2008

Technologies for Proteomic and Genomic Biomarker Analysis

Yiding Liu
Cleveland State University

TECHNOLOGIES FOR PROTEOMIC AND GENOMIC BIOMARKER ANALYSIS

YIDING LIU

BACHELOR OF SCIENCE IN CHEMISTRY

ZHENGZHOU UNIVERSITY

ZHENZHOU, HENAN PROVINCE, P.R.CHINA

SEPTEMBER 1999

submitted in partial fulfillment of requirements for the degree

DOCTOR OF PHILOSOPHY IN CLINICAL-BIOANALYTICAL CHEMISTRY

at the

CLEVELAND STATE UNIVERSITY

DECEMBER 2008

This dissertation has been approved

for the Department of CHEMISTRY

and the College of Graduate Studies by

_____

Dissertation Chairperson, Baochuan Guo
Department of CHEMISTRY

_____

Date

_____

Lily M Ng
Department of CHEMISTRY

_____

Date

_____

Crystal M. Weyman
Department of BIOLOGICAL, GEOLOGICAL, and ENVIRONMENTAL SCIENCES

_____

Date

_____

Aimin Zhou
Department of CHEMISTRY

_____

Date

_____

Xiang Zhou
Department of CHEMISTRY

_____

Date

TECHNOLOGIES FOR PROTEOMIC AND GENOMIC BIOMARKER ANALYSIS

YIDING LIU

ABSTRACT

In the first part of this dissertation, we systematically validated the application of molecular weight cut-off ultrafiltration in separation and enrichment of low-molecular-weight peptides from human serum. Under optimized conditions, both free-phase and bound LMW peptides could be separated and enriched. The method proved to be highly efficient and reproducible coupled with MALDI-TOF MS proteomic pattern analysis. Three marker peaks were found to be eligible for distinguishing normal and ovarian cancer samples.

A novel organic solvent precipitation method coupled with enzymatic deglycosylation was also developed for biomarker detection from human serum. This method allowed us to generate reproducible free-phase peptide patterns comparing with the ultrafiltration method. A potential marker was found up-regulated in benign and ovarian cancer patients. It was further identified as des-alanine-fibrinopeptide A using LC tandem mass spectrometry.

In the second part of this dissertation, a new sample preparation procedure was developed to improve the MALDI-TOF analysis of low-concentration oligonucleotides. The oligonucleotide solutions are first dispensed and allow shrinking onto a small spot on an anchoring target. A small volume (0.1uL) of saturated 3-HPA matrix solution is then

added on top of each dried oligonucleotide spot. Samples prepared by this procedure are homogenous and reduces the need to search for ‘sweet’ spots. The increased shot-to-shot and sample-to-sample reproducibility makes it useful for high-throughput quantitative analysis. This procedure allowed robust detection of oligonucleotides at 0.01μM level and mini-sequencing products produced using only 50 fmol of extension primer.

And a strategy called probe-clamping-primer-extension-PCR (PCPE-PCR) was developed to detect MRS alterations in a large background of wild-type DNA. PCR errors often generate false positive mutant alleles. In PCPE-PCR, mutant single-strand DNA molecules are preferentially produced and enriched. Thereafter, the remaining single-strand mutant DNA molecules were amplified using PCR and analyzed to reveal the MSI-H status. We showed that PCPE-PCR could detect both mutated BAT26 and TGF-h-RII(A)10 markers in the presence of 500-fold excess of normal DNA and that as few as three copies of mutated DNA could be detected. We also showed that this technology could detect MSI-H colorectal cancer by fecal DNA analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART I

TECHNOLOGY DEVELOPMENT FOR PROTEOMIC BIOMARKER ANALYSIS

CHAPTER I

GENERAL INTRODUCTION

1.1 Proteomics Basics

After the human genome sequence was determined in 2000, molecular medicine research moved further beyond genomics to proteomics.

What is proteomics? Although the exact definition of proteomics evolves along with research progress [Wasinger et al., 1995; Loo et al., 1996; Wilkins et al., 1996; Hochstrasser, 1998], it is widely accepted now that proteomics is "the study of all the protein forms expressed within an organism as a function of time, age, state, external factors, etc." [Reynolds, 2002]. In other words, the types of expressed proteins, their abundance, state of modification, sub-cellular location, etc. are dependent on the physiological state of the cell or tissue. Therefore, proteomic analysis can be considered as an assay surveying throughout the whole genome to

differentiate and study cellular states and to determine the molecular mechanisms behind them [Haynes et al., 1998].

Why proteomics? After the Human Genome Project was completed, researchers noted that the genomic information itself can not accurately predict what is actually occurring at the protein level within a certain cell type. The expression of genes in a specific cell is affected by numerous factors as a result of tissue complexity, inter- or intracellular signaling and responding to changes in their environment. Since proteins are the working units to perform functions of a cell, and since their expression, structure, and activity can be changed any time through post translational modifications, proteomics, a new paradigm comparable to genomics, is therefore needed. Proteomics provided a powerful approach for differentiating normal and transformed cell formations, identification of molecular targets from complex mixtures which are specific to disease, evaluating protein-based drug targets with novel endpoints for better treatment of various malignant diseases [Verma et al., 2001].

1.2 Cancer Detection and Molecular Biomarkers

Cancer remains to be a major challenge to public health despite progress in detection and therapy. A large portion of the US population will develop cancer during their lifetime [Chaurand et al., 2001], with ~50,000 individuals dying annually from the disease [American Cancer Society, 1996]. With the better understanding of molecular basis of malignant diseases, cancer diagnosis and treatment has become

more target-oriented but still limited in the late invasive and metastatic stages. This in turn greatly limited the success of treatment. It is generally accepted that if molecular targets in the earliest stage of tumor growth can be identified, the treatment of cancer is likely to be more effective and successful [Negm et al., 2002].

Molecular markers are defined as detectable species that are characteristic in case of an organism's state of health or disease, or are responsive to a particular clinical treatment [Negm et al., 2002]. It is known that changes in molecular levels occur at the beginning of the tumor growth. To detect these changes, diagnostic assays using molecular biomarkers have considerable potentials for early cancer detection.

Proteomics, especially serum proteomics, is unquestionably valuable in the discovery of biomarkers. Proteomes could reflect both the genetic coding of the cell and the impact of their surrounding microenvironment. Due to altered expression, posttranslational modification and other intra- or inter-cellular events, distinguishable changes will occur at the protein level during the transformation of healthy cells into cancer cells [Banks et al., 2000]. Therefore, the ability to make accurate measurement about the state of pathology (or health) within organ systems by analyzing the fluid that perfuses them, especially as in serum/plasma, has several major advantages. Firstly, for a biomarker (or set of biomarkers) to be clinically valuable, it must be disease-specific and be accurately detected. To some extent, components found in the serum represent the cumulative effect of disease processes in organ tissue, and some of those were not recognized by previous clinical practices. This makes serum a valuable source of molecular biomarkers with high expected specificity. And

moreover, because traces of the molecular footprints of on-going disease process are expected to equilibrate in the circulation system, the strong possibility of sampling error that is often associated with tissue biopsy is reduced substantially. Secondly, to be applicable for clinical uses, any source of biomarkers must be readily obtainable from the patient. Serum or plasma is one of the best sources for biomarker discovery and screening because phlebotomy can be repeated as often as needed. Therefore, serum investigations provide relative ease of sampling compared with the biopsy of solid organs.

There are several clinical cases where molecular markers from serum are already being used. For example, increased serum levels of prostate-specific antigen (PSA) and cancer antigen-125 (CA-125) are routinely used for the detection of cancer in the prostate and ovary, respectively [Grossklaus et al., 2002; Guppy et al., 2002]. However, there are no molecular markers clinically available for most types of cancer in terms of early stage diagnosis, especially for several most deadly cancers such as lung cancer and breast cancer.

1.3 Technologies Involved in Cancer Proteomic Research:

1.3.1 Two-Dimensional Gel Electrophoresis

Two-dimensional gel electrophoresis (2D-GE) has been the mainstay of electrophoretic technology for a decade and is one of the most widely used tools for separating proteins. In 2D-GE, proteins are first separated by isoelectric focusing (IEF) based on the charge state of proteins and then further resolved by sodium dodecyl

sulfate polyacrylamide gel electrophoresis (SDS–PAGE) in a second, perpendicular dimension according to their size. Separated proteins can then be visualized by all kinds of staining methods (Coomassie Blue and Silver stain, for example), or by autoradiography, producing a two-dimensional image that contains thousands of proteins [Klose et al., 1995]. The identification of proteins in each stained spot could be achieved using immunoblotting, N-terminal sequencing or tandem mass spectrometry [Hunter et al., 2002]. In many applications, 2D-GE was used to evaluate cell lysates, tissue, or biofluid protein extracts [Jiang et al., 2003; An et al., 2004; Vitorino et al., 2004]. The commercialized use of narrow, immobilized PH gradients for the first dimension can increase resolving power and help detect low-abundance proteins [Görg et al., 2000]. Differences between the samples can be compared and relative quantities determined by quantifying the spot intensity ratios in different 2D gels. Matrix-assisted desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and high resolution LC tandem mass spectrometry allow the identification and quantification of very small amounts of protein isolated from the gel. During last several years, increasing numbers of researches in cancer biomarker screening and discovery were reported using 2D differential gel electrophoresis (2D-DIGE, see **Figure 1.1**) coupled with mass-spectrometry [Arthur et al., 2002; Schulenberg et al., 2003; Friedman et al., 2004].

However, in spite of all these achievements made in the past decade, 2D-GE still suffers from several major limitations. These limitations includes: requirement of relatively large sample quantities (the limit for the detection of proteins with silver

staining is approximately 1 ng, i.e., 20 fmol for a 50 kDa protein [Shevchenko et al., 1996], difficulties in detecting proteins with extreme molecular mass (large or small) and isoelectric point, limited dynamic range according to staining method used, poor reproducibility, labor intensive and difficulties in automating [Beranova-Giorgianni, 2003].

