization mass spectrometry. *Clin Chem* 50:391-402
72. Kim S, Edwards JR, Deng L, Chung W, Ju J. 2002. Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Res.* 30:e85
73. Kim S, Ulz,M.E., Nguyen,T., Li,C., Sato,T., Tycko,B. and Ju,J. 2004. Thirtyfold multiplex genotyping of the *p53* gene using solid phase capturable dideoxynucleotides and mass spectrometry. *Genomics* 83:924-31
74. Rettie AE, Jones JP. 2005. Clinical and toxicological relevance of CYP2C9: Drug-drug interactions and pharmacogenetics. *Annu. Rev. Pharmacol. Toxicol.* 45:477-94
75. Xie HG, Prasad HC, Kim RB, Stein CM. 2002. CYP2C9 allelic variants: ethnic distribution and functional significance. *Advanced Drug Delivery Reviews* 54:1257-70
76. Sherry ST, Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M., Sirotkin,K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29:308-11
77. Ingelman-Sundberg M, Oscarson M. 2002. Human CYP allele database: submission criteria procedures and objectives. *Methods Enzymol* 357:28-36
78. Su T, Bao,Z.P., Zhang,Q.Y., Smith,T.J., Hong,J.Y. and Ding,X. 2000. Human cytochrome P450 CYP2A13: predominant expression in the respiratory tract and its high efficiency metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Cancer Res.* 60:5074-79
79. He XY, Tang,L., Wang,S.L., Cai,Q.S., Wang,J.S. and Hong,J.Y. 2006. Efficient activation of aflatoxin B_1 by cytochrome P450 2A13, an enzyme predominantly expressed in human respiratory tract. *Int. J. Cancer* 118:2665-71
80. Zhang X, Chen Y, Liu Y, Ren X, Zhang QY, et al. 2003. Single nucleotide polymorphisms of the human CYP2A13 gene: Evidence for a null allele. *Drug Metabolism and Disposition* 31:1081-85
81. Benson DA, Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Wheeler,D.L. 2005. GenBank. *Nucleic Acids Res.* 33:D34–D38
82. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-10
83. Wen SY, Wang H, Sun OJ, Wang SQ. 2003. Rapid detection of the known SNPs of CYP2C9 using oligonucleotide microarray. *World Journal of Gastroenterology* 9:1342-46
84. Hruska MW, Frye RF, Langaee TY. 2004. Pyrosequencing method for genotyping cytochrome P450 CYP2C8 and CYP2C9 enzymes. *Clin Chem* 50:2392-5
85. Mas S, Crescenti,A., Vidal-Taboada,J.M., Bergonon,S., Cuevillas,F., Laso,N., Molina,R., Ballesta,A., Lafuente,A. 2005. Simultaneous genotyping of CYP2C9*2, *3 and 5' flanking region (C-1189T) polymorphisms in a Spanish population through a new minisequencing multiplex single-base extension analysis. *Eur. J. Clin. Pharmacol.* 61:635-41
86. Sanchez JJ, Børsting,C., Hallenberg,C., Buchard,A., Hernandez,A and Morling,N. 2003. Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci. Int.* 137:74-84

