

QUANTITATIVE ANALYSIS OF GENETIC VARIATIONS USING MOLECULAR
AFFINITY AND MALDI-TOF MASS SPECTROMETRY

by

AANA MOON KIM

A Thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Biomedical Engineering

written under the direction of

Professor Sobin Kim

and approved by

New Brunswick, New Jersey

May 2007

ABSTRACT OF THE THESIS

Quantitative analysis of genetic variations using molecular affinity and MALDI-TOF
mass spectrometry

By Aana Moon Kim

Thesis Director:
Professor Sobin Kim

A system for DNA quantification is demonstrated utilizing matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and a molecular affinity system. MALDI-TOF MS can detect nucleic acids in milliseconds with high resolution, which allows for a high throughput sample analysis. The resulting mass spectra can produce more accurate results than conventional methods based on gel electrophoresis especially when secondary structures exist in the sample DNA molecules. Secondary DNA structures can lead to incomplete separation in the gel and reduce the accuracy in DNA sequence determination. MALDI-TOF MS measures the intrinsic property of DNA mass, and thus allows unambiguous detection of DNA fragments. MALDI TOF MS can be further facilitated by molecular affinity systems such as biotin-streptavidin as they increase the accuracy and throughput of MS.

Our approach is based on SPC-SBE, a previously developed method for high throughput genotyping employing molecular affinity, and involves optimization in both sample preparation and MALDI-TOF MS spectral analysis by signal processing. Creating small sample spots with a stainless steel needle to reduce sample crystal heterogeneity improved the DNA quantification system. To further reduce the impact of salt adduct peaks, signal-processing techniques were used for better signal separation. A basic signal processing technique involved ensemble average and ensemble standard deviations of spectra. Ensemble averaging is a simple and powerful signal processing technique to reduce noise. Additionally, the microbead pipette tip device was tested as a new method to rapidly isolate biotinylated DNA fragments and increase throughput of the molecular affinity system. We demonstrate our method provides high quantification

accuracy, and therefore can be efficiently used for the application of rapid high-throughput gene expression monitoring.

Table of Contents

1. Preliminary	
1.1. Abstract	page ii
1.2. List of tables	page v
1.3. List of figures	page vi
2. Background	page 1
3. Materials and Methods	
3.1. Synthetic system	page 12
3.2. PCR products	page 12
3.3. Developed new quantitative analysis technique for biotinylated DNA fragments	page 13
4. Results and Discussion	
4.1. Matrix optimization	page 15
4.2. Synthetic system results	page 16
4.3. Improvements in quantitative analysis	page 16
4.4. Small spotting technique and MATLAB 6.0 results	page 18
4.5. Microbead pipette tip device	page 19
5. Conclusions	page 22
6. References	page 23
7. Matlab program	page 36

Lists of tables

1. Table 1: DNA sequences of the templates and primers utilized in experiments. Page 25.

List of figures

1. Figure 1 A schematic of a MALDI-TOF MS. Page 26
2. Figure 2 The effect of matrix composition on the peak signal quality. Page 27
3. Figure 3 A calibration curve using a different system of synthetic templates and primer. Page 28
4. Figure 4 MATLAB 6.0 graphs illustrate the ensemble-averaged signals that are being subtracted to obtain a more accurate representation of each signal. Page 29
5. Figure 5 Images were taken to demonstrate the size differences when using the improved DNA quantification system with small sample spots. Page 30
6. Figure 6 Calibration curve generated from 2 synthetic templates based on a short region of the CYP2C9 gene. Page 31
7. Figure 7 Quantification of PCR products generated from genomic DNA of a HepG2 cell line. A synthetic template of 2 pmol was used as an internal standard. Page 32
8. Figure 8 Prototype of a microbead pipette tip device. Page 33
9. Figure 9 Initial data to compare the incubation and binding step of biotinylated extension primers to streptavidin coated microbeads in a pipette tip to the conventional method using a centrifuge tube. Page 34
10. Figure 10 Different lengths of DNA primer extension products were also tested in a centrifuge tube versus a pipette tip. Page 35

©2007

Aana Moon Kim

ALL RIGHTS RESERVED

1. Background:

The ability to perform with a high level of throughput or multiplexing is critical in a wide range of biomedical applications²⁸. For instance, in clinical genetic studies one major requirement is the analysis of sequence variations on a large scale in a cost and time efficient manner. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) MS has the capability to perform highly multiplexed experiments because a single mass spectrum has many detection channels or many data signals can be obtained from one sample. The high throughput capability originates from the high duty cycle of the TOF analyzer. Data acquisition can also be automated for increased time efficiency³¹.

For this reason, several MALDI-TOF MS approaches have been developed in biomedical research, particularly for single nucleotide polymorphism (SNP) analysis. MALDI-TOF MS operates to generate the spectra, which identifies the gene fragments in a sample (figure 1). A DNA sample must be purified and then mixed with a matrix solution, before being applied to a sample plate as a 0.5-1 μ L droplet. A matrix solution is a UV absorbing weak organic acid, such as 3-hydroxypicolinic acid²⁶. It was not until after the discovery that 3-hydroxypicolinic acid as a matrix could ionize DNA fragments in a MS that genetic research with MALDI-TOF MS could take place. The matrix functions with the laser to ionize the DNA fragments²⁶. The matrix also serves to protect the analyte from being fragmented by the excess laser energy²⁷. According to Beer's Law, the wavelength absorption from the UV laser in a MALDI-TOF MS to a sample mixed with matrix, should be enough energy to cause ionization of the analyte²⁴. These ions then move through the flight tube to the detector. The masses of the ions are

calculated based on the kinetic energy of these ions and the amount of time the ions take to move from a fixed point to the detector. A voltage grid ensures that all the ions travel up the flight tube at the same time. The detector is connected to a computer where the resulting mass spectrum is displayed, and where the user can operate the instrument. With the use of a TOF mass analyzer or detector, the mass determination of a 25-50mer oligonucleotide sample can be measured with an accuracy of $\pm 0.01\%$ ²⁶. The explanation for why or how MALDI-TOF MS permits the analysis of such large biomolecules is due to the “soft ionization” mechanism between the matrix and the UV laser¹. Essentially the soft ionization term refers to the fact that the analyte is not destroyed during the ionization process. MALDI-TOF MS has applications in both DNA analysis such as SNP genotyping, and proteins.

Several multiplexing techniques have been developed with MALDI-TOF MS as the basis for genetic analysis. In 1998, Ross *et al.* have demonstrated that their PinPoint assay can be used for a 12-fold multiplexing of SNPs²⁸. SNPs were selected from different locations on a BRCA1 gene. All 12 SNPs were PCR amplified in a single tube. The PCR products were unambiguously identified in the resulting mass spectrum. In contrast, current genotyping technology based on MALDI-TOF has progressed to 50-fold multiplexing as shown by Misra *et al.*³². They have utilized a molecular affinity system to isolate only the extended primers as a result of single base extension reactions. Biotin and streptavidin is a specific and strong binding affinity. All 50 SNPs were selected from a CYP 450 gene and also clearly detected in the mass spectra³². In comparing both studies, SNP genotyping is rapidly progressing to become automated, high-throughput,

and multiplexed. Both studies also demonstrate the importance of MALDI-TOF MS as an analytical molecular biology tool.

1.1 Molecular Affinity

The molecular affinity system of biotin and streptavidin has expanded to include many diverse applications, such as immunological assays, electron microscopy, and affinity chromatography. This particular molecular affinity system is applied in conjunction with MALDI-TOF MS for DNA quantification. Originally, the biotin-streptavidin binding system was a natural defense mechanism to inactivate biotinylated enzymes thus causing growth inhibition of bacteria that depend on those enzymes¹². Avidin can be found in egg whites, while streptavidin can be found in *Streptomyces avidinii*. Biotin is also a vitamin that can be found in small quantities in all living cells. Generally, biotin acts as a cofactor for carboxylating enzymes. Biotin can be derivatized for further labeling purposes¹². There are several properties of this particular affinity system that makes it very useful in biological research fields. The binding constant is $10^{15} \text{ L} \cdot \text{mol}^{-1}$, which is such a high constant that the binding will not easily change as a result of pH or repeated washings with buffers. Biotin also does not change the biological activity or characteristic of the molecule it is bound to. The biotin-streptavidin affinity also has applications with nucleic acids. These applications include a broad range of experimental techniques for instance; nick translation, random priming, transcription, or PCR¹². The enzymatic reaction with DNase I and DNA polymerase I with biotinylated dNTP are used in nick translation. The use of biotinylated dATP with DNA polymerase I was demonstrated by a study described by Wu *et al*¹³. Nick translation was used to detect DNA damage in erythrocyte lysis after intracerebral

hemorrhage and whether this causes oxidative brain damage. Although at least one of the components of the technique must be biotinylated, for example dideoxynucleotides, the reagents to create biotinylated components are commercially available or can be purchased already in a biotinylated state. Another advantage of the biotin-streptavidin affinity is that once both molecules are bound, both molecules were found to be very stable under harsh chemical conditions. Streptavidin and avidin have four binding sites, which makes it possible to have multiple biotin molecules attach to these molecules for greater detection sensitivity. DNA hybridization assays can also be performed using the biotin-streptavidin system, by using biotinylated probes for Southern and Northern blotting. DNA sequencing techniques have also been described for biotinylated primers. Colorimetric detected DNA sequencing assay utilizes biotinylated primers with direct blotting electrophoresis to relocate DNA fragments to a nylon membrane. The biotinylated primers are then viewed with streptavidin-alkaline phosphatase. This biotin-streptavidin affinity has been employed for DNA quantification by mass spectrometry.