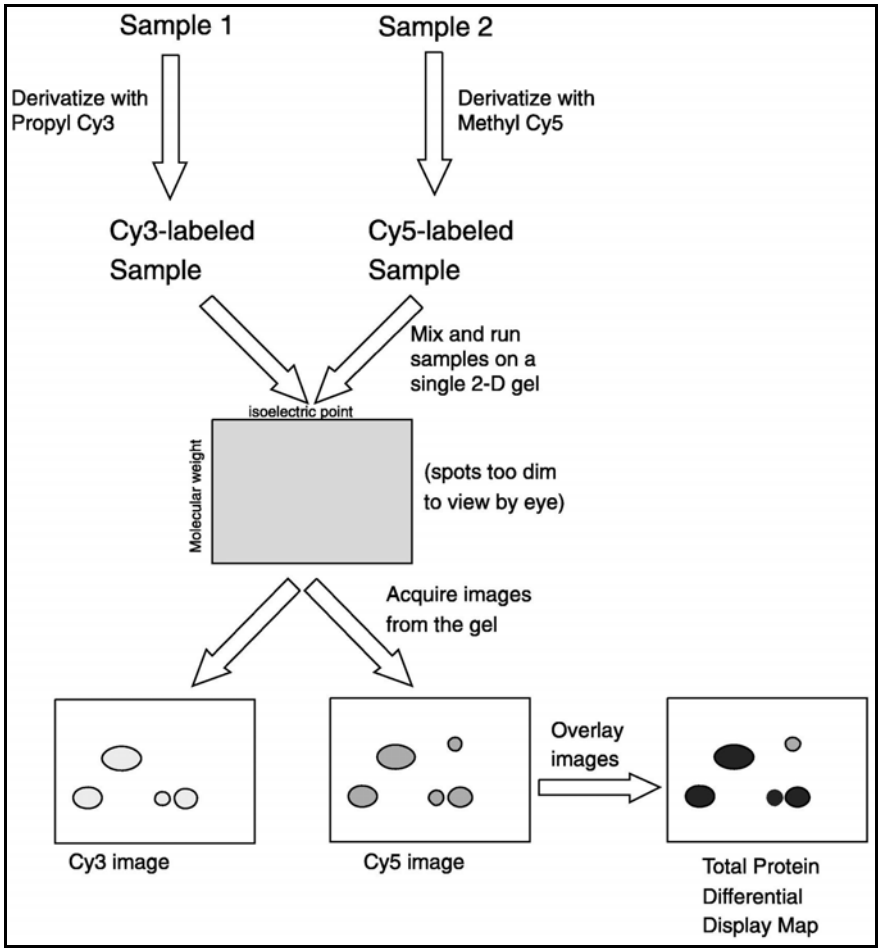


Figure 1.1 Schematic illustration of the difference gel electrophoresis (DIGE) technology platform ["Detection technologies in proteome analysis", Wayne F. Patton, 2002]. Two different samples are derivatized with two different fluorophores, combined and then run on a single 2D gel. Proteins are detected using a dual laser scanning device or xenon arc-based instrument equipped with different excitation /emission filters in order to generate two separate images. Differences in protein expression are identified by evaluation of a pseudo-colored image and data spreadsheet. DIGE technology can maximally evaluate three different samples using Cy2-, Cy3- and Cy5-based chemistries.

1.3.2 Multidimensional Protein Identification Technology

Multidimensional protein identification technology (MudPIT) is a method that couples a microscale two-dimensional chromatography system to a tandem mass spectrometer, which allows high-throughput analysis of a complex proteome (see **Figure 1.2**). First introduced by A. J. Link and colleagues, strong cation exchange (SCX) resin is placed upstream of the RP portion of the HPLC column [Link et al., 1999], forming a in-line 2D HPLC separation system.

Peptides are bound to the SCX phase first under acidic/low-salt environment and subsequently eluted off using multiple high salt pulses with higher PH buffers, followed by reversed phase separation of each subset of peptides. For analysis of even more complex protein mixtures, such as whole cell extracts and serum/plasma, greater separation and data acquisition capacity is required. Further optimizations in combinations of various HPLC methods, buffers, and even enhancement of general strategy have resulted in a much more advanced MudPIT [Washburn et al., 2001; Wolters et al., 2001]. All kinds of multidimensional chromatography coupled with tandem mass spectrometry have been developed recently. With the high efficiency micro capillary LC-MS or even nanoscale LC-MS, a lot of achievements have been reported in cell or serum/plasma proteome research [Chen et al., 2002; Wagner et al., 2003; Chen et al., 2004]. Hundreds of proteins, including ~pg/ml low abundance (angiotensinogen precursor, cytokins, interleukin) proteins, can be detected and identified [Tirumalai et al., 2003]. Although mixed-bed packing in one HPLC column was a patented setup for the original MudPIT technology, the idea of

multi-dimensional separation coupled with MS/MS detection was well adapted for analyzing highly complicated proteomes. This strategy has played a very important role on proteomic biomarker discovery.

1.3.3 Isotope-coded Affinity Tag

The prototypical method to generate quantitative protein profiles based on stable isotope affinity tagging and MS is the isotope-coded affinity tag (ICAT) reagent method [Gygi et al., 1999]. The first generation reagents (see **Figure 1.3**) consist of a cysteine-reactive group, a linker that contains either heavy or light isotopes, and a biotin affinity tag.

In ICAT reagent method (see **Figure 1.4**), two protein extracts representing different cell states (i.e. healthy and diseased) are treated with the isotopically light or heavy ICAT reagents, respectively. In this labeling step, ICAT reagents are covalently attached with cysteine residues. The labeled protein mixtures are then combined and digested using enzyme of choices (usually by trypsin). ICAT labeled peptides which contain cysteine residue(s) could be selectively isolated using affinity chromatography and analyzed by mass spectrometry. The relative abundance is determined by the ratio of signal intensities of the tagged peptide pairs. The MS/MS spectra are recorded and searched against protein sequence databases to identify the protein. Therefore, in a single operation, the relative abundance and sequence of a peptide can be determined [Tao et al., 2003].

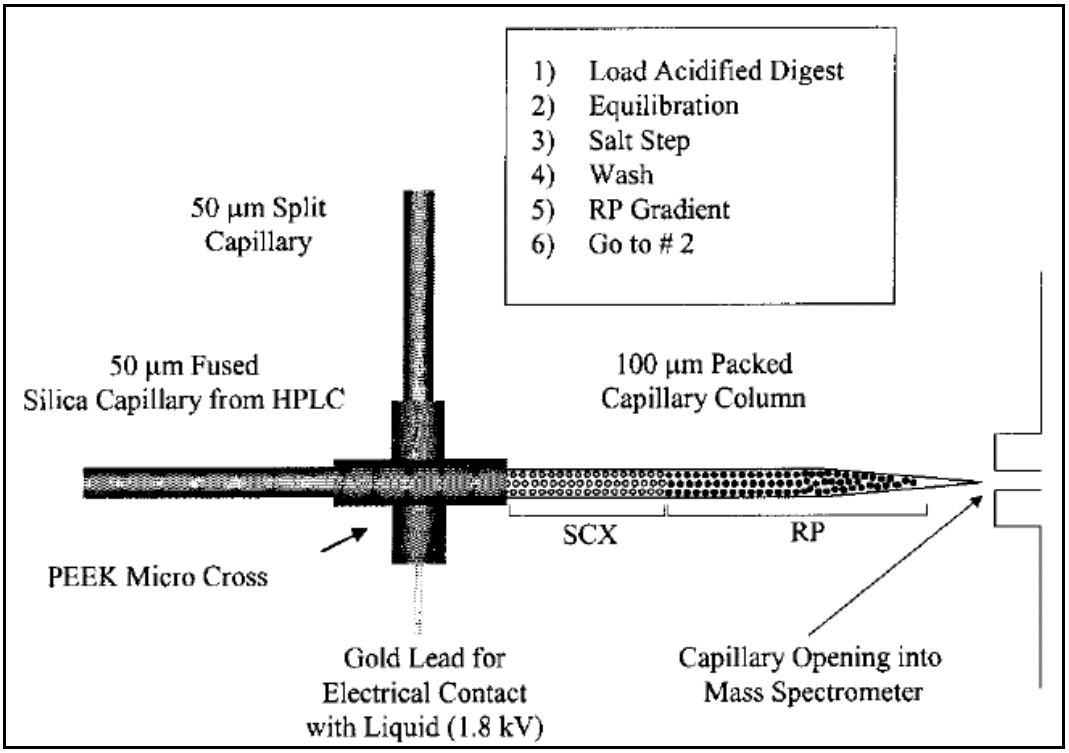


Figure 1.2 The online MudPIT interface developed by Yates and colleagues ["Direct analysis of protein complexes using mass spectrometry", A. J. Link et al., 1999]. It includes biphasic microcapillary column, packed with strong cation-exchange (SCX) and reversed-phase (RP) packing material connected to a microcross. The microcross splits the flow from the HPLC column to 0.15-0.25 uL/min and serves as the connection for the electrospray voltage.

Cysteine residues represent a good target to reduce the complexity of digests from large proteomes or other complex protein samples. This is because 96.1% of all human proteins are known to have cysteine residue(s) but only 26.6% of the peptides contain cysteine after tryptic digestion [Zhang et al., 2003]. Furthermore, there are other chemical methods available to target the SH group on the side chain of reduced cysteine [Zhou et al., 2002]. Therefore, cysteine becomes one of the most useful targets for residue-specific enrichment. And the isotope-coded affinity tag (ICAT) method has become widely used for quantitative proteome profiling.

The second generation of ICAT reagent contains an acid-cleavable linker within the ICAT reagent molecule. This allows removal of the biotin affinity tag before MS and MS/MS analysis. The improved MS/MS performance will significantly increases the number of proteins identified in a single experiment and quantify them with higher confidence scores. Incorporation of ^{13}C rather than deuterium into the ICAT heavy reagent molecule overcomes isotope-co-elution effect in reversed phase chromatography, thereby increases the accuracy of quantification measurements using LC-MS. Recently, Kubota et al. reported proteome analysis of secreted proteins during osteoclast differentiation using ICAT, together with 2D-GE and 2D LC-MS [Kubota et al., 2003]. Meehan et al. reported quantitative profiling of LNCaP prostate cancer cells using ICAT and mass spectrometry [Meehan et al., 2004].

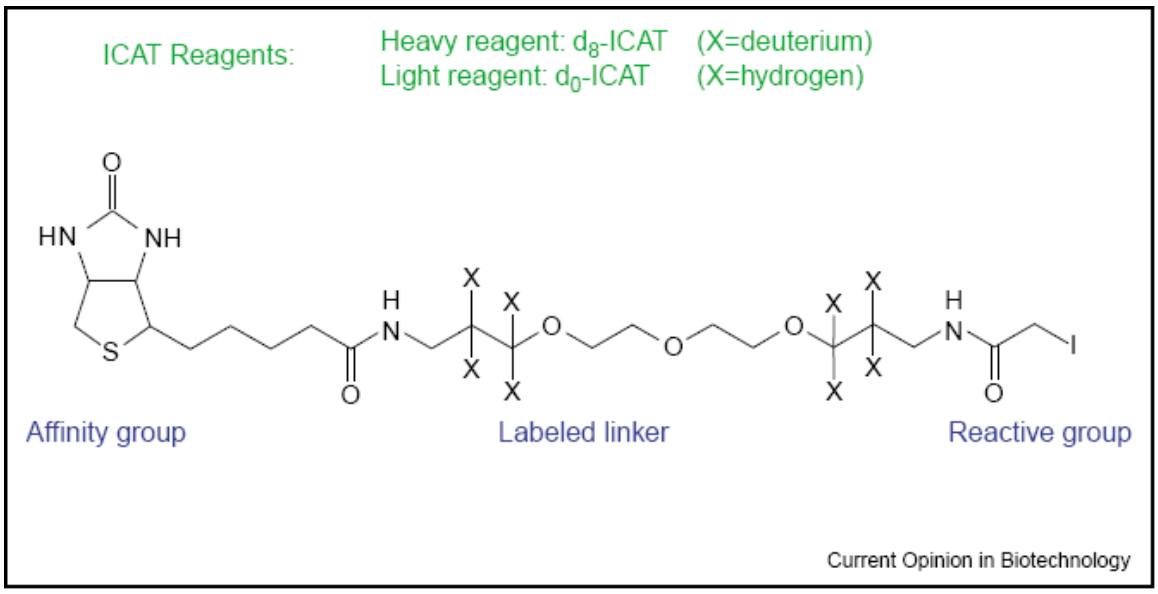


Figure 1.3 The structure of ICAT reagents ["Advances in quantitative proteomics via stable isotope tagging and mass spectrometry",W Andy Tao and Ruedi Aebersold, 2003]. ICAT reagents comprise a cysteine-reactive group, a linker containing either heavy or light isotopes and a binding affinity tag.