87. Kim S, Shi,S., Bonome,T., Ulz,M.E., Edwards,J.R., Fodstad,H., Russo,J.J. and Ju,J. 2003. Multiplex genotyping of the human beta2-adrenergic receptor gene using solid-phase capturable dideoxynucleotides and mass spectrometry. *Anal. Biochem.* 316:251-58
88. Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. 2004. MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome single-nucleotide polymorphisms. *Anal Chem* 76:6039-45
89. Shahgholi M, Garcia,B.A., Chiu,N.H.L., Heaney,P.J. and Tang,K. 2001. Sugar additives for MALDI matrices improve signal allowing the smallest nucleotide change (A:T) in a DNA sequence to be resolved. *Nucleic Acids Res.* 29:e91
90. Monforte J, Becker C. 1997. High-throughput DNA analysis by time-of-flight mass spectrometry. *Nat. Med.* 3:360-62
91. Giacomini K, Brett C, Altman R, Benowitz N, Dolan M, et al. 2007. The Pharmacogenetics Research Network: From SNP Discovery to Clinical Drug Response. *Clin. Pharmacol. Ther.* 81:328-45
92. Kim S, Misra A. 2007. SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng.* 28:289-320
93. Griffin TJ, Smith LM. 2000. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* 18:77-84
94. Ding C, Cantor C. 2004. Quantitative analysis of nucleic acids–the last few years of progress. *J Biochem Mol Biol* 37:1-10
95. Sauer S. 2007. The essence of DNA sample preparation for MALDI mass spectrometry. *J. Biochem. Biophys. Methods* 70:311-18
96. Holmberg A, Blomstergren A, Nord O, Lukacs M, Lundeberg J, Uhlen M. 2005. The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* 26:501-10
97. Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. 2004. MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome single-nucleotide polymorphisms. *Anal. Chem.* 76:6039-45
98. Misra A, Hong J, Kim S. 2007. Multiplex Genotyping of Cytochrome P450 Single-Nucleotide Polymorphisms by Use of MALDI-TOF Mass Spectrometry. *Clin. Chem.* 53:933-39
99. Green NM. 1990. Avidin and streptavidin. *Methods Enzymol.* 184:51-67
100. Tong X, Smith LM. 1992. Solid-phase method for the purification of DNA sequencing reactions. *Anal. Chem.* 64:2672-77
101. Beavis RC, Chait BT. 1990. Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc. Natl. Acad. Sci.* 87:6873-77
102. Cornett D, Reyzer M, Chaurand P, Caprioli R. 2007. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods* 4:828-33
103. Griffiths J. 2008. A Brief History of Mass Spectrometry. *Anal. Chem.* 80:5678-83
104. Chen W-Y, Chen Y-C. 2007. MALDI MS Analysis of Oligonucleotides: Desalting by Functional Magnetite Beads Using Microwave-Assisted Extraction. *Anal. Chem.* 79:8061-66
105. Roque ACA, Lowe CR. 2006. Advances and applications of de novo designed affinity ligands in proteomics. *Biotechnol. Adv.* 24:17-26
106. Lee W-C, Lee KH. 2004. Applications of affinity chromatography in proteomics. *Anal. Biochem.* 324:1-10
107. Hu B-H, Jones MR, Messersmith PB. 2007. Method for Screening and MALDI-TOF MS Sequencing of Encoded Combinatorial Libraries. *Anal. Chem.* 79:7275-85
108. Diamandis E, Christopoulos T. 1991. The biotin-(strept) avidin system: principles and applications in biotechnology. *Clin. Chem.* 37:625-36

109. Schriemer DC, Yalcin T, Li L. 1998. MALDI Mass Spectrometry Combined with Avidin-Biotin Chemistry for Analysis of Protein Modifications. *Anal. Chem.* 70:1569-75
110. Wang H, Tseng K, Lebrilla CB. 1999. A General Method for Producing Bioaffinity MALDI Probes. *Anal. Chem.* 71:2014-20
111. Abel K, Reneland R, Kammerer S, Mah S, Hoyal C, et al. 2006. Genome-wide SNP association: Identification of susceptibility alleles for osteoarthritis. *Autoimmun. Rev.* 5:258-63
112. Griffin TJ, Smith LM. 2000. Genetic Identification by Mass Spectrometric Analysis of Single-Nucleotide Polymorphisms: Ternary Encoding of Genotypes. *Anal. Chem.* 72:3298-302
113. Misra A, Kim S. 2008. Microbead device for isolating biotinylated oligonucleotides for use in mass spectrometric analysis. *Anal. Biochem.* 384:96-100
114. Fortier M-H, Bonneil E, Goodley P, Thibault P. 2005. Integrated Microfluidic Device for Mass Spectrometry-Based Proteomics and Its Application to Biomarker Discovery Programs. *Anal. Chem.* 77:1631-40
115. Bhattacharyya A, Klapperich CM. 2006. Thermoplastic Microfluidic Device for On-Chip Purification of Nucleic Acids for Disposable Diagnostics. *Anal. Chem.* 78:788-92
116. Hellard SL, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, et al. 2002. SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res* 30:e74

CURRICULUM VITA

ASHISH MISRA

EDUCATION

2002-2009 **Rutgers, The State University of New Jersey, New Brunswick, NJ**
Ph.D. in Chemical & Biochemical Engineering

2002-2005 **Rutgers, The State University of New Jersey, New Brunswick, NJ**
M.S. in Chemical & Biochemical Engineering

1998-2002 **University Institute of Chemical Technology (UICT), Mumbai, India**
B.E. in Chemical Engineering

PUBLICATIONS

Misra A and Kim S (2008) Microbead device for isolating biotinylated oligonucleotides for use in mass spectrometric measurement. *Anal. Biochem*. 384:96-100 (2008)

Misra A, Hong JY and Kim S (2007) Multiplex Genotyping of Cytochrome P450 Single-Nucleotide Polymorphisms by Use of MALDI-TOF Mass Spectrometry. *Clin. Chem*. 53:933-939

Kim S and Misra A (2007) SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng.* 28:289-320