1.2 Other Modified Dideoxynucleotides

Other recently developed techniques for DNA analysis involve the use of modified dideoxynucleotides. There is continued interest in high-throughput DNA sequencing for applications in further understanding the human genome for clinical medicine and health care. Reversible terminators were created from the attachment of a fluorophore to 3'-*O*-allyl-modified nucleotides¹⁴. Despite this modification, DNA polymerase will still recognize these nucleotides and incorporate them appropriately. These newly modified nucleotides will then be used in conjunction with synthesis on a solid surface. DNA strands will be immobilized onto a DNA chip and the modified

nucleotides will be identified using a four-color fluorescent scanner. A different fluorophore will be attached to each of the four types of nucleotides such that the fluorescent emission differs based on the identity of the nucleotide. A small reversible moiety is attached to the 3'-OH group of the nucleotide so that after incorporation, the identity of the nucleotide can be determined with the scanner and then another round of incorporation can begin again. In this way the DNA sequences immobilized on the DNA chip can be identified through multiple rounds of nucleotide incorporation. The DNA chip allows for parallel identification of many different DNA strands for increased efficiency. Validation of the extension products with these new nucleotides was performed using MALDI-TOF MS. Finn *et al.* at Amersham Biosciences¹⁵ have also developed another new DNA sequencing technique using modified dideoxynucleotides. The dideoxynucleotides were modified using a negatively charged linker between the base and a fluorescent label. The negatively charged linker removes the need for further purification of DNA extension products so that the samples can be directly loaded for gel electrophoresis. The use of these modified nucleotides not only prevents more purification steps but also reduces gel compression artifacts in DNA sequencing when used with dGTP. DNA polymerase can still incorporate these modified nucleotides.

1.3 Other Quantitative Analysis Techniques

Since the completion of a draft of the human genome, quantifying gene expression levels has been of particular interest as it could reveal gene functions. The quality of gene expression data is critical to interpretation. Investigation into genes is essential to understanding processes at a cellular level for antibiotic design, new drug targets, and stem cells for therapeutic applications¹⁷. Real-time quantitative PCR (rtPCR)

has been recognized to be a popular standard technique for gathering gene quantification data¹⁹. It also requires more controls than a standard PCR due to the other variables involved in analysis. With rtPCR, the amount of PCR products accumulating during the exponential phase of PCR is measured by fluorescent signals. A plot of fluorescent signal versus number of PCR cycles is used to calculate the amount of starting DNA template or mRNA expression levels. There are three general methods to detect the PCR products and mRNA expression levels: TaqMan probes, molecular beacons, and SYBR Green I intercalating dyes¹⁹.

Liu *et al.* have reported a TaqMan probe microarray for quantitative analysis using real-time PCR¹⁸. This probe array combined the use of TaqMan probes and real-time PCR by using a glass microarray to increase the signal strength of the activated fluorescent label on the Taqman probe. The TaqMan probes were modified by adding a 3'- amino group for immobilization onto a glass slide in a microarray format. Then real-time PCR is performed on the glass slide. The 5'- nuclease activity of DNA polymerase will cleave the quencher component when in contact with the nucleic acid targets. As a result of the increased physical distance between the quencher and the fluorescent dye, there is a significant increase in the signal strength from the fluorescent dye after cleavage of the quencher. Here the quencher has been relocated to the 5' end of the modified TaqMan probe so it can be cleaved off during PCR. Data acquisition can be obtained with a laser scanning confocal microscope. This is in contrast to the conventional TaqMan probe. The fluorescent label is located at the 5' ends and a quencher dye at the 3' ends. During PCR, the quencher will still be released from the probe.

1.4 Other Quantitative Analysis Techniques Using MALDI-TOF MS

Other DNA quantification techniques have also been developed using the MALDI-TOF MS as a detection system. One such technique is the quantification of CpG methylation with the GOOD assay²¹ as described by Tost and Gut *et al*²⁰. The GOOD assay is based on the principle that the analysis of DNA molecules depends on the charge state of those molecules. The GOOD assay begins with a standard PCR reaction to produce enough DNA templates for subsequent reactions. Modified primers are then used to generate charge tagged DNA products. These primers have 3 phosphorothioates and an optional quaternary ammonium charged group with ddNTPs or α -S-ddNTPs. The uncharged portions of the primers are then digested by phosphodiesterase II, followed by an alkylation reaction that leaves only charged DNA extension products. MS can then detect the positively charged DNA molecules in positive ion mode, while negatively charged DNA molecules are detected in negative ion mode. The resulting final products have enough mass separation for adequate peak resolution in the MS spectra. By generating charged DNA extension products, the signals remain resolved and provide a strong signal despite the presence of salts or detergents²¹. The GOOD assay was then further modified to analyze CpG positions both in single and multiplexed reactions²⁰. DNA methylation has been found to be associated with a number of diseases, most notably in tumor cells. It often occurs on mammalian DNA strands on the 5' position of cytosine. Specific methylation patterns have been noticed for different types of tumors as well as being tumor-type specific. Therefore highly accurate quantization of these methylated sites is required for further analysis, which could lead to efficient prostate cancer diagnosis. Genomic DNA has been first treated with bisulphite, after which the

GOOD assay is applied to the treated genomic DNA samples. Bisulphite treatment results in the conversion of all non-methylated cytosines to uracils, but methylated cytosines remain unchanged by the treatment.

Another MALDI-TOF MS based technique was developed by Sequenom, Inc (San Diego, CA)²². They have developed the MassEXTEND system for a chip-based assay that first binds PCR products to a silicon chip. This is followed by a primer-oligo extension reaction (PROBE) that occurs on the surface of the chip. The PROBE reaction is designed to extend primers specifically designed to bind next to sites of interest and to efficiently react on the surface of the chip. Matrix is then directly added onto each spot where the PROBE reaction has occurred for MS analysis. Peak resolution and separation was achieved by designing the primers to be extended by either one nucleotide or two nucleotides depending on the sequence of the DNA template. The MassEXTEND reaction was also modified for gene expression analysis by performing real competitive PCR (rcPCR). This particular method starts with an RNA sample that is converted to cDNA by reverse transcription. A known amount of competitor is added to the sample for PCR amplification with the cDNA sample. The competitor acts as an internal standard for quantification analysis. Then the PROBE reaction is performed for detection and quantification by MALDI-TOF MS. It should be noted that this particular method does not remove leftover reactants from the reaction, including unextended primers.

1.5. SPC-SBE

Here we report a high-throughput DNA quantification method that can be further developed for the application of gene expression monitoring, utilizing MALDI-TOF MS and a molecular affinity system. Previous studies have proven that highly accurate

frequency estimates for pooled samples can be obtained with MALDI-TOF MS detection. The difference between pooled and individual genotype frequencies is 1-2%. MALDI-TOF MS has also been shown to be a high-throughput platform with an estimated 10,000-500,000 genotypes per day. The multi-channel detector also allows for high levels of multiplexing, or multiple polymorphisms can be detected in a single reaction. This high multiplexing capability results in a cost-effective and efficient technique. In contrast, conventional gel electrophoresis-based techniques require several hours for completion.

However, there are still a few challenges that are associated with genetic variation analysis with MALDI-TOF MS. Current limitations in MALDI-TOF MS include the requirement of stringent sample preparations. Salts must be removed before analyses to reduce the formation of salt adduct peaks. DNA can be quantified with MALDI-TOF MS by measuring the DNA peak area in the spectra. The quantification accuracy can be decreased as results of the presence of salt adduct peaks, leftover single base extension (SBE) reactants, and the heterogeneous crystallization of sample spots with matrix². Salt adducts contribute to the error by causing inaccurate peak area measurement specifically by broadening peaks². These salt adduct peaks are formed due to the phosphodiester bonds in the backbone of the DNA molecule³. Peak broadening can cause overlap of adjacent peaks. Peak broadening decreases resolution between peaks of similar mass. Leftover SBE reactants will also inhibit detection of correct SBE products located next to SNP sites by again reducing peak resolution or actually overlapping with the correct SBE product signal.

These limitations can be addressed through the utilization of a molecular affinity system as shown in the SPC-SBE approach. SPC-SBE utilizes biotinylated dideoxynucleotides (ddNTPs) in SBE reactions to isolate only the DNA products of interest, which distinguishes this technique from other available DNA analysis methods. DNA extension products are therefore biotinylated on the 3'-end. Streptavidin-coated magnetic microbeads capture only the DNA extension products, and other unwanted leftover reagents are washed away with buffers. SPC-SBE facilitates MALDI-TOF MS by increasing the accuracy and throughput. Owing to the use of biotinylated ddNTPs, the smallest mass difference between two DNA fragments is 16 Da compared to 9 Da with standard ddNTPs. This increases the mass resolution and increases accuracy of sequence identification. Further, this approach permits for higher degree of multiplexing and minimal effort into primer and assay design optimization. The higher multiplexing format provided by SPC-SBE decreases the overall cost of the technique by using smaller amounts of reagents and equipment. By utilizing the molecular affinity of biotin and streptavidin, DNA extension products are isolated before MALDI-TOF MS analysis. The resulting mass spectrum is now free of non-extended primer peaks and associated dimers, which increases the accuracy and scope of multiplexing. Also, non-extended primers compete for ion current that reduces the detection sensitivity of DNA extension products of interest. Polymorphisms can be identified by the difference between the primer mass and the detected extension product.

1.6. Improved Quantification of Biotinylated DNA Fragments

Another advantage of SPC-SBE utilization is the higher DNA quantification accuracy because of the isolation of DNA extension products of interest. Higher DNA

quantification accuracy can be particularly useful for gene expression analysis. Gene expression patterns in complicated diseases, such as cancers, are believed to profile the activity of related gene functions and can be used to illustrate the physiological characteristics of the disease at the molecular level⁷. Our approach involves optimization in both sample preparation and signal processing and shows high accuracy in the quantitative analysis of the biotinylated DNA fragments produced in SPC-SBE, which has not been demonstrated previously. The signal processing technique of ensemble averaging has yet to be applied to MALDI-TOF MS spectra. It is used for spectral noise reduction and to obtain a representative signals for the signal subtraction of one extension product signal from a mixture of extension products data. Here using a model gene system, we demonstrate that reducing sample spot size and applying new spectral processing techniques can accurately quantitate biotinylated DNA molecules.