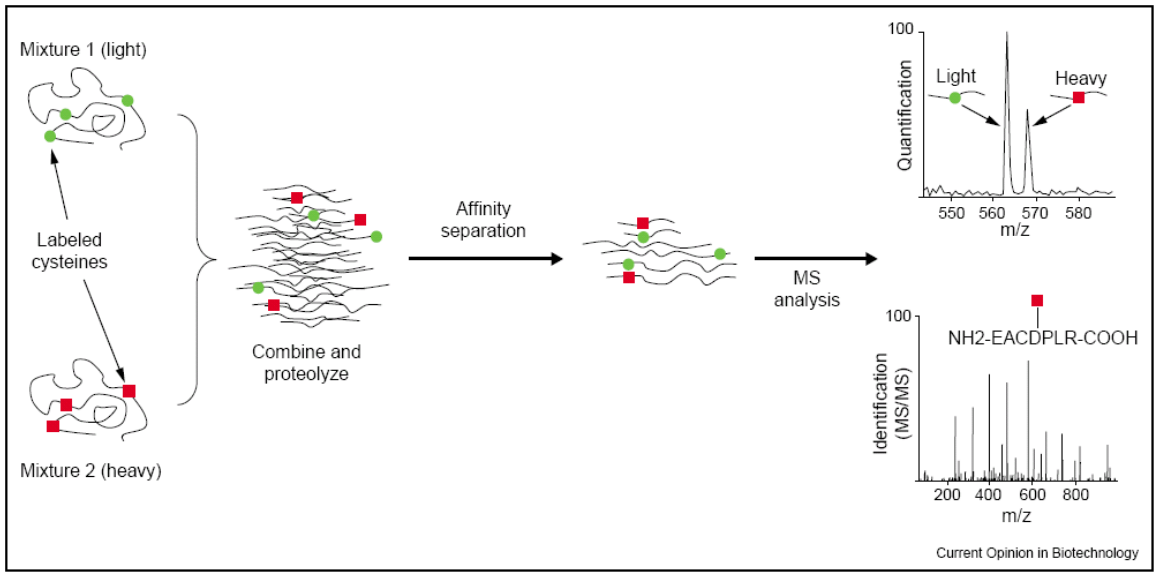


Figure 1.4 The ICAT reagent strategy of quantifying differential proteins ["Advances in quantitative proteomics via stable isotope tagging and mass spectrometry",W Andy Tao and Ruedi Aebersold, 2003].

A more recent study conducted by Molloy and colleagues pointed out an issue that using ICAT-MS technology tends to under represent low molecular weight basic proteins in the proteome [Molloy et al., 2005]. Their research also indicates that ICAT-MS technology was not much superior than 2-DE coupled with MS on detecting highly hydrophobic transmembrane proteins. This will surely limit the application of ICAT-MS technology in these disciplines.

All in all, ICAT MS technology provided essential quantitative reproducibility and sensitivity for general proteomic analysis.

1.3.4 SELDI and Proteomic Pattern Analysis

Surface-enhanced laser desorption/ionization time of flight MS (SELDI-TOF) is a novel approach introduced by Hutchens and Yip [Hutchens et al., 1993]. It combines various retention surfaces with MALDI mass spectrometry. Samples are captured by adsorption, partition, electrostatic interaction, or affinity chromatography on a solid-phase protein chip surface (see **Figure 1.5**) before MALDI-TOF MS analysis.

Liotta and Petricoin's group pioneered the use of SELDI-TOF MS for the detection of prostate and ovarian cancer based on proteomic pattern diagnostic approach [Petricoin III et al., 2002a,b]. It is suggested that this approach can successfully distinguish protein counterparts that differ by expression level, providing a fingerprint map of proteomic contents within different specimens. Its core hypothesis is that proteins or protein fragments produced by neoplastic cells may

eventually enter the blood circulation system. The abundance of these proteins/fragments could be analyzed directly with SELDI-TOF MS. A well designed mathematical algorithm is a must for data interpretation and the extracted infomation could be used for diagnostic applications. In addition to the original FDA group's results, other groups also reported biomarker analysis of prostate cancer using SELDI-TOF proteomic pattern approach [Adam et al., 2002; Qu. Y et al., 2002; Banez et al., 2003; Lehrer et al., 2003]. A similar approach has also been utilized in a study on ovarian cancer [Kozak et al., 2003]. Metal affinity (IMAC-Cu), hydrophobic (C16 or H4) or weak cation exchange (WCX2) chips were used in these researches. In general, it has been suggested that this technology can achieve much higher diagnostic sensitivity and specificity (approaching 100%) in comparison to the classical cancer biomarkers.

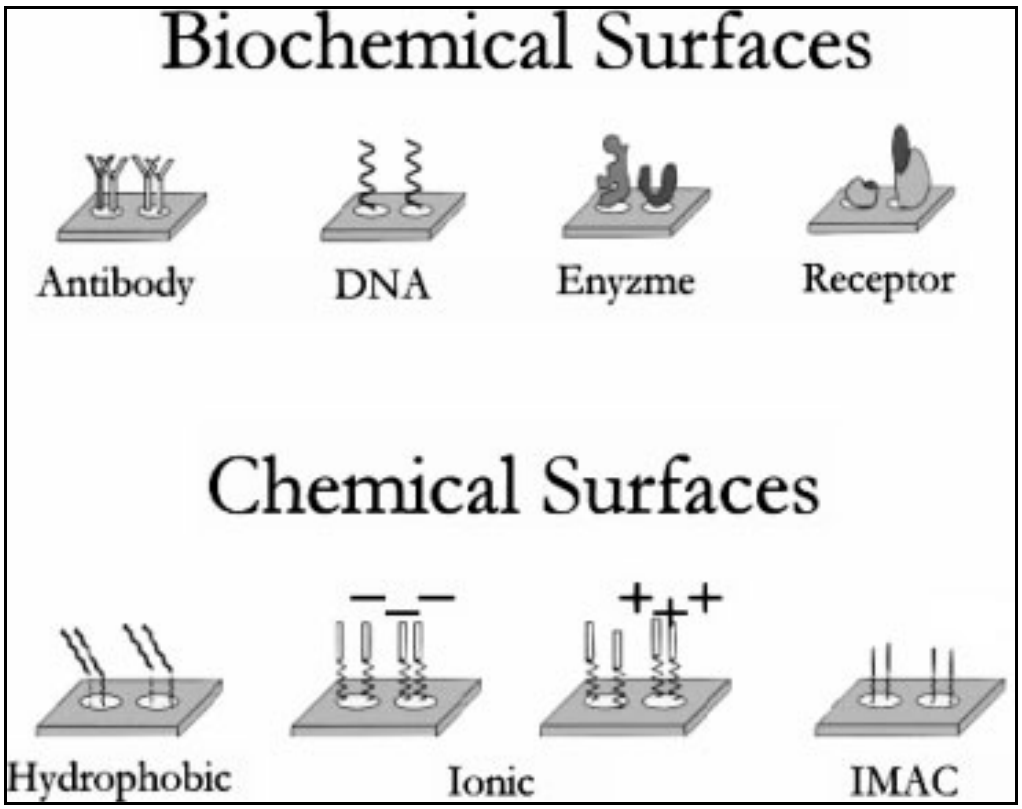


Figure 1.5 The arrays consist of either 8 or 16 spots composed of a specific chromatographic surface ["Current developments in SELDI affinity technology", Ning Tang et al., 2004].

1.4 Issues Associated With Current Human Serum/Plasma Proteome Research and Our Strategy

1.4.1 Dynamic Range Issue

The concentration range of serum/plasma proteome is one of the major challenges in cancer marker detection. Among the high abundance group, serum albumin has a concentration of about 35-50mg/mL. At the low abundance end, interleukin 6 has a normal concentration range between 0-5pg/ml [Anderson et al., 2002]. This extraordinary high dynamic range (10^6~10^9) exceeds most of the contemporary analytical methods, including high efficiency micro capillary LC-MS. Therefore, when targeting low abundance biomarkers, proper fractionation of proteins prior to the final detection and identification is a must for serum/plasma proteome researches. It has been reported that removal of serum albumin and immunoglobulins using affinity chromatography could enhance the detection of low molecular weight/low abundance proteins [Sato et al., 2002; Adkins et al., 2002]. However, those major protein components, especially albumin, in serum are well known to be carrier or transporter of other smaller proteins [Curry et al., 2002]. Therefore, many important low-molecular-weight (LMW) proteins/peptides could have been removed by affinity depletion process. R. S. Tirumalai and colleagues reported an ultrafiltration approach to deplete high abundance proteins and thus to enrich LMW portion of serum proteome [Tirumalai et al., 2003]. With 341 proteins identified, including very low abundance interlukins, their result is pretty attractive. However, among all the proteins identified, the recovery rate of LMW proteins was only 15-20%, and most (83%) of proteins were identified by only 1 unique peptide.

1.4.2 Quantification/Data Interpretation Issue

The SELDI based approach is simple, can be high-throughput, and could be easily automated. However, despite the reported success on detecting cancer from serum, the SELDI based approach using affinity protein chip has several intrinsic limitations, which may compromise the clinical use of this approach.

The most serious issue encountered is its poor reproducibility. In fact, the findings from one lab are essentially not repeatable by others [See **Table 1.1**]. For example, the SELDI approach has been used by several groups to detect prostate cancer [Adam et al., 2002; Petricoin III et al., 2002; Qu. Y et al., 2002; Banez et al., 2003; Lehrer et al., 2003]. Adam et al. found nine peaks, while Petricoin used 7 peaks to differentiate cancer from noncancer. Interestingly, none of the peaks were identical. Qu et al. who originally belonged to Adam's group conducted the research using the same chip and the same instrument. A comparison of the results from these two groups yielded different sets of diagnostic peaks (in fact, only two peaks were identical). In addition, although Adam et al. identified a peak as a marker distinguishing between cancer and noncancer, the same peak was used by Qu et al. to distinguish a healthy individual from a patient with benign hyperplasia.

The reproducibility issue is believed to be caused by nonspecific interaction and saturation of protein chip surface with high abundance interfering proteins in a random manner. Without proper fractionation, analyzing serum samples directly on SELDI chips will not be able to overcome this obstacle.