Materials and Methods

1) *Single Base Extension reaction and Solid Phase Capture:*

We have demonstrated our DNA quantification system based on MALDI-TOF MS and molecular affinity using 2 different systems: a synthetic system and PCR products. The DNA templates of the synthetic system mimicked a region of the HIV gene (table 1). The reaction mixture contained 20 pmol of SNP primer, 20 pmol of one template, 1 U of ThermoSequenase enzyme, and 1 μ L of reaction buffer in a 10 μ L total volume. The reaction mixture was subjected to 35 thermal cycles of 94.0°C for 10 seconds and 44.0 °C for 30 seconds. For each reaction, the yield of the reaction was checked to ensure that the reaction went to completion. After SBE, the samples were purified using SPC as previously described by Kim *et al*³⁰.

1.1) *PCR products system:*

The other system that was used to demonstrate the quantification system was with PCR products generated from a region of the CYP2C9 gene (exon 2). PCR was performed to amplify one 500 base pair region of the CYP 2C9 gene exon 2 (table 1). The 10 μ L PCR reaction consisted of 300 ng of genomic DNA isolated from a HepG2 liver cell line, 1 nmol of dNTP, 1 μ L of 10X PCR buffer, 7.5 pmol of reverse and forward primers, 0.5 U of JumpStart RedAccuTaq DNA polymerase (Sigma, MO). After a 2 minute hot start, the PCR program was performed with 40 cycles of 94.0°C for 30 seconds, 62.0°C for 30 seconds, 72.0°C for 2 minutes, and a final extension of 72.0°C for 10 minutes. After PCR, 2.0 μ L of PCR product from a sample was run on a 2% agarose gel with 1X TAE buffer for visualizing the PCR products to check for product quality. The region was successfully amplified without an indication of nonspecific amplification.

Excess dNTP and primers were degraded using 1U of shrimp alkaline phosphatase (USB Corporation, Cleveland OH) and 1U of *Escherichia coli* exonuclease I (USB Corporation, Cleveland OH) in 1X phosphatase buffer, incubated at 37.0°C for 1.5 hours, and then heated to 94.0°C for 15 minutes for inactivation of the enzyme. A DNA synthetic system with 2 SBE primers was designed based on CYP2C9 exon 2. Both SBE primers were designed to anneal next to 2 SNP sites and were checked for cross-hybridization. The sequence of both templates (table 1) was based on a section of the CYP 2C9 region that was amplified in the previously described PCR reaction. The concentration ratio between both templates was varied and the peak area ratio for each concentration ratio was calculated to obtain a calibration curve. The 10 µL single base extension reaction consisted of 10 pmol each of 2 primers, 25 pmol each of biotinylated dideoxynucleotides, 1 U of ThermoSequenase enzyme and 1 µL of reaction buffer. The reaction mixture was subjected to 40 cycles of 94°C for 30 seconds and 51.0°C for 30 seconds. For each reaction, the yield of the reaction was checked to ensure that the reaction went to completion. After SBE, the samples were again purified using SPC as previously described by Kim *et al*³⁰.

2) *Sample spot preparation for MALDI-TOF MS:*

Small sample spots were achieved by using a stainless steel needle to deposit the sample solution of purified DNA and matrix onto a MS sample plate. After the needle is dipped into the sample solution, the needle is then quickly touched onto the surface of a MALDI-TOF MS sample plate to create multiple small crystal spots. Proof-of-concept studies using a stainless steel needle and the synthetic DNA system were performed to

prove the viability of this approach. SPC-SBE will be performed to generate purified DNA extension products that carry information on the identity of nucleotides at mutation sites. A spectrum from each spot was acquired with 100 laser shots and then all the spectra are accumulated per replicate. Once verified, the process was further optimized to ensure reproducibility between the smaller sample spots.

3) *Data procession on mass spectra:*

Ensemble averaging is a simple and powerful signal processing technique to reduce noise⁹. An ensemble or groups of signals, in this case MS spectrum, are averaged on a point-by-point basis, over all the signals in the ensemble. Ensemble standard deviation is similarly calculated by determining the standard deviation on a point-by-point basis. Ensemble averaged signal subtraction has been demonstrated with primer extensions of ddC and ddA using synthetic DNA templates. Both primer extensions were generated through separate SBE reactions, and then mixed in known concentration ratios. The salt adduct peak of ddC extension overlapped with the major peak of the ddA extension. The signal subtraction, in MATLAB 6.0 (The Mathworks, Inc., Natick, MA), reduced the effect of salt adduct peaks from the ddC extension, thereby increasing quantification accuracy of ddA extension.

Results and Discussion:

1) *Matrix optimization:*

The matrix formulation was first optimized for quantitative analysis. The matrix commonly utilized for DNA analysis is 3-hydroxypicolinic acid. It was selected for DNA analysis based on its ability to produce spectrum of high signal to noise ratio from lower chemical noise²³. Signal saturation, unresolved overlapping ion signals, and noise broadening distort peak shapes and thus the mass accuracy. Different matrices can have different sensitivities to interference from impurities such as buffer salts. An additive can be mixed with the matrix to enhance soft ionization mechanism of MALDI-TOF MS. Absorption of laser wavelength is not a restrictive criterion; instead matrices are found empirically²⁴. Additives can assist in the ionization processes. Another essential matrix property is that it prevents analyte aggregation. Individual analyte molecules should be well incorporated into the matrix. The matrix should have a high tolerance to possible contaminants because of the role matrix has in ionization. The matrix should also have no chemical reactivity and vacuum stable. It must also provide fast, easy, and reproducible results. A commonly used additive is diammonium citrate. The concentration ratio between additive and matrix was varied from 0.011M diammonium citrate with 0.34M 3-hydroxypicolinic acid to 0.036M diammonium citrate with 0.29M 3-hydroxypicolinic acid to determine the optimal additive to matrix ratio. In these experiments the matrix formulation of 0.034M diammonium citrate and 0.32M 3-hydroxypicolinic acids showed the most consistent ion count and peak area signals as shown in figure 2, and thus was selected for future analysis.

2) *Results from synthetic system:*

A calibration curve was generated from SBE-SPC using the synthetic system (figure 3). Two synthetic templates and one primer were used in the SBE reaction. The primer was designed to anneal next to the SNP site. The amounts of each synthetic template were varied to generate the calibration curve. This synthetic template system produced primer extension products of biotin-ddC and biotin-ddA with masses of 6107 Da and 6131 Da respectively. From the data, the peak area of the C extension correspondingly decreased as the amount of added synthetic template decreased in the reaction. The concentration ratio ranged from 100% to 5%, where for 100% equal amounts of each template were added to the SBE reaction. With this calibration curve, the need for signal separation was apparent because the 2 SBE extension products were close in mass. The salt adduct peaks from the C extension would overlap with the major peak signal for the A extension. Since the salt adduct peaks are associated with the major peak signal, then the salt adduct peak of the C extension is contributing to the major peak signal of the A extension. In order to increase quantitative accuracy, the signals must be separated and also measures the peak area of the accompanying salt adduct peaks. This was performed by ensemble average of mass spectra and then subtraction of signals in MATLAB 6.0.

3) *Improvements in quantitative analysis:*

The signal isolation graphs (figure 4) are the result of performing a subtraction of biotin-ddC extension signal from a mixed 100% concentration ratio of biotin-ddC and biotin-ddA extensions with the use of MATLAB 6.0 programming. The raw data is

shown in the first step. All spectra are converted into .txt files for input into MATLAB. In step 2, the ensemble average of biotin-ddC extension, biotin-ddA extension, and the equal concentration mixture of both was calculated. The final step is to align the biotin-ddC extension with the mixed data and subtract the two data sets. The salt adduct peak from biotin-ddC extension is subtracted from the major peak of the biotin-ddA extension. The baseline was adjusted by calculating the location of the valleys on either side of the peak signal. The subtraction was performed using the ensemble average of accumulated spectra. The end result is 2 separated signals for the C and A primer extensions.

All the samples for testing the small spot technique are purified DNA primer extensions mixed with matrix (figure 5). Image A depicts a sample spot made using the standard spotting technique and without purification with SPC-SBE. The standard technique results in large and heterogeneous crystals. Image B is a sample spot made after purification using SPC-SBE. The sample size decreases as a result of employing only SPC-SBE. SPC-SBE by itself can reduce the sample spot size and reducing the area of a spot. Image C is a sample spot made by dipping a stainless steel pin into a mixture of purified DNA primer extensions and matrix. The coated pin is then quickly touched onto the surface of the sample plate. The amount of downward pressure with the pin onto the stainless steel MS sample plate regulates the resulting size of the sample spot. The sample spots are now close to the size of the MALDI-TOF MS laser diameter. The diameter of the sample spot made with a needle is now significantly smaller than using the standard spotting technique²⁵. This is an improvement over standard spotting techniques, because it eliminates the need for scanning the laser across an entire area to find a “hot spot”, where a strong peak signal can be found on the sample. This “hot spot”

effect is due to the heterogeneity of the sample crystal. Now that the sample size has decreased to the diameter of the laser, the laser can ionize the entire sample without the need for scanning. Reducing the need for scanning the laser across a sample crystal also decreases the amount of time spent in search of a hot spot. This increases the overall efficiency of the analysis by eliminating the amount of time needed for searching the sample spot. The small sample spotting technique was modified from protein nanoarrays made by dip-pen nanolithography²⁵. A robotic arm could make multiple small sample spots. Image D shows the size of a MALDI-TOF MS sample plate. The ruler is 6 inches or 15.24 cm long.

4) *Results of application of small spots and signal separation on PCR products data:*

The calibration curve (figure 6) was generated from varying the concentration ratio between 2 synthetic templates. These synthetic templates were based on a short region of CYP2C9 gene. The concentration ratios of G allele to A allele range from 0.01 to 3. Both synthetic templates have 2 SNP sites. One SNP is homozygous for C and the other is heterozygous for G/A. A linear relationship was found between the concentration ratios and peak area ratios. The linear trend closely correlates with the experimental data as indicated by the R^2 value of 0.9. At least 3 trials were completed to reduce random variability. The peak area ratios for both the calibration curve and PCR products were calculated using the MATLAB code for signal separation.