Reproducibility could be somewhat improved by the use of better algorithms

[Alexe et al., 2004; Petricoin et al., 2004]. It is also suggested that better quality-control and quality-assurance are critical for reducing variability and potential bias [Wulfkuhle et al., 2003; Petricoin et al., 2004]. However, what has been overlooked thus far is the non-quantitative and non-reproducible nature of the currently used SELDI-based serum proteomic pattern approach itself. The proteomic pattern analysis requires at least a semi-quantitative comparison of the relative intensity of a marker peak in both control and disease samples. Nevertheless, both direct serum protein extraction on chip and the following MALDI detection used in the SELDI approach are often biased and not quantitative, thus leading to poor reproducibility.

Table 1: Comparison of 5 reports for prostate cancer diagnosis based on SELDI-TOF technology [a]

Study	Chip Type	Distinguishing Peaks M/Z[b]	Diagnostic Sensitivity and Specificity
Petricoin et al. (39)	Hydrophobic C16	2092, 2367, 2582, 3080, 4819, 5439, 18220	95%; 78-83%
Adam et al. (40)	IMAC-Cu	4475, 5074, 5382, **7024,** 7820, 8141, 9149, 9507, **9656**	83%; 97%
Qu et al. (41)	IMAC-Cu	Non-cancer vs cancer 3963, 4080, 6542, 6797, 6949, 6991, **7024**, 7885, 8067, 8356, **9656**, 9720, Healthy vs BPH[c] 3486, 4071, 4580, 5298, 6099, 7054, 7820, 7844, 8943	97-100%; 97-100%
Banez et al. (42)	WCX2	3972, 8226, 13952, 16087, 25167, 33270	63%; 77%
	IMAC-Cu	3960, 4469, 9713, 10266, 22832	66%; 38%
Lehrer et al. (43)	Hydrophobic H4	Cancer and BPH vs Controls 15200, 15900, 17500 Cancer vs BPH	100% (specificity)
		15200	82%; 67%
		15900	82%; 100%
		17500	64%; 67%

a. This table is from "*Mass Spectrometry as a Diagnostic and a Cancer Biomarker Discovery Tool: Opportunities and Potential Limitations*"*[Eleftherios P. Diamandis, 2004]*.

b. M/Z ratios were rounded to whole numbers for simplicity. M/Z ratios in bold font represent those identified by Adam et al. and Qu et al. for differentiating cancer from non-cancer patients. The underlined M/Z ratio represents a peak identified by Adam et al. for differentiating cancer from non-cancer patients and by Qu et al. for differentiating healthy individuals from patients with benign prostatic hyper plasia.

c. BPH, benign prostatic hyperplasia.

The goal of the first part of my dissertation is to develop a multidimensional fractionation strategy involving molecular weight cut-off (MWCO) ultrafiltration and/or other separation methods which can deplete high abundance proteins as well as retain and enrich low molecular weight (LMW) or low abundance proteins/peptides. Instead of struggling on perfect molecular weight cut-off, we will focus on efficient and quantitative recovery of targeted sub-proteome from human serum samples.

CHAPTER II

SEPARATION AND ENRICHMENT OF FREE-PHASE AND BOUND LOW-MOLECULAR-WEIGHT FRACTION FROM HUMAN SERUM USING ULTRAFILTRATION FOR MALDI-TOF-MS PATTERN ANALYSIS

2.1 Introduction

After completed the Human Genome Project, researchers started a new journey exploring proteomes in various organism. Since proteins play a major role in the molecular pathways that drive all kinds of cellular functions, the elucidation of diseases' mechanisms, such as cancer, requires further study and identification of these proteins [Ashcroft, 2002]. It was widely accepted that early detection is critical in cancer control and prevention. Proteomics study could help finding novel biomarkers and pathological pathways for early detection and prognosis. One of the major parts of cancer proteomics involves the identification of differentially expressed

proteins and quantitative comparison of those proteins at various disease stages with their counterparts in healthy specimen [Srinivas et al., 2001].

Matrix-assisted desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) has been an important tool for proteomic research. Cell [Lee et al., 2004], tissue [Pierre et al., 2001], cerebrospinal fluid [Davidsson et al., 1999], urine [Lin et al., 2006], plasma/serum [Kiernan et al., 2004; Wang et al., 2003], virtually any specimen that are commonly used for clinical research and disease diagnosis could be analyzed using MALDI-TOF-MS.

Surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS), a modified version of MALDI-TOF-MS, has also been used to generate protein profiles directly from plasma/serum. Promising results have been reported on ovarian cancer, prostate cancer, and lung cancer samples, several candidate markers have been found [Petricoin et al., 2002; Kozak et al., 2003; Banez et al., 2003; Zhukov et al., 2003].

However, using MALDI-TOF-MS to generate complete protein expression profiles of these complex biological samples encountered some difficulties, and they are either intrinsic from MALDI-TOF-MS, or directly derived from the complexity of the sample themselves.

Serum, derived from plasma after clotting, contains 60-80mg/ml proteins. The major protein constituents of serum include albumin, immunoglobulins, transferrin, haptoglobin, and lipoproteins [Turner et al., 1970; Burtis et al., 2001]. In addition to these major constituents, serum also contains many other proteins that are

synthesized and secreted, shed, or lost from cells and tissues throughout the body [Wrotnowski, 1998; Kennedy, 2001]. It is estimated that up to 10,000 proteins may be commonly present in serum [Wrotnowski, 1998]. This makes serum a perfect reservoir of finding disease related biomarkers for various organs, including cancer.

However, there is one major limitation for using MALDI-TOF-MS in serum proteomics. All species are not ionized equally. Species with higher abundance or better ionization will tend to depress the signal of those with much lower relative abundance. Therefore, with the presence of high abundance proteins in serum (albumin, for example, 35-50mg/ml in serum), it will be impossible to detect those low abundance proteins (e.g. C-reactive protein, osteoponin, and prostate-specific antigen, which are below 10ng/ml in serum) using MALDI-TOF-MS.

One of the most widely adapted strategies for overcoming this dynamic range issue is to remove high molecular weight proteins, enabling the detection of remaining low molecular weight (LMW) proteins present in much lower concentration. Various immuno-affinity columns have been developed for high abundance protein depletion. Multiple affinity removal system (MARS) from Agilent Technologies could effectively deplete 7 major interfering high-abundant proteins (albumin, IgG, IgA, transferrin, haptoglobin, antitrypsin, and fibrinogen) from human plasma samples, by which eliminates approximately 88-92% of total protein. Removal of these abundant proteins improves the subsequent analysis of the serum sample by effectively expanding the dynamic range of the analysis. However, it is well known that HSA and immunoglobulins are functioning as carrier proteins in serum; many

important low molecular weight peptides/proteins (e.g. hormones) bound to them would be removed with their carrier proteins simultaneously [Zhou et al., 2004]. Because affinity methods used to deplete high abundance proteins under non-denaturing conditions, this may result in the loss of potential disease-informative LMW species.

Ultrafiltration, using molecular weight cut-off (MWCO) membrane, is widely used to concentrate and purify macromolecules. Among these methods of protein separation and fractionation, ultrafiltration is relatively simple, less expensive, and can be used as a continuous process with high throughput at ambient temperature. However, previous studies using ultrafiltration were not very successful. Georgiou and colleagues [Georgiou et al., 2001] demonstrated that they had failed to completely remove HSA and other high abundance proteins using centrifugal ultrafiltration. But it was later pointed out by Tirumalai [Tirumalai et al., 2003] that using undiluted plasma and centrifuging under high speed (12,000xg) may be the cause of their failure. Triumalai and colleagues had compared the protein recovery under non-denaturing (i.e., 25 mM NH4HCO3, pH 8.2) and denaturing conditions (i.e., 25 mM NH4HCO3, pH 8.2 with 20% (v/v) acetonitrile). They found that the LMW fraction was successfully enriched. Highly diluted serum and much slower centrifuge speed (3000xg) were also proved to be effective for preventing HSA from leaking through the filter membrane. They managed to identify 341 unique proteins from low molecular fractions of serum using this improved procedure.

Here we demonstrated the separation and enrichment of free-phase and

bound low molecular weight fraction in serum using MWCO ultrafiltration. Both elution and denaturing conditions were optimized to achieve robust proteomic pattern analysis using MALDI-TOF-MS and high quality protein identification using LC-MS/MS. The result showed that under optimized conditions, ultrafiltration may provide enough separation power to reproducibly elucidate potential biomarker information from low molecular weight fraction of human serum.

2.2 Materials and Methods

2.2.1 Materials and Reagents

Milli-Q water was used for all solutions. Microcon® centrifugal filters (spin column) with a molecular weight cut-off (MWCO) of 30kDa and ZipTip® C^{18} reversed phase pipette tips for sample preparation were purchased from Millopore Corporation (Berford, MA). Phosphate-buffered saline (PBS) solution, trifluoroacetic acid (TFA), HPLC grade acetonitrile, α-cyano-hydroxycinnamic acid (CHCA) were purchased from BIO-RAD (Hercules CA), Pierce (Rockford, IL), Aldrich Chemicals (Milwauke, WI), Bruker Daltonics (Billerica, MA), respectively.

2.2.2 Human Serum Resources and Information

Human serum from healthy donor was purchased from Sigma (St. Louis, MO). Serum was thawed upon arrival and aliquot into smaller fractions to minimize future freeze/thaw cycles. Most of method developments were done using this serum. Serum from ovarian cancer patients and normal controls was obtained from GOG

bank. Totally 20 ovarian cancer patient sample and 19 normal controls was used in this study. These serum samples were also divided into smaller fractions upon receiving and stored under -80 degree before further processing.

2.2.3 Centrifugal Serum Ultrafiltration

2.2.3.1 Free Phase Low Molecular Weight Fraction Collection

The centrifugal filter membrane in Microcon® spin columns were rinsed and used according to the manufacturer's instructions. 50µl serum was diluted with 250µl PBS (pH 7.4), and gently mixed on a vortexer for 2 minutes. The sample was spun at a speed of 14,000 rpm (~16,000xg on Eppendorf Centrifuge 5415C) for 10 minutes before performing ultrafiltration to precipitate any insoluble particles that may clog the membrane. The supernatant was then carefully pipetted into prepared centrifugal filter. The sample was centrifuged (Eppendorf Centrifuge 5415C, Brinkman Instruments, Westbury, NY, USA) at 5000 rpm for 30 minutes until about 90% of the input serum solution had passed through the membrane, meanwhile, make sure there was enough retentate left to avoid membrane dryness. An aliquot of 1xPBS solution was then added into the sample reservoir to dilute the retentate to 300µl, and the retentate was gently mixed by performing 10-15 aspirate-dispense cycles with pipette tip before another consecutive centrifuge. Such operation was performed 8 times totally. The filtrates in separate tubes were lyophilized to dry in Speed-Vac and resuspended in 50µl of 0.1% TFA solution before MS analysis.

2.2.3.2 Bound Low Molecular Weight Fraction Collection

After free-phase low molecular weight (LMW) fraction being eluted, the retentate was collected by reversing the sample reservoir and centrifuging into a new collection tube at proper lower speed according to the instruction manual. The total volume of each retentate was about 30uL. It was then incubated with 100µl 8M urea for at least 2 hours under ambient temperature. After that, the elution of bound LMW fraction from denatured retentate was carried out in a new 30k MWCO spin column, and all the eluting operations were the same as described in section **2.3.1**.

2.2.4 MALDI-TOF-MS Pattern Analysis

2.2.4.1 Sample Preparation

ZipTip® C^{18} reversed phase pipette tips (Millopore, Billerica, MA) were used for de-salting and concentrating of peptides/proteins prior to MALDI-TOF-MS analysis. 10µl of the resuspended filtrates (free-phase or bound) were used for each clean-up process. The elution buffer is 10ul 50% acetonenitrile with 0.1% TFA if not otherwise mentioned. The reproducibility of this process was also tested using MALDI-TOF-MS pattern analysis of resuspended filtrates from three separate Microcon® spin columns.

2.2.4.2 MALDI-TOF-MS

Dried-droplet deposition of the sample was achieved by depositing 1µl (0.5µl each time, apply the rest after the first drop dried) of sample solution onto stainless

steel MALDI target, and then overlaid with 0.5µl of CHCA (α-Cyano-4-Hydroxycinnamic Acid) matrix solution (5mg/ml in 50% acetonitrile, 0.1% TFA). For each sample, three duplicate spots were prepared to eliminate possible fluctuations.

All mass spectra were collected using OminiFLex^{TM} MALDI-TOF mass spectrometer (Bruker, Bilerica, MA, USA) operated in the linear, positive mode with delayed extraction turned on. OminiFLex^{TM} MALDI-TOF mass spectrometer uses a 337nm pulsed nitrogen laser, which has a duration of 3ns and a full power of 175µJ. MS spectra were collected in the range of 700-10000 m/z ratio. All experiments were performed under 75% of full laser power with sampling rate of 5Hz if not otherwise specified. 10 sampling points were randomly picked on each sample spot. One hundred shots per sampling points were accumulated and all 10 sets of spectra were combined to eliminate on-target fluctuation.

2.2.5 LC-MS/MS Analysis and Protein Identification

2.2.5.1 Trypsin Digestion

The collected LMW species were concentrated to 80µL/sample using speed-vac and the PH of the solutions is adjusted to 7.8 using freshly made 100mM NH_4HCO_3 if necessary. Trypsin digestion was conducted in solution using standard procedure with some modifications. In general, there was no denaturation, reduction, and alkylation before adding trypsin. The first dose of trypsin was added at a ratio of 1:100 (w/w) and incubated under 37℃ for 4 hours. The second dose of trypsin (1:50

w/w) was added afterwards and incubated overnight under 37℃. Tryptical digests were frozen under -20℃ before further analysis.

2.2.5.2 LC-MS/MS Analysis and Protein Identification

The LC-MS/MS analysis was performed on Agilent 1100 HPLC system coupled in-line with a Bruker HCT 3000plus ESI-IonTrap Mass Spectrometer. 20μL tryptic digested sample was injected and eluted from a Vydac C18 capillary RP-HPLC column (10cm, 300μm ID) using the following gradient: 2% B, 10min; 5%-45% B, 200min; 45%-85% B, 30min; 85%B, 30min. Solvent A is 0.1% formic acid and 1% acetonitrile in Milli-Q H_2O, solvent B is 0.1% formic acid in acetonitrile.

Tendem MS data were merged into one single data file for both free-phase and bound species. It was searched against NCBInr protein database using an in-house Mascot server. The search was performed using no enzyme, trypsin, trypsin+chymotrypsin separately and the results were combined. No taxonomy or modifications were specified. Peptide mass tolerance is ±1.0Da, MS/MS tolerance is ±0.5Da, and searched for all possible charge states (i.e. +1, +2, +3). The final result was manually inspected to remove all false positives.

2.3 Results and Discussion

2.3.1 Centrifugal Ultrafiltration

Centrifugal ultrafiltration has long been used as clean-up/concentrating method on bio-macromolecules such as DNA and proteins. The unique molecular

weight cut-off membrane provided fast sample processing and high recovery rate. However, when used on serum proteomics, especially for high-abundance protein depletion, this method required special optimization.

Tirumalai [Tirumalai et. al., 2003] and Harper [Harper et al., 2004] had investigated how experimental conditions, e.g. sample preparation, centrifugation time and speed, sample size and protein concentration, would effect the result of centrifugal ultrafiltration fractionation of serum. Some key factors had been pointed out but the results were still far from satisfactory.

In our study, we introduced a systematic way of evaluating the performance of centrifugal ultrafiltration in low molecular weight serum proteome studies. The schematic demonstration of the strategy is shown in **Figure 2.1**.

In previous study [Tirumalai et. al., 2003], low molecular weight fractions collected under non-denaturing and denaturing conditions were compared. As a conclusion, much more proteins/peptides could be detected under denaturing condition. Therefore denaturing conditions should be preferred and free-phase species eluted under non-denaturing conditions was neglected. However, our study showed that the amount of free-phase species could be much larger than people had thought before.

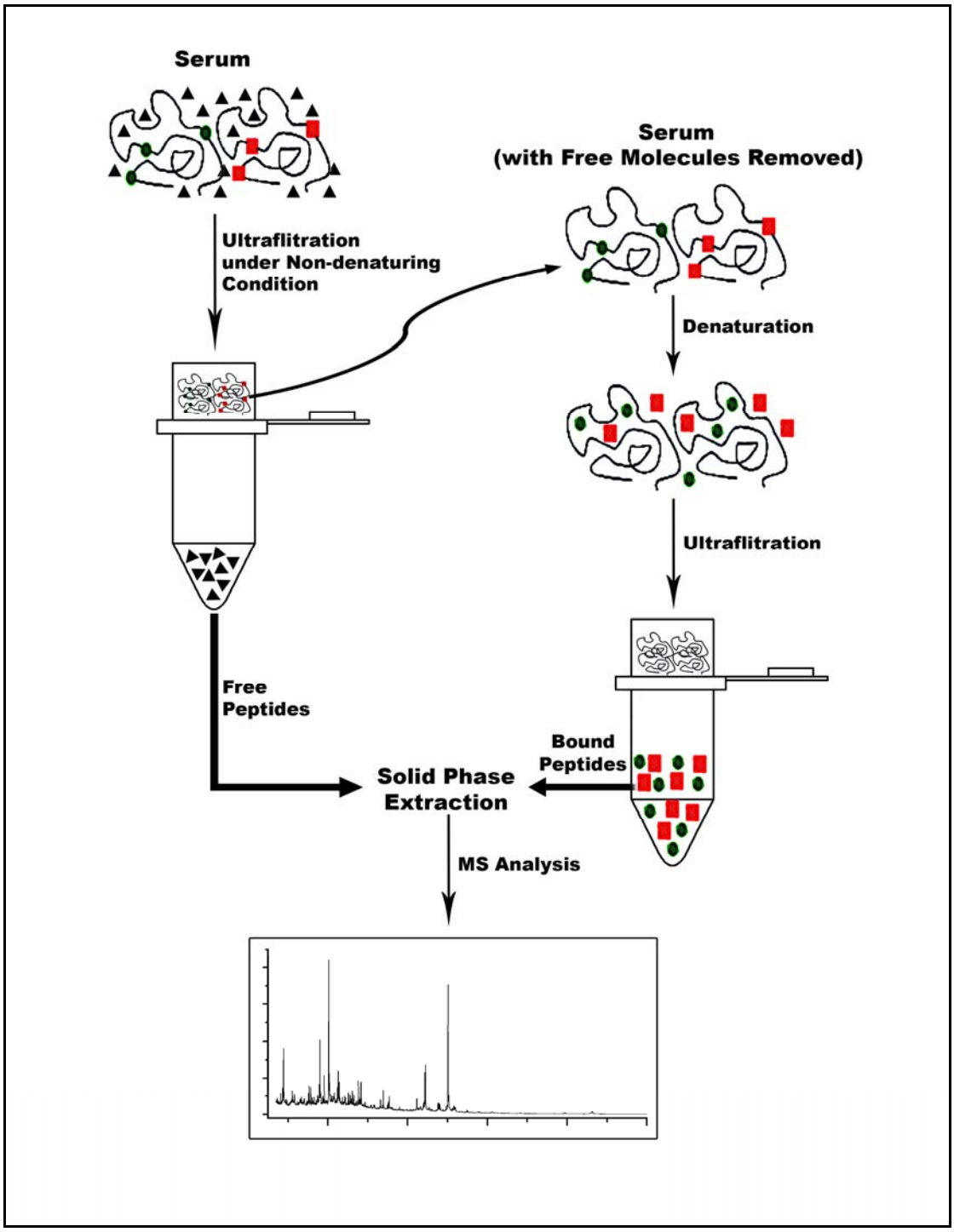


Figure 2.1 A schematic representation of separating and enriching free-phase and bound peptides from serum using centrifugal ultrafiltration and MALDI-TOF pattern analysis.

In our study, serum sample was diluted with PBS to prevent denaturing serum proteins, which minimized the disassociation of low molecular weight fraction from major carrier proteins. A low centrifuge speed was also adapted from previous studies to avoid protein leaking. Furthermore, we have found that only filtering once was not sufficient enough to completely elute free peptide/proteins. A total of 8 consecutive ultrafiltrations under non-denaturing condition were performed to maximally separate free-phase peptide/proteins. A SDS-PAGE analysis was performed to demonstrate the effect of this process (See **Figure 2.2**).

According to the results in **Figure 2.2**, it was after 6 consecutive non-denaturing ultrafiltrations when the amount of LMW proteins could be neglected. During consecutive centrifuge process, the retentate was carefully resuspended by gently performing aspirate-dispense cycles with pipette tip while not touching the membrane, which could maximally avoid protein accumulation on the membrane and pore clogging. The aspirate-dispense cycles made the solution be homogeneous without concentration gradient, which resulted in uniform experimental conditions for each ultrafiltration.