PCR products were generated from genomic DNA samples extracted from a HepG2 liver cell line; amplified a short region of CYP2C9 exon 2 gene (figure 7). The 2 SNPs were selected based on their inclusion in the R144C allelic variant¹¹. This

particular variant was found to be important to the change in the phenotype of an enzyme to convert therapeutically active warfarin to inactive components of 6-hydroxy and 7-hydroxy metabolites¹⁰. A decrease in warfarin metabolism was catalyzed by the R144C allelic variant. The V_{\max} decreased about 2-3 times lower than the wild-type enzyme. Warfarin is an anticoagulant used to treat and prevent harmful blood clots resulting from for instance, a recent heart attack. The R144C allelic variant has been shown to elicit an Arg¹⁴⁴ to Cys¹⁴⁴ change that occurs in a large proportion of the population. The amount of PCR product was increased by increasing the volume added to the SBE reaction. The resulting primer extensions indicate that the genomic DNA has a C allele, while the synthetic template yields a T allele at the same SNP location. Both templates have a G extension at the second SNP site. The amount of synthetic template remained fixed with each SBE reaction while the volume of PCR product added to the same SBE reaction was increased from 5 μ L to 30 μ L. SPC was performed on each reaction after verification that the SBE reaction produced primer extension products. As seen in figure 7, the peak area ratio of the C extension to the U extension, calculated from the purified DNA samples, increased as more PCR product was added to the reaction mixture. This increase in peak area ratio was expected as more PCR product was added to the reaction mixture, which would produce more C extension products. As a result of adding more PCR product, the concentration ratio of C to U increased linearly from 0.08 to 0.19. The difference between calculating peak area using Matlab in comparison to the standard technique was calculated to determine the improvement in quantitative analysis of biotinylated DNA fragments (table 2).

5) *A microbead pipette tip device:*

Prototypes of a microbead pipette tip device, to increase the throughput of the SPC-SBE approach, were built to test the concept (figure 8). It also provided an opportunity to test the selected membrane filter and epoxy to construct the device. Monomeric avidin-coated beads are confined between a membrane filter and another filter. A 1-200 μ L pipette tip with a built-in filter was used to construct a prototype device to test the selected membrane filter and the adhesive used to attach the membrane filter. The Loctite Extreme Repair adhesive successfully formed an airtight seal between the membrane filter and the plastic pipette tip. Loctite Extreme Repair was chosen based on its ability to maintain its stability and adhesive properties through temperature changes. Monomeric avidin-coated beads need to be stored below 4°C. The membrane filters used in these prototypes are Millipore Express Plus (Millipore Corp. Bedford, MA). Each filter is polyethersulfone, hydrophilic, 13 mm in diameter, 160-190 μ m thickness and has a pore size of 0.45 μ m. These filters are low protein binding and are commonly used to sterile filter tissue culture media, additives, buffers and other aqueous solutions. It was calculated that these filters would be appropriate for use with a pipettor because it has a water flow rate of 23 mL/minute \times cm² at 10 psi. One filter was cut into fourths to simply cover the opening at the end of the tip.

An initial investigation was performed to determine to test a pipette tip as an incubation container for the binding of biotinylated extension products to streptavidin coated microbeads. Two synthetic templates were used to generate 2 different extension products. Each template had a SNP located at the same site but with a different allele. The extension primer used was a 14-mer oligonucleotide. The equal amounts of each

extension product were incubated for 30 minutes with the beads in a pipette tip or a centrifuge tube. As a result of measuring the peak area ratio, it appears as though the binding efficiency of biotinylated extension products to streptavidin coated beads is not hindered by incubation in a pipette tip (figure 9). DNA extension products of different lengths were incubated for 30 minutes with streptavidin coated microbeads in a pipette tip or a centrifuge tube (figure 10). The same synthetic templates were used as in figure 3. Two different primers were used in the SBE reaction. In comparing both plots, the binding of both extension products did not change significantly between using a pipette tip or a centrifuge tube as incubation vessels. With further development, the device will provide a rapid on-line isolation system for biotinylated DNA prior to MALDI-TOF MS.

Conclusions:

The effect of matrix formulation was first investigated on DNA quantization and the appropriate diammonium citrate to 3-hydroxypicolinic acid was found based on the gathered data. Then, we have shown that improvements in sample preparation and data analysis enhanced the DNA quantification analysis system based on SPC-SBE. Creating small sample spots with a stainless steel needle to reduce sample crystal heterogeneity and signal processing to reduce the effect of salt adduct peaks further optimized the DNA quantification system. Additionally, a prototype of a microbead pipette tip device was built to test the selected adhesive and membrane filter to construct it. Preliminary data also suggests that extension products will still be similarly captured in a pipette tip versus a standard centrifuge tube. In summary, we demonstrate here our method to optimize DNA quantification using MALDI-TOF MS and molecular affinity provides high quantification accuracy and therefore can be efficiently used for multiple applications, such as gene expression analysis.

References:

- 1) TJ Griffin, LM Smith. "Single nucleotide polymorphism analysis by MALDI-TOF mass spectrometry." *Mass Spectrometry and Genomic Analysis*. Book edited by JN Housby. 2001. 1-15.
- 2) JA Ragas, TA Simmons, PA Limbach. "A comparative study on methods of optimal sample preparation for the analysis of oligonucleotides by matrix-assisted laser desorption/ionization mass spectrometry." *The Analyst*. 2000. 125: 575-581.
- 3) J Glokler, P Angenendt. "Protein and antibody microarray technology." *Journal of Chromatography B*. 2003. 797: 229-240.
- 4) Z Meng, TA Simmons-Willis, PA Limbach. "The use of mass spectrometry in genomics." *Biomolecular Engineering*. 2004. 21:1-13.
- 5) J Luo, WB Issacs, JM Trent, DJ Duggan. "Looking beyond morphology: cancer gene expression profiling using DNA microarrays." *Cancer Investigation*. 2003. 21: 937-949.
- 6) Ding and Cantor. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *PNAS* (2003) 100: 3059-3064.
- 7) M Hollstein, D Sidransky, B Vogelstein, CC Harris. "*p53* mutations in human cancers." *Science*. 1991. 253:49-53.
- 8) S Kim, JR Edwards, L Deng, W Chung, J Ju. "Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry." *Nucleic Acids Research*. 2002. 30: e85.
- 9) JL Semmlow. *Biosignal and Biomedical Image Processing: MATLAB-based Applications*. Marcel Dekker, Inc. First edition. 2004. 34-36.
- 10) AE Rettie, LC Wienkers, FJ Gonzalez, WF Trager, KR Korzekwa. "Impaired (S)-warfarin metabolism catalyzed by the R144C allelic variant of CYP2C9." *Pharmacogenetics*. 1994. 4:39-42.
- 11) M Ingleman-Sundberg, AK Daly, DW Nebert. "Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee." <http://www.cypalleles.ki.se/cyp2c9.htm>.
- 12) EP Diamandis, TK Christopoulos. "The biotin-(strept)avidin system: principles and applications in biotechnology." *Clinical Chemistry*. 1991. 37: 625-636.
- 13) J Wu, Y Hua, RF Keep, T Schallert, JT Hoff, G Xi. "Oxidative brain injury from extravasated erythrocytes after intracerebral hemorrhage." *Brain Research*. 2002. 953: 45-52.
- 14) J Ju, DH Kim, L Bi, Q Meng, X Bai, Z Li, X Li, MS Marma, S Shi, J Wu, JR Edwards, A Romu, NJ Turro. "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators." *Proceedings of the National Academy of Sciences*. 2006. 103: 19635-19640.
- 15) PJ Finn, L Sun, S Nampalli, H Ziao, JR Nelson, JA Mamone, G Grossman, PK Flick, CW Fuller, S Kumar. "Synthesis and application of charge-modified dye-labeled dideoxynucleoside-5'-triphosphates to 'direct-load' DNA sequencing." *Nucleic Acids Research*. 2002. 30: 2877-2885.

- 16) H Fakhrai-Rad, N Pourmand, M Ronaghi. "Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms." *Human Mutation*. 2002. 19: 479-485.
- 17) CM Roth, "Quantifying gene expression." *Current Issues in Molecular Biology*. 2002. 4: 93-100.
- 18) H Liu, H Wang, Z Shi, H Wang, C Yang, S Silke, W Tan, Z Lu. "TaqMan probe array for quantitative detection of DNA targets." *Nucleic Acids Research*. 2006. 34: e4.
- 19) DG Ginzinger. "Gene quantification using real-time quantitative PCR: an emerging technology hits mainstream." *Experimental Hematology*. 2002. 30: 503-512.
- 20) J Tost, P Schatz, M Schuster, K Berlin, IG Gut. "Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry." *Nucleic Acids Research*. 2003. 31: e50.
- 21) S Sauer, IG Gut. "Extension of the GOOD assay for genotyping single nucleotide polymorphisms by matrix-assisted laser desorption/ionization mass spectrometry." *Rapid Communications in Mass Spectrometry*. 2003. 17: 1265-1272.
- 22) C Jurinke, D van den Boom, CR Cantor, H Koster. "The use of MassARRAY technology for high throughput genotyping." *Advances in Biochemical Engineering/Biotechnology*. 2002. 77: 57-74.
- 23) Koomen JM, WK Russell, SE Tichy, DH Russell. "Accurate mass measurement of DNA oligonucleotide ions using high-resolution time-of-flight mass spectrometry." *Journal of Mass Spectrometry*. 2002. 37:357-371.
- 24) M Karas, U Bahr. *Mass spectrometry in biomolecular sciences*. 1996. 33-39.
- 25) K Lee, SJ Park, CA Mirkin, JC Smith, M Mrksich. "Protein nanoarrays generated by dip-pen nanolithography." *Science*. 2002. 295: 1702-1705.
- 26) G Corona, G Toffoli. "High throughput screening of genetic polymorphisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry." *Combinatorial Chemistry and high throughput screening*. 2004. 7: 707-725.
- 27) R Bakhtiar, FLS Tse. "Biological mass spectrometry: a primer." *Mutagenesis*. 2000. 15: 415-430.
- 28) P Ross, L Hall, I Smirnov, L Haff. "High level multiplex genotyping by MALDI-TOF mass spectrometry." *Nature Biotechnology*. 1998. 16: 1347-1351.
- 29) J Tost, IG Gut. "Genotyping single nucleotide polymorphisms by mass spectrometry." *Mass Spectrometry Reviews*. 2002. 21: 388-418.
- 30) S Kim, ME Ulz, T Nguyen, C-M Li, T Sato, B Tycko, J Ju. "Thirtyfold multiplex genotyping of the *p53* gene using solid phase capturable dideoxynucleotides and mass spectrometry." *Genomics*. 2004. 83: 924-931.
- 31) S Kim, HD Ruparel, TC Gilliam, J Ju. "Digital genotyping using molecular affinity and mass spectrometry." *Nature Reviews Genetics*. 2003. 4: 1001-1008.
- 32) A Misra, Hong, S Kim. *Clinical Chemistry*. 2007. In Press.

	DNA sequence, 5' to 3'	Extension product mass, Da
Synthetic system:		
Template 1	TAAAGCTATAGGTACAGTATTAGTAGGACCTACACC-TGTCAAGATAATGGTCCAGGTCGGT	
Template 2	TAAAGCTATAGGTACAGTATTAGTAGGACCTACACC-TGTCAATATAATGGTCCAGGTCGG	
SBE primer	ACCGACCTGGACCATTAT	6108.6, 6131.6
PCR products:		
PCR primer	GTA ACTTCGTTTGCTGTTATCTCT	
PCR primer	TCCATATCACTGACCTTACTGGA	
Template 1	TCCCTCCACAAGGCAGCGGGCT-TCCTCTTGAACACAGTCCTCAATGCTCCTCTTCCCCATC	
Template 2	TCCTCCACAAGGCAGCGGGCTT-CCTCTTGAACACGGTCCTCAATGCTCCTCTTCCCCATC	
SBE primer 1	AGAGGAGCATTGAGGAC	5978 (C), 6063 (U)
SBE primer 2	GTGTTCAAGAGGAAGCCC	6264 (G)

Table 1 DNA sequences of the templates and primers utilized in experiments (Midland Certified Reagent Company Midland, TX).

Primer+C : Primer+U			
	Standard	With Matlab	
Concentration Ratio	Peak Area Ratio	Peak Area Ratio	Difference
0.01	0.22	0.16	0.06
0.0125	0.28	0.20	0.08
0.02	0.32	0.21	0.11
0.05	0.47	0.35	0.12
0.1	0.44	0.27	0.17
0.2	0.91	0.86	0.05

Table 2 This table shows the difference between the peak area ratio using the standard method in comparison to using signal processing with Matlab.

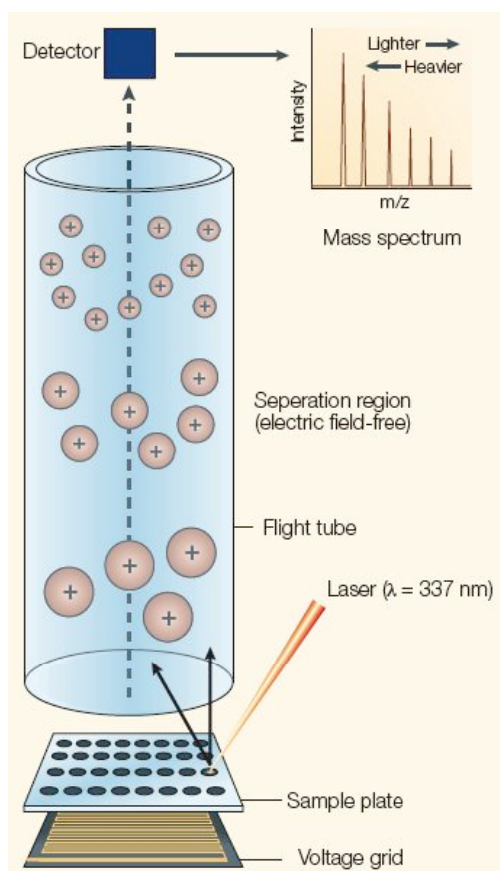


Figure 1 A schematic of a MALDI-TOF MS³¹.

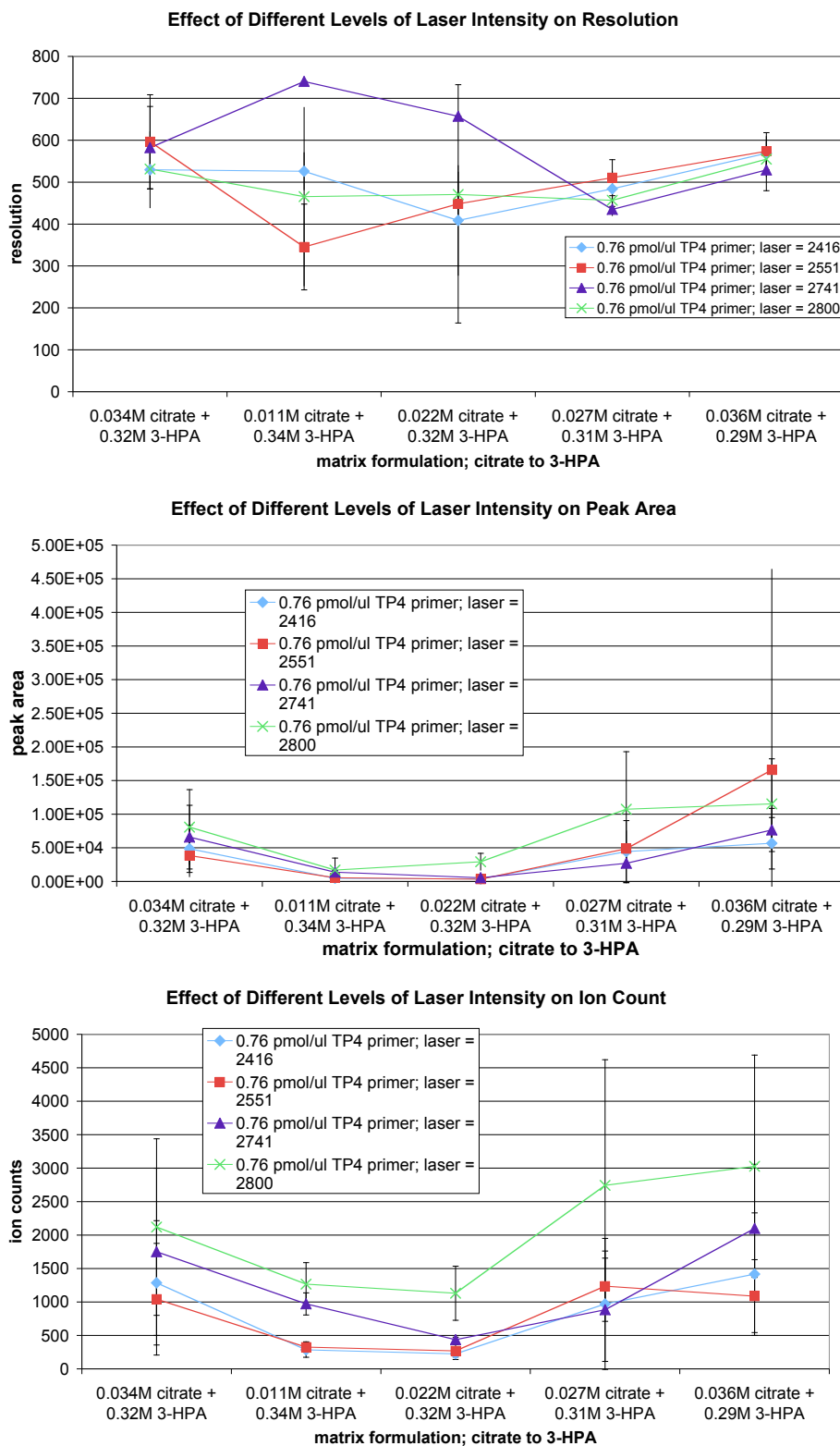


Figure 2 The effect of matrix composition on the peak signal quality.