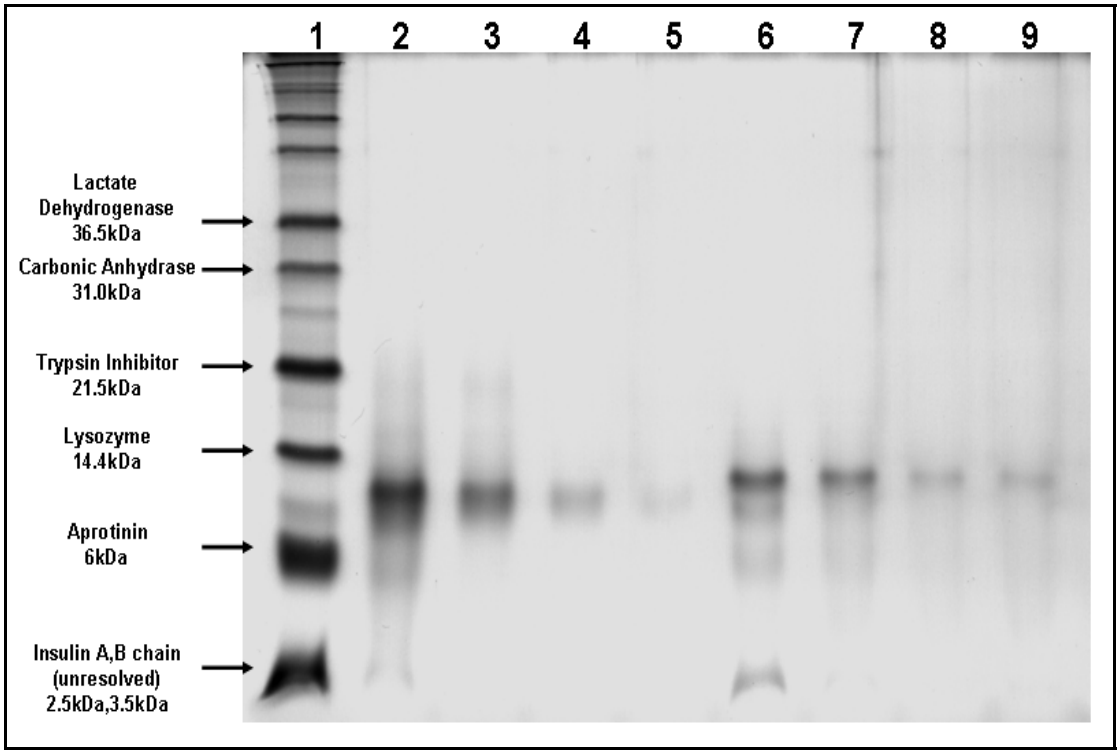


Figure 2.2 SDS-PAGE analysis of free-phase and bound peptides in serum using ultrafiltration with 30kDa cut-off spin column.
Lane 1: Molecular weight marker;
Lane 2 to Lane 5: Free-phase peptides, eluant 1 and 2, 3 and 4, 5 and 6, 7 and 8 combined respectively;
Lane 6 to 9: Bound peptides, eluant 1 and 2, 3 and 4, 5 and 6, 7 and 8 combined respectively.

The retentate part containing large carrier proteins was then incubated with 8M urea solution, resulting in the disruption of the protein/protein interaction and separation of bound peptide/proteins from their carrier proteins. Thereafter, the denatured retentate was filtrated with a new 30kDa molecular weight cut-off spin column to collect the bound LMW species. PBS solution was used to dilute the denatured protein solution and the elution of dissociated species. This could ensure similar elution efficiency and minimize buffer effect during ultrafiltration procedure. No gel-like precipitate or membrane clogging was observed during the ultrafiltration of denatured proteins.

As could be seen in **Figure 2.2**, the amount of low molecular weight proteins/peptides greatly increased after denaturing ultrafiltration. Moreover, most of LMW fraction disassociated from their carrier proteins by denaturing also can be completely eluted after 8 consecutive filtrations. And the components in free-phase LMW fractions were dramatically different (smaller molecular weight) from those eluted in bound fractions. The result clearly indicated the selective enrichment of both free-phase and bound species. The reproducibility of this consecutive ultrafiltration process was examined by MALDI-TOF MS pattern analysis using combined mixture of 8 eluants for free-phase and bound peptides, respectively. The results were shown in **Figure 2.3**.

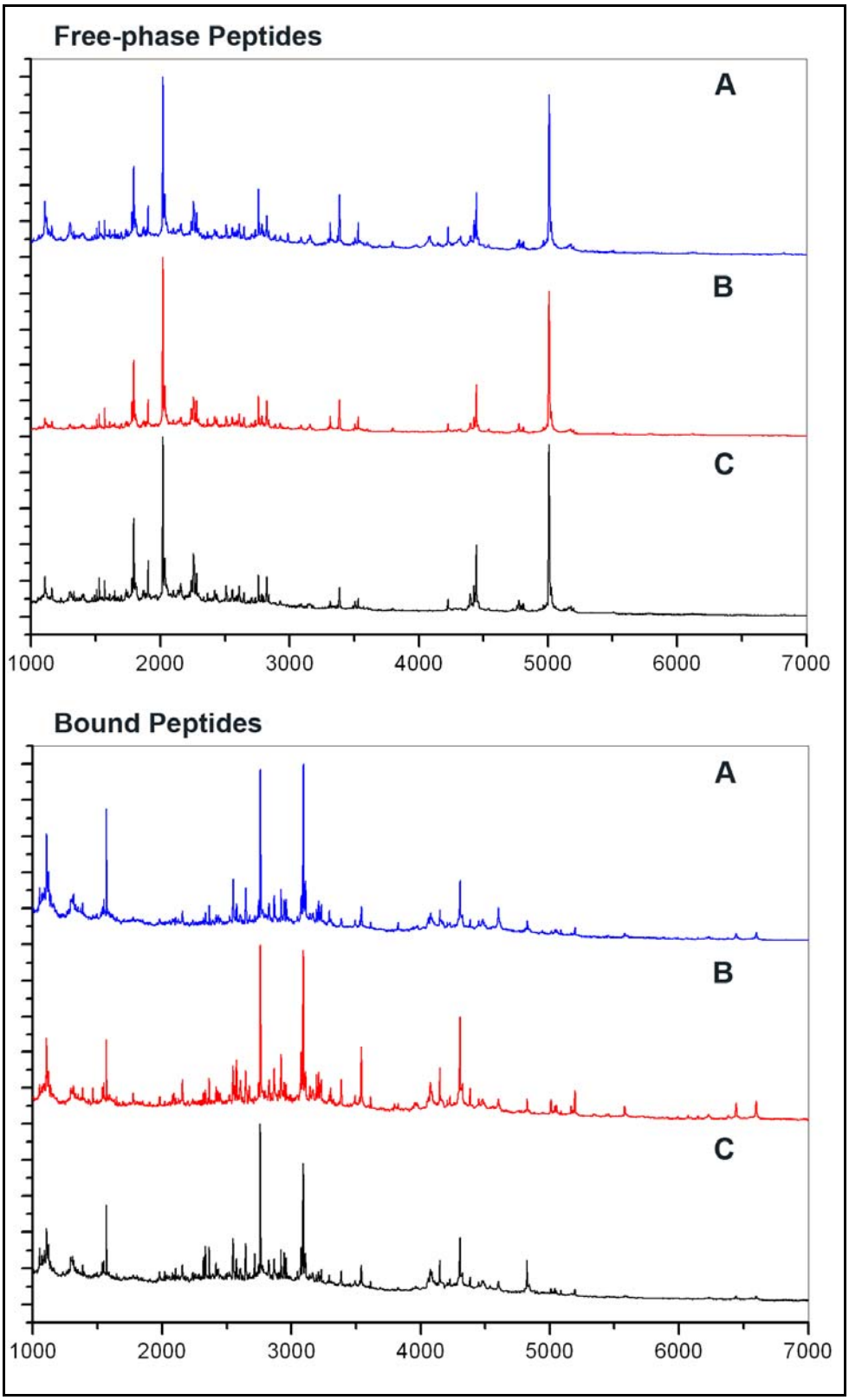


Figure 2.3 MALDI-TOF spectra obtained from the samples filtrated by three different spincolumns.

As indicated in MALDI-TOF MS pattern analysis, the free-phase peptides demonstrated major difference from bound fraction. A side by side comparison indicated that some major small molecular weight peaks in free-phase protein/peptide samples disappeared or become weaker in bound peptide spectra, and many peaks with larger M/Z ratio were much more abundant in bound protein/peptide spectra. This is consistent with SDS-PAGE results. In summary, the ultrafiltration strategy we developed can lead to both enrichment of LMW serum peptides and separation of free-phase proteins/peptides from bound ones. Another important observation is that the total quantity of the free-phase proteins/peptides is comparable to that of the bound species in serum.

A series of test were carried out to verify that the increased quantity of proteins after denaturation was neither yielded by the effect of membrane clogging during centrifugal ultrafiltration (free-phase proteins/peptides leftover on membrane), nor by the solvent elution efficiency difference between PBS and 8M urea.

Firstly, the retentate containing high molecular weight carrier proteins was collected by placing the sample reservoir of Microcon spin column upside down, and spun at a proper low speed according to the instruction manual. One part of collected retentate was denatured and filtered in the same used Microcon spin column. Before reusing, the used empty column was spun twice using PBS to remove possible proteins/peptides leftovers on the membrane. In fact, no bands were observed for collected filtrates on SDS-PAGE analysis (data not shown). And then the denatured retentate was spun 8 times in washed used column using PBS as elution buffer. The

other part of retentate was spun using used Microcon spin column directly without washing. A third part of retentate was denatured and filtered using new clean Microcon spin column, following procedure described before. MALDI-TOF MS pattern analysis results were compared for these 3 differentially collected bound peptide samples (see **Figure 2.4A**, **B**, **C**, respectively).

Secondly, the combined free-phase filtrates in PBS solution were also tested in a similar manner, except one part was eluted directly with PBS for 8 times, the other was incubated with 8M urea and then eluted 8 times with PBS. Both of them were filtrated with using new spin columns. MALDI-TOF MS pattern analysis results were also compared for these 2 differentially collected free-phase peptide samples in **Figure 2.4D and 2.4E**.

As seen in **Figure 2.4**, with a slight difference of peak intensity, almost identical proteomic patterns had been generated in each group of tests. Denatured retentates generated completely deferent pattern from free-phase eluants. The patterns are almost identical with their counterparts in **Figure 2.3**. These results suggested that the decreasing quantity of free-phase peptides after multi-centrifugal ultrafiltration and the increase in quantity of low molecular weight fraction after urea denaturation were not resulted from membrane clogging or solvent effect. Separation of free-phase LMW species, dissociation and enrichment of bound proteins/peptides from large carrier proteins were achieved effectively using urea denaturatoin and centrifugal ultrafiltration.

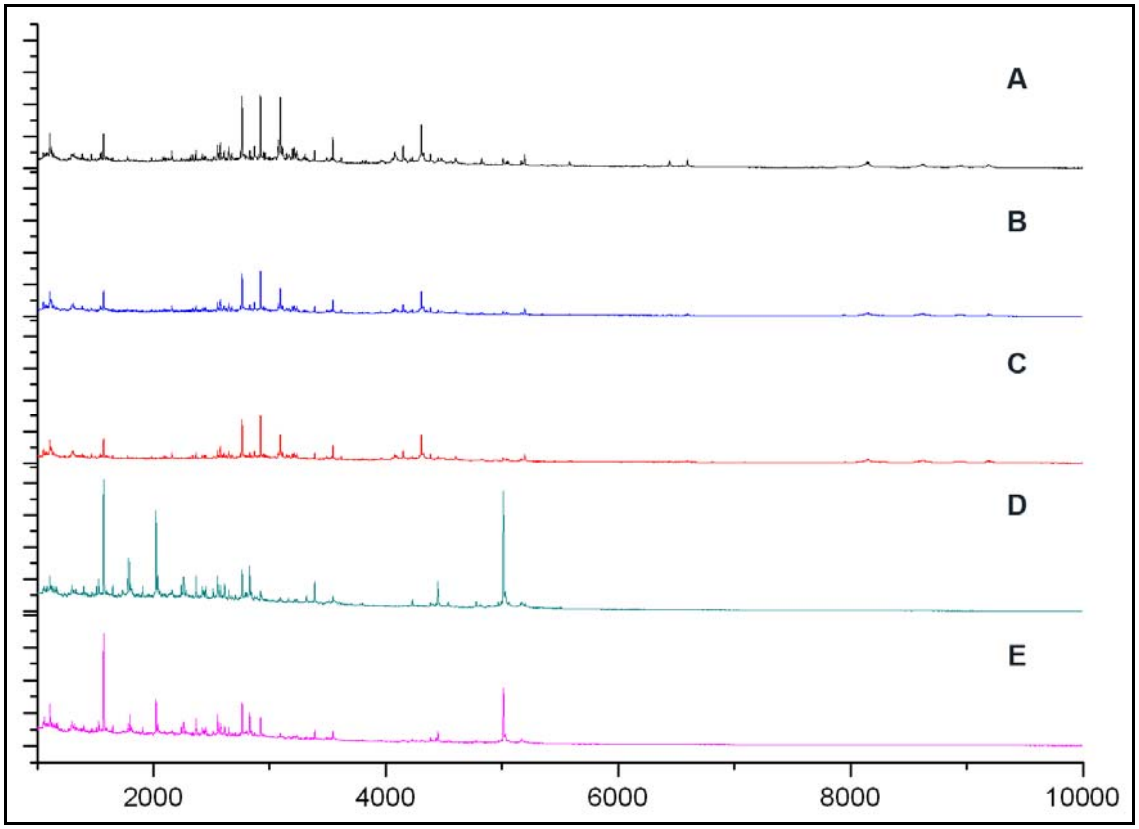


Figure 2.4 Test of membrane clogging and solvent efficiency after 8 consecutive ultrafiltration.