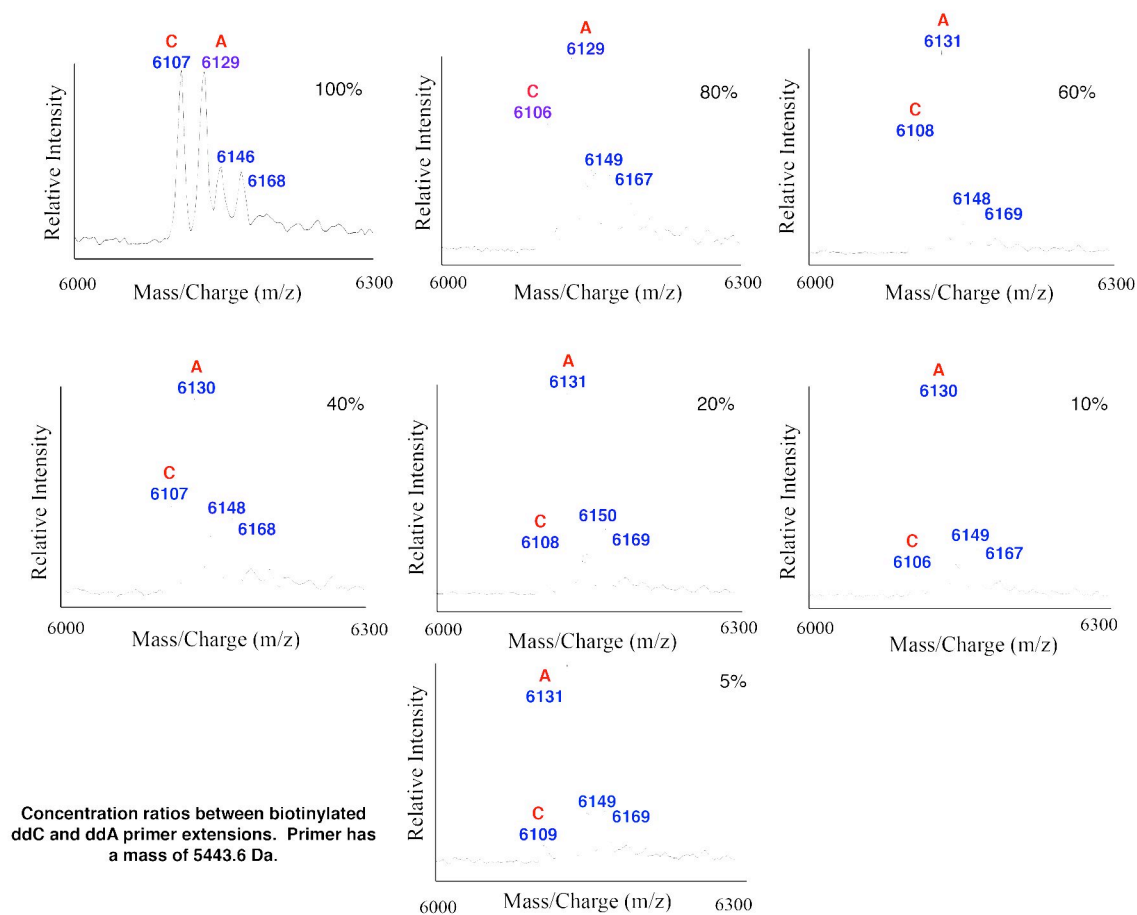


Figure 3 Another calibration curve using a different system of synthetic templates and primer.

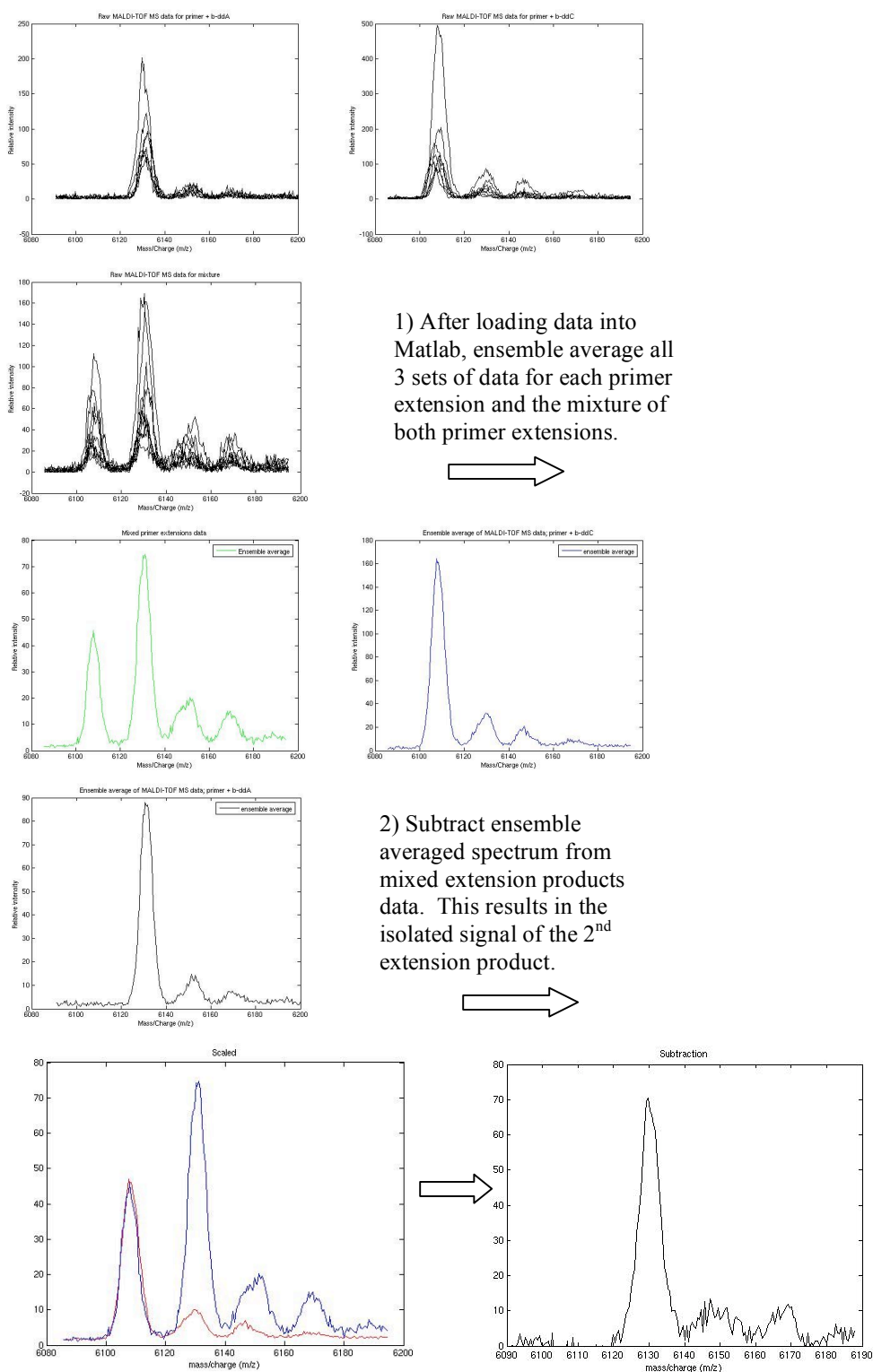


Figure 4 MATLAB 6.0 graphs illustrate the ensemble-averaged signals that are being subtracted to obtain a more accurate representation of each signal.

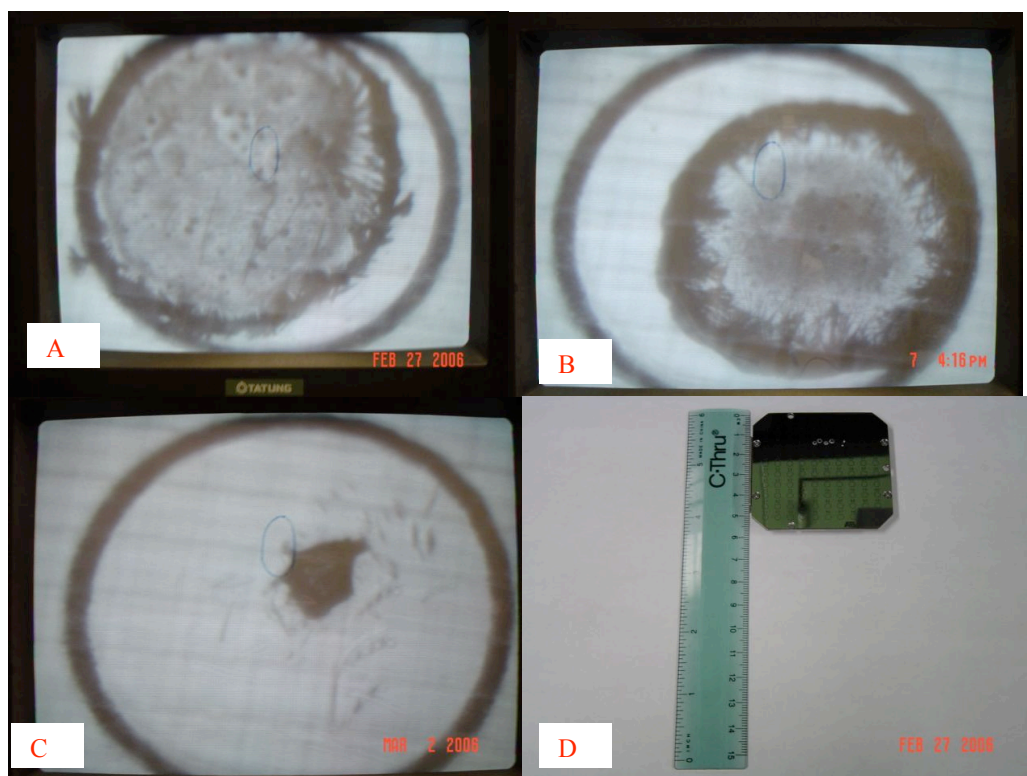


Figure 5 Images were taken to demonstrate the size differences when using the improved DNA quantification system with small sample spots.

CYP2C9 Calibration Curve

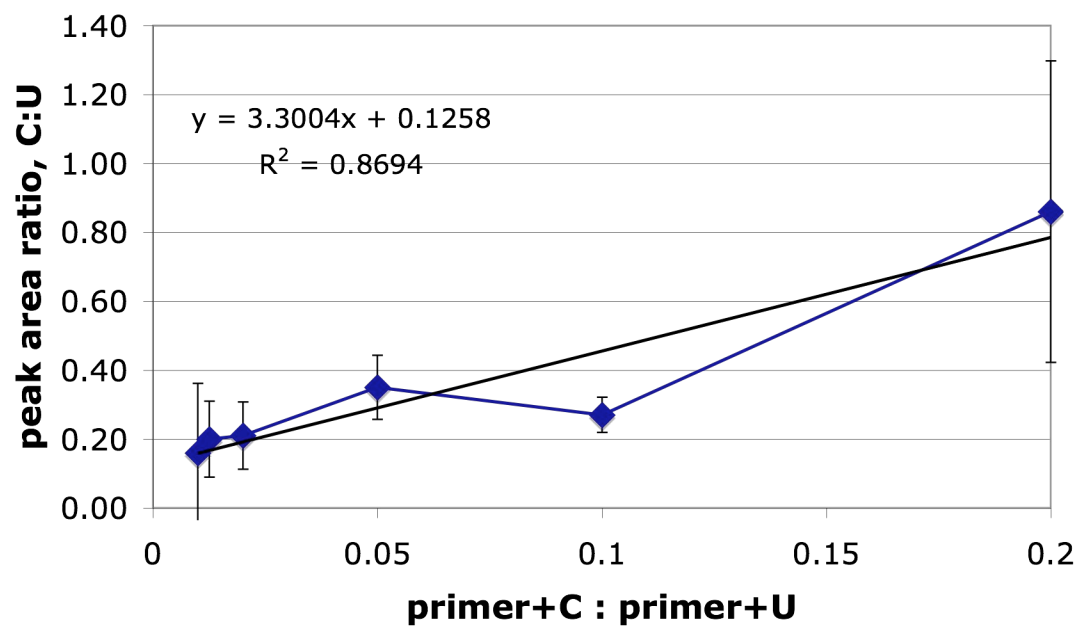


Figure 6 Calibration curve generated from 2 synthetic templates based on a short region of the CYP2C9 gene.

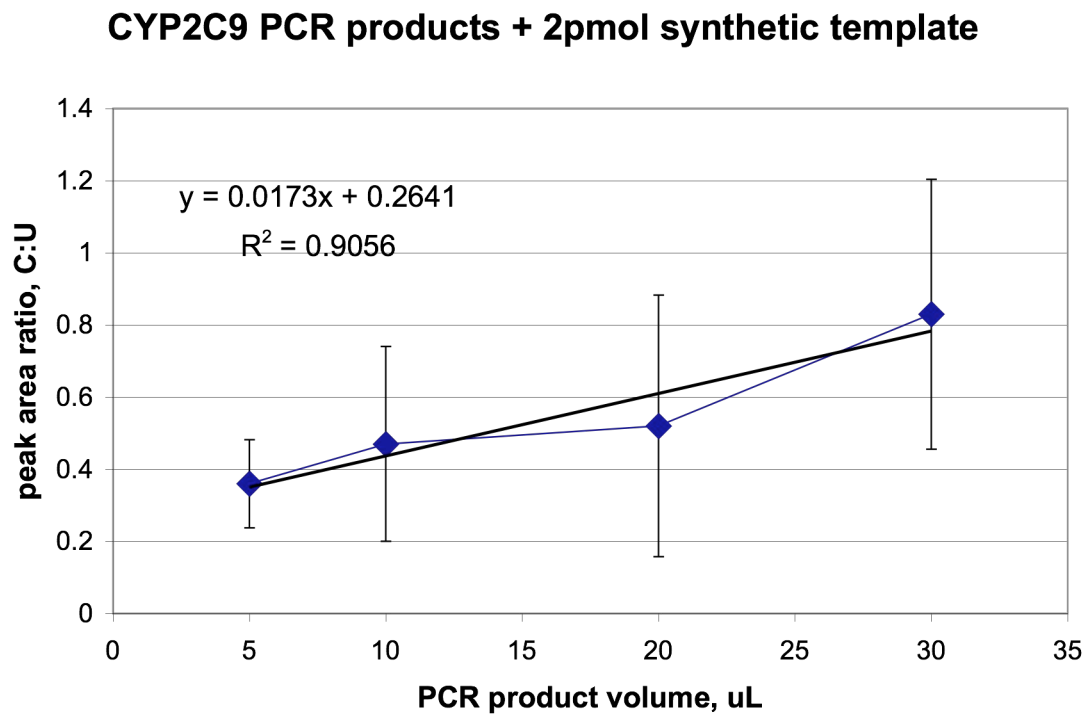


Figure 7 Quantification of PCR products generated from genomic DNA of a HepG2 cell line. A synthetic template of 2 pmol was used as an internal standard.

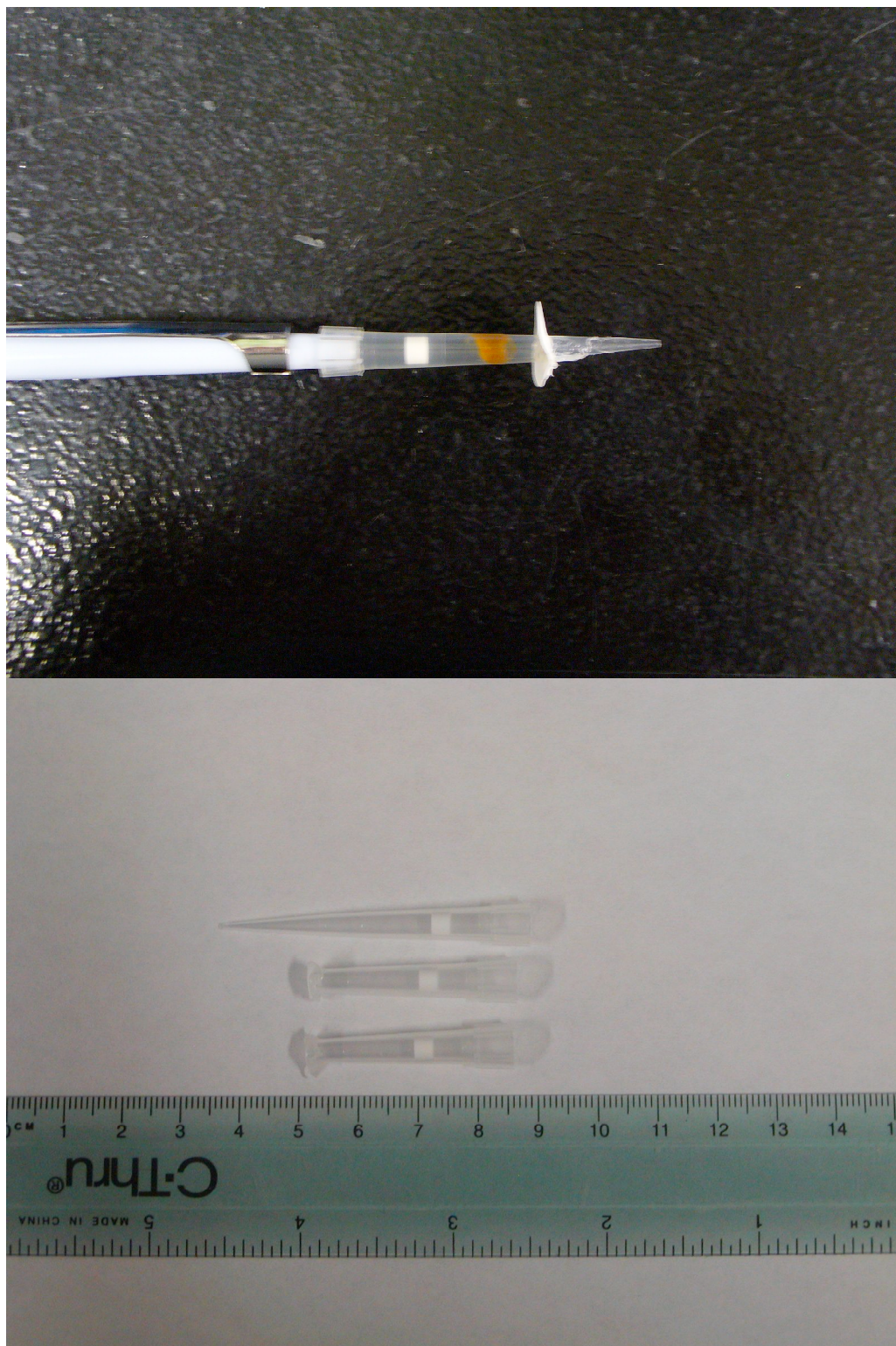


Figure 8 Prototype of a microbead pipette tip device.

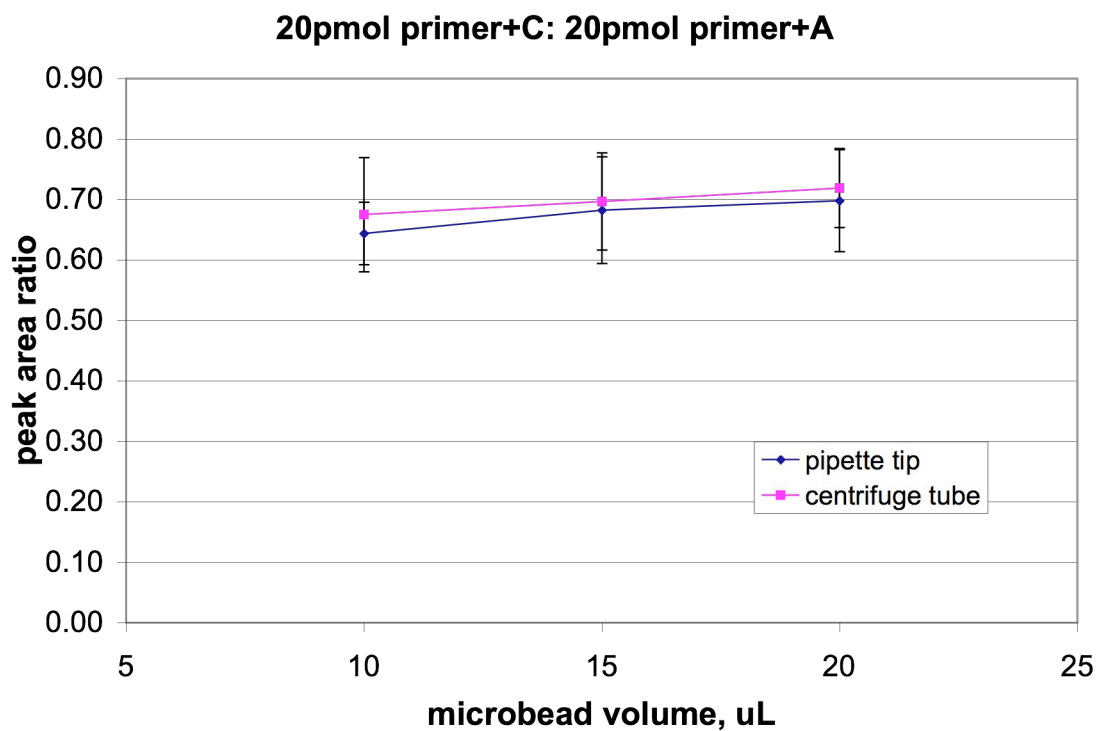


Figure 9 Initial data to compare the incubation and binding step of biotinylated extension primers to streptavidin coated microbeads in a pipette tip to the conventional method using a centrifuge tube.