A. MALDI-TOF spectra of the re-centrifugal ultrafiltration of bound peptides from serum, using used spin column without additional wash.
B. MALDI-TOF spectra of the re-centrifugal ultrafiltration of bound peptides from serum, using used spin column after additional wash.
C. MALDI-TOF spectra of the re-centrifugal ultrafiltration of bound peptides from serum, using new spin column.
D. MALDI-TOF spectra of the re-centrifugal ultrafiltration of free-phase peptides from serum, eluted after 8M urea incubation.
E. MALDI-TOF spectra of the re-centrifugal ultrafiltration of free-phase peptides from serum, eluted directly with PBS solution.

2.3.2 MALDI-TOF Pattern Analysis - Method Development

MALDI-TOF pattern analysis was introduced in human serum proteomic research by Patricoin and Liotta [Petricoin III et al., J. Natl. Cancer Inst. 2002, Lancet 2002]. Although the original SELDI approach resulted in debatable data [Diamandis, 2004], the idea of using differential comparison of healthy and patient proteomic patterns for biomarker discovery is still a widely accepted strategy.

We developed a data acquisition procedure to robustly generate easy to read proteomic patterns using MALDI-TOF MS. Every pattern was a combined spectrum from 10 randomly selected sampling spots with 100 shots per spot for each sample deposited on target. This would reduce spot to spot variation during sample preparation in MALDI-TOF MS (known as “hot-spot effect”).

One set of MALDI-TOF MS spectra were collected from three MALDI samples deposited on target using the same eluted LMW proteins/peptides. It was seen that with the averaging effect, both the proteomic pattern and the relative intensity of peaks were highly reproducible. And all combined spectra had shown strong signal intensity with high resolution (data not shown). This made visual identification of potential biomarker peaks a much easier job to do.

2.3.3 Protein Identification Using LC/MS/MS with Database Searching

Protein identification results using LC/MS/MS with database searching were summarized in **Table 2.1**.

At first look, we identified less proteins compared with previous study. But

our strategy has several advantages. Firstly, the data quality and the confidence level of identification were much better than previous study. As mentioned before, our 8-cycle elution strategy could ensure complete collection of LMW species. Enrichment of free-phase and bound peptides in two different step also could be considered as an extra fractionation besides reversed-phase HPLC separation. Without interferences from large high abundance proteins, the sensitivity of high resolution mass spectrometry was further extended to detect low molecular weight fragments. Secondly, since the strategies used in both studies were focused on low molecular weight species in human serum, the number of LMW proteins detected should be an important factor to evaluate the efficacy of the method. As indicated in table 1, our method demonstrated much better result than previous study (90 LMW proteins identified in our study v.s. 69 in previous study). Lastly, fragments from same protein could be collected separately in both free-phase and bound fractions. So combining MS/MS data before database searching could be one of the essential factors for protein identification. Searching against non-tryptic fragmentation also could help identify LMW species naturally present in serum.

Table 2.1 Protein Identification Result Compared With Previous Study [Tirumalai et al., 2003]

	Previous Study	Our Result
Total Proteins Identified	341	226
# of proteins Identified by only one unique peptide	285 (83.5%)	36(16%)
# of proteins with predicted MW ***higher*** than Cut-off MW (30kDa) of spin colum	282 (82.7%)	136 (60.2%)
# of proteins with predicted MW lower than 20k	38 (11%)	32 (14%)

2.3.4 Biomarker Discovery Using MALDI-TOF Pattern Analysis

We were able to test our whole proteomic strategy on some clinical samples obtained from GOG bank. A total of 39 serum samples were tested in our project. There are 20 patient samples with ovarian cancer and 19 normal serum samples as control set.

All experimental setup was the same as described above. MALDI-TOF pattern analysis was performed for free-phase LMW peptides, free-phase-peptide-removed denatured LMW peptides and directly denatured LMW peptides. All spectra were externally calibrated and normalized before bioinformatic analysis.

Significant differences were found between cancer patient and normal control when comparing free-phase peptide patterns. Same results were also found when using directly denatured peptide patterns. Based on previous test results on centrifugal ultrafiltration method, we could conclude that these differences were contributed solely by free-phase peptides.

A typical pattern comparison was shown in **Figure 2.5a**. As could be seen, three peaks were identified to differentiate cancer patients and normal control. To further compensate for the intrinsic disadvantage of quantification using MALDI-TOF MS, we used relative peak intensity ratio instead of peak intensity itself to define cut-off thresholds. So there are 2 marker ratios used to distinguish cancer and normal samples: M1/M2>1 and M1/M3>1. Both of them have to be satisfied for positive classification of cancer. It could not be classified in either of the two groups if

only one of them is satisfied.

Using these markers, 16 out of 20 cancer patients (80%) were correctly identified. In the 4 misclassified samples, only 1 demonstrated a “normal” pattern (M1/M2 and M1/M3 are all smaller than 1). We failed to detect any marker peaks in the other 3 samples (as could be seen in **Figure 2.5b**).

Correct marker ratios (M1/M2<1 and M1/M3<1) were found in 79% of normal samples. There were 2 samples that demonstrated false positive signals, and the other 2 samples could not be classified in either group.

A t-test was also performed to verify the statistical significance of our results. With an assigned p-value threshold as 0.05, marker ratio 1 (M1/M2) between two sample groups has a p-value of 0.0009, and marker ratio 2 (M1/M3) is 0.04, all indicating significant differences between patient and normal samples. Better results could be obtained if we discard the 3 patient samples without any marker peaks (M1/M2 p-value=0.0002 and M1/M3 p-value=0.025).

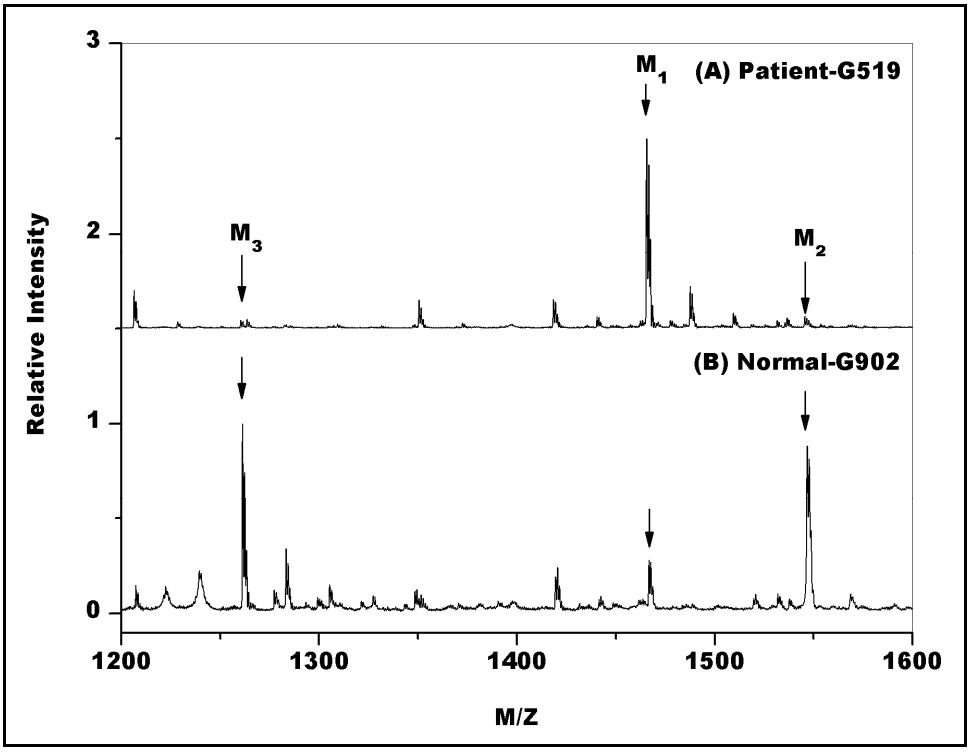


Figure 2.5a. MALDI-TOF-MS pattern analysis of ovarian cancer patient serum and normal serum.

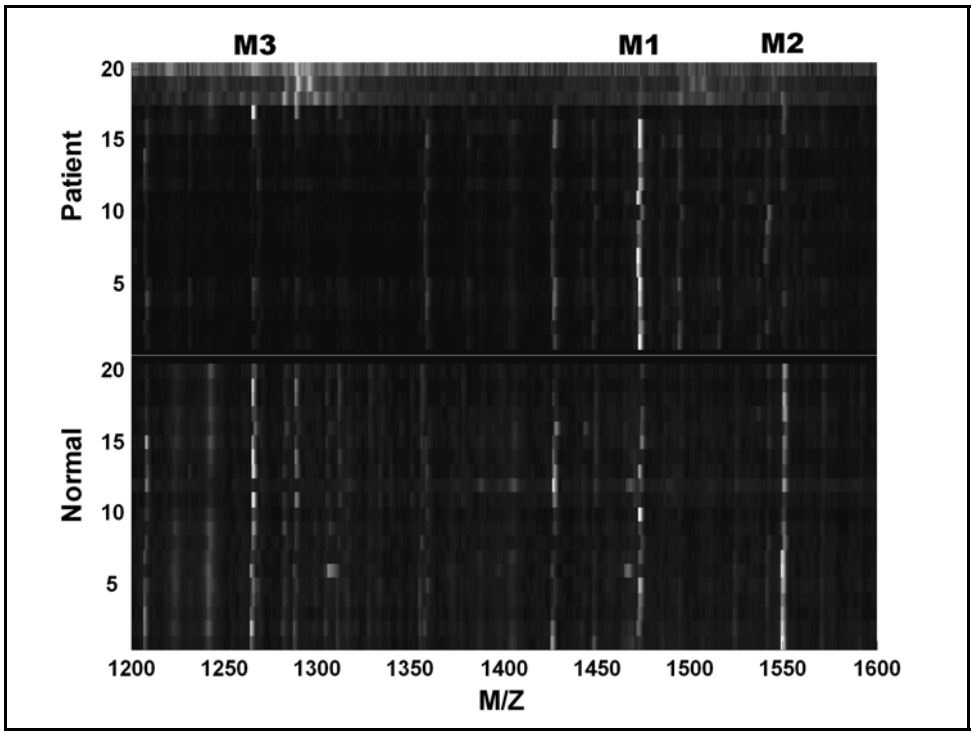


Figure 2.5b. Software generated "virtual gel" indicating selected marker peaks in 2 sample sets.