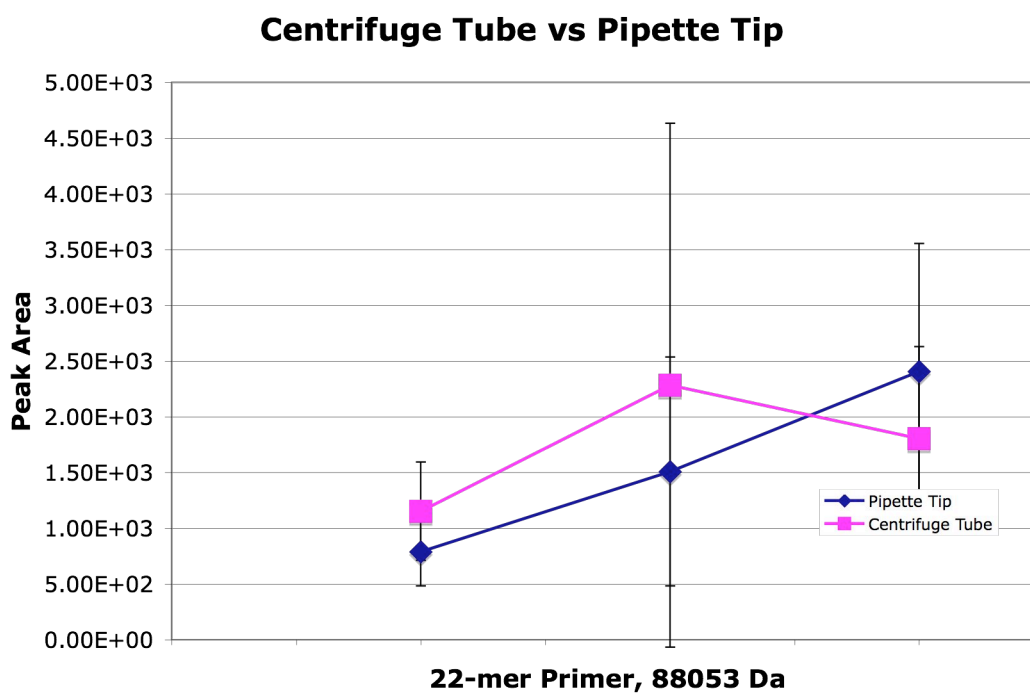
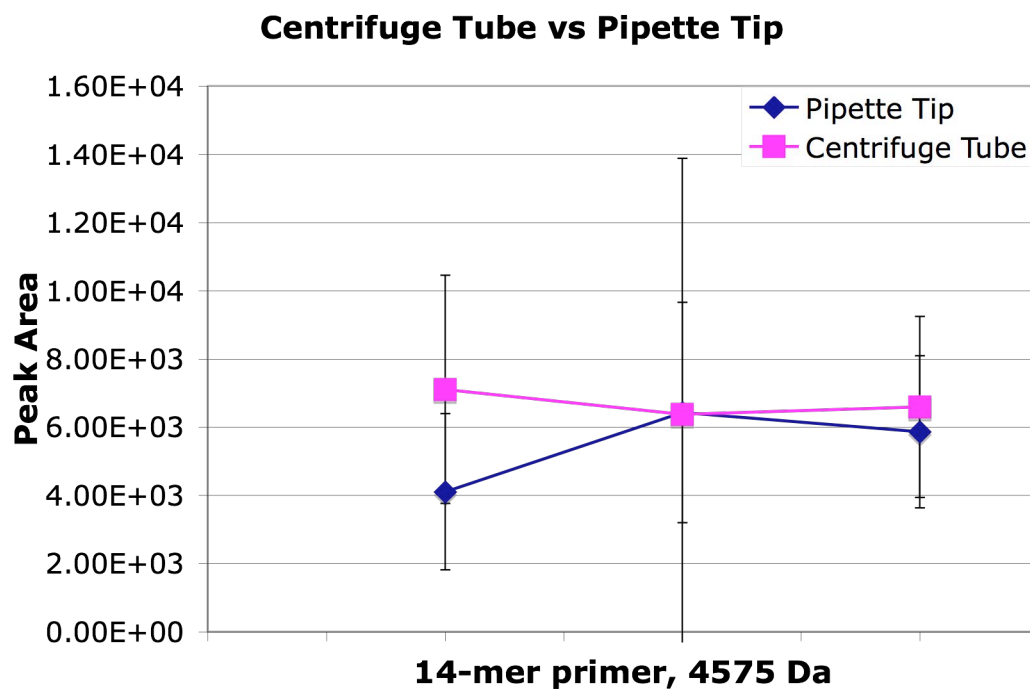


Figure 10 Different lengths of DNA primer extension products were also tested in a centrifuge tube versus a pipette tip.

Matlab 7.0 Program Code:

This is the program code to ensemble average and calculate peak area ratio for MALDI-TOF MS spectra. Specific sections are labeled in bold text for further clarification.

```

clear all; close all;
% Load data (Load data into Matlab as text files.)
a1 = textread('2C9_PCR_30_0001.txt');
a2 = textread('2C9_PCR_30_0002.txt');
a3 = textread('2C9_PCR_30_0003.txt');
a4 = textread('2C9_PCR_30B_0001.txt');
a5 = textread('2C9_PCR_30B_0002.txt');
a6 = textread('2C9_PCR_30B_0003.txt');
a7 = textread('2C9_PCR_30C_0001.txt');
a8 = textread('2C9_PCR_30C_0002.txt');
a9 = textread('2C9_PCR_30C_0003.txt');
A = [a1(:,2),a2(:,2),a3(:,2),a4(:,2),a5(:,2),a6(:,2),a7(:,2),a8(:,2),a9(:,2)];
massa = (a1(9700:10800,1))';
xa = [A(9700:10800,:)];
Xa = xa';
figure;
plot(massa,Xa,'k');
xlabel('Mass/Charge (m/z)');
ylabel('Relative intensity');
title('Raw MALDI-TOF MS data');

b1 = textread('2C9_PCR_0006.txt');
b2 = textread('2C9_PCR_0007.txt');
b3 = textread('2C9_PCR_0009.txt');
B = [b1(:,2),b2(:,2),b3(:,2)];
mass = (b1(9700:10800,1))';
x = [B(9700:10800,:)];
X = x';

figure;
plot(mass,X,'k');
xlabel('Mass/Charge (m/z)');
ylabel('Relative intensity');
title('Raw MALDI-TOF MS data');
%
(Calculate the ensemble average for each set of data.)
% %Ensemble average
avg_a = mean(Xa);
avg_c = mean(X);
%
% %Ensemble standard deviation
% stdev_a = std(Xa);
% stdev_c = std(X);

```

```
%
figure;
plot(massa,avg_a,'k');
% hold on;
% plot(massa,stdev_a,'k. ');
xlabel('Mass/Charge (m/z)');
ylabel('Relative intensity');
title('Ensemble average 2C9');
% legend('ensemble average','ensemble standard deviation');
%
figure;
plot(massa,avg_c, 'b');
% hold on;
% plot(massa,stdev_c,'b. ');
xlabel('Mass/Charge (m/z)');
ylabel('Relative intensity');
title('Ensemble average of MALDI-TOF MS data');
% % legend('ensemble average','ensemble standard deviation');
```

(Align one set of data to the other set to get ready for subtraction.)

```
% %Scaling
scaled_vector = 2.*avg_c + 9;
```

```
figure;
plot(massa,scaled_vector,'r');
hold on;
plot(massa, avg_a,'b');
xlabel('mass/charge (m/z)');
title('Scaled');
```

```
subtract = avg_a - scaled_vector;
```

```
figure;
plot(massa,subtract,'m');
title('Subtraction')
xlabel('mass/charge (m/z)');
```

(Calculate peak area.)

```
%Peak area calculation:
subtract_sum1 = sum(scaled_vector(100:350))-20
subtract_sum2 = sum(subtract(300:600))
Peak_area_ratio = subtract_sum1./subtract_sum2
```

Baseline adjustment (separate file):

```
%%curve for baseline is subtract_new from calling file
```

```

% line_vect=subtract_new;
% line_vect=scaled_vector(100:350);
line_vect=subtract(300:600);
thresh= 100;
base_vals=find(line_vect<thresh);
mask=zeros(length(line_vect),1);
mask(base_vals)=1;

figure; plot(thresh*mask); hold on; plot(line_vect,'r');
on_pts=find(diff(mask)<0);
off_pts=1+find(diff(mask)>0);

%%%m%%% for verification
hold on;
figure;
for i=1:length(on_pts)
plot(on_pts(i),thresh,'-kx','markersize',15,'linewidth',3);
plot(off_pts(i),thresh,'-kx','markersize',15,'linewidth',3);
end

for i=1:length(on_pts)
avg_before(i) = mean(line_vect(on_pts(i)-10:on_pts(i)));
end

for i=1:length(off_pts)
avg_after(i) = mean(line_vect(off_pts(i):off_pts(i)+10));
end

for i=1:length(avg_before)
baseline_mean(i)=mean([avg_before(i), avg_after(i)])
end

subtract_adjust = line_vect - baseline_mean;
figure; plot(subtract_adjust, 'b'); title('Baseline adjusted');

subtract_adjust_sum = sum(subtract_adjust)
% ratio = scaled_sum/subtract_adjust_sum;

```