2.4 Conclusion

In this project, we evaluated a new application of centrifugal ultrafiltration in human serum proteome study. Our results demonstrated successful separation and enrichment of free-phase proteins/peptides and proteins/peptides bound to large carrier proteins in serum.

Combining centrifugal ultrafiltration with MALDI-TOF MS pattern analysis, we selected 3 marker peaks for differentiating ovarian cancer patients and healthy controls. And we believe that these markers were found from free-phase fractions. The sensitivity and specificity are 80% and 79%, respectively (using peak intensity ratios). This is only a small pilot study with limited number of samples. More validation should be done with larger sample sets using more standardized procedures.

Based on the procedure we used in centrifugal ultrafiltration, a more accurate definition for so called “free-phase” species should be “free and less tightly bound molecules”, and this part of serum proteome was somehow ignored in previous studies. Our results agreed with previous studies on the supreme complex nature of serum proteome but provided new possibilities of finding potential biomarkers.

It will never be emphasized enough about the importance of removing high abundance proteins before further study of the rest of serum proteome. However, removal should not be the ending of these hard-working carriers. Each one of them might represent a whole new sub-proteome and further thorough exploring may lead to a world we never know.

CHAPTER III

A NOVEL PROTEOMIC APPROACH USING ORGANIC SOLVENT PRECIPITATION AND ENZYMATIC DEGLYCOSYLATION FOR ELUCIDATING OVARIAN CANCER MARKERS IN HUMAN SERUM

3.1 Introduction

During the last decades, proteomics has been well accepted as a new systematic tool of understanding various biological disorders. High sensitivity and automated mass spectrometric techniques have been developed to enhance the capability of analyzing highly complex biological samples.

However, there are still considerable limitations. When encountering whole proteomes, the extraordinary dynamic range of protein components in a single cell, or in plasma and serum has long been a huge obstacle before researchers. A 7-10 orders' difference in concentration is surely out of the reach of any single contemporary

proteomic techniques [Stasyk and Huber, 2004].

Pre-fractionation of the whole proteome before further analysis has been a widely used strategy especially when focusing on biomarker discovery in low molecular weight (LMW) fraction of human plasma/serum. Centrifugal ultrafiltration has been used to separate and enrich the LMW serum proteome [Tirumalai et al., 2003]. And it has also been used to demonstrate the complicated interactions between serum proteins [Zhou et al., 2004]. Various immuno-affinity columns have been developed to conveniently remove high abundance proteins from human serum [Bjöhall et al., 2005], and a similar approach using affinity spin tube filter was also shown to be effective [Wang et al., 2003].

However, there is one limitation associated with direct removal of large high abundance proteins. In serum, dominating proteins such as albumin are known to be carriers and transporters of proteins and other species [Schussler, 2000; Curry, 2002; Dea et al., 2002]. Removal of LMW species spontaneously with their carriers may greatly effect further identification and quantitative analysis. Therefore, disrupting protein-protein interactions before removal of high abundance carrier proteins is an essential step for LMW serum proteomic research. Tirumalai and Zhou demonstrated that using centrifugal ultrafiltration under denaturing conditions could effectively deplete high abundance large proteins as well as separate and enrich LMW proteins in human serum; several new low abundance proteins in serum proteomic study has been identified, too [Tirumalai et al., 2003; Zhou et al., 2004]. In our previous study (see **Chapter II**), centrifugal ultrafiltration was used under optimized conditions to

separate and enrich both free-phase and bound LMW species in human serum. Furthermore, a series of potential markers were identified for distinguishing ovarian cancer patients and normal controls using MALDI-TOF MS pattern analysis.

A novel approach using organic solvent extraction as a sample preparation method for mass spectrometry based proteomic analysis has been reported recently [Chertov et al., 2004]. They were using 50% acetonitrile solution with 0.1% trifluoroacetic acid as denaturing reagent to extract LMW proteins/peptides from both turmor challenged mice serum and normal human serum samples. The following surface enhanced laser desorption/ionization (SELDI) TOF MS analysis identified several differentially expressed protein markers. Compared with centrifugal ultrafiltration, organic solvent precipitation is simple and effective on high abundance protein removal and LMW protein enrichment.

In this study, we demonstrated an organic solvent precipitation approach on detecting serum biomarkers for ovarian cancer. High concentration of acetonitrile solution (80%) was used to improve high abundance protein depletion. An enzymatic deglycosylation step was added before MALDI-TOF MS pattern analysis. It was proved that for complex samples such as serum extract, deglycosylation could reduce suppression effect and improve signal to noise ratio within low molecular weight range. Biomarkers found from previous study were further evaluated and verified. One of the major marker peaks was identified as des-alanine fibrinopeptide A and this marker could be used to distinguish benign, carcinoma and non-cancer samples successfully.

3.2. Methods and Materials

3.2.1 Materials

HPLC grade acetonenitrile was purchased from sigma (St. Louis, MO). PNGase F (500 units/μL) was purchased from New England Biolabs (Ipswich, MA). An aliquot of enzyme was diluted to 5units/μL immediately upon receiving with supplied 1x G7 reaction buffer (50 mM sodium phosphate, pH 7.5) and stored at 4℃ until use. Microcon® centrifugal filters (spin column) with a molecular weight cutoff (MWCO) of 30kDa and ZipTip® C^{18} reversed phase pipette tips were purchased from Millopore Corporation (Berford, MA). Peptide calibration standard (1000~4000Da) and ultra-pure α-CHCA (α-cyano-hydroxycinnamic acid) matrix were purchased from Bruker Daltonics (Billerica, MA).

3.2.2 Serum Sample Information

Serum samples were obtained from multiple institutions. The first sample set was obtained from Gynecologic Oncology Group tissue bank (18 normal, 20 patients). All patients were diagnosed to have advanced stage ovarian cancer: 17 papillary serous carcinomas, 1 endometrioid carcinoma, 1 clear cell carcinoma and 1 unknown type.

The second sample set was obtained from Cooperative Human Tissue Network – University of Pennsylvania Medical Center. There are 19 samples totally, and they may have diseases other than cancer but no detailed information available. One of them was from male donor and therefore excluded from our study.

The third sample set was again obtained from Gynecologic Oncology Group tissue bank (40 carcinomas and 20 benign). In cancer patient samples, there are 20 papillary serous carcinomas, 10 mucinous carcinomas, 5 endometrioid carcinomas and 5 clear cell carcinomas. In samples from benign diseases, there are 10 serous benign, 5 mucinous benign and 5 other benign.

All serum samples were aliquot to small fractions and stored under -80℃ before use.

3.2.3 Denaturation of Serum and Separation of LMW Species

40μL of serum was diluted and denatured in 2mL of 80% acetonenitrile in a 15mL centrifuge tube. After immediate vortexing, the solution was centrifuged (Beckman Coulter Allegra™ X-22R centrifuge with Beckman F0685 fixed angle rotor) at 10,000 rpm (~10500xg) for 3 minutes to collect precipitated proteins. 1500μL of supernatants (equivalent to 30μL serum extracts) were carefully transferred into a 2mL centrifuge tube and evaporate to 250μL using speed-vac. The sample is now ready for deglycosylation.

3.2.4 Deglycosylation

The pH of extracted sample was adjusted to 7.8 using 100mM NH_4HCO_3 solution. 25 units (5ul 5units/ul solution) of PNGase F were added to each sample and gently mixed. All samples were incubated at 37℃ overnight. The next day, the reaction were stopped by freezing samples under -20℃.

After deglycosylation, 20μL of each sample (equivalent to ~3μL serum extract) was desalted using C^{18} Ziptip™ pipette tips following procedures slightly modified from manufacturer's instruction. The only change is that we used 10μL of 50% and 80% acetonenitrile each to elute the desalted peptides off Ziptip™ and 2 aliquots were then combined and concentrated to ~2μL before applied on MALDI target.

3.2.5 MALDI-TOF Pattern Analysis

Desalted samples were applied on stainless steel MALDI target first. After all samples are applied and dried, 0.3μL of matrix solution (5mg/mL α-CHCA in 50% acetonenitrile, 0.1% TFA) was applied on top of each sample.

All mass spectra were collected using OminiFLex™ MALDI-TOF mass spectrometer (Bruker, Bilerica, MA, USA) operated in the linear, positive mode with delayed extraction turned on. OminiFLex™ MALDI-TOF mass spectrometer uses a 337nm pulsed nitrogen laser, which has a duration of 3ns and a full power of 175μJ. MS spectra were collected in the range of 700-10000 m/z ratio. All experiments were performed under 77% of full laser power with sampling rate of 2Hz if not otherwise specified. 3 sampling points were randomly picked on each sample spot. One hundred shots per sampling points were accumulated and all 3 sets of spectra were combined to eliminate on-target fluctuation. For those samples with low signal intensity, the first 10 spectra were discarded to minimize matrix effect and get better signal to noise ratio within targeted m/z range.

The mass-spectrometer was calibrated with peptide calibration standard (1000~4000Da) from Bruker, and all samples were then calibrated externally and normalized.

3.2.6 LC-MS/MS Analysis and Protein Identification

3.2.6.1 Trypsin Digestion

After MALDI-TOF-MS pattern analysis, samples from GOG bank (18 normal samples and 20 cancer patient samples) were combined into 2 general mixtures: patient and normal. One half of combined LMW species were concentrated to 80μL/sample using speed-vac and the PH of the solutions is adjusted to 7.8 using freshly made 100mM NH_4HCO_3 if necessary. Trypsin digestion was conducted in solution using standard procedure with some modifications. In general, there was no denaturation, reduction, and alkylation before adding trypsin. The first dose of trypsin was added at a ratio of 1:100 (w/w) and incubated under 37℃ for 4 hours. The second dose of trypsin (1:50 w/w) was added afterwards and incubated overnight under 37℃. Tryptical digests were frozen under -20℃ before further analysis.

3.2.6.2 LC-MS/MS Analysis and Protein Identification

The LC-MS/MS analysis was performed on Agilent 1100 HPLC system coupled in-line with a Bruker HCT 3000plus ESI-IonTrap Mass Spectrometer. 20μL tryptic digested sample was injected and eluted from a Vydac C18 capillary RP-HPLC column (10cm, 300μm ID) using the following gradient: 2% B, 10min; 5%-45% B,

200min; 45%-85% B, 30min; 85%B, 30min. Solvent A is 0.1% formic acid and 1% acetonitrile in Mili-Q H_2O, solvent B is 0.1% formic acid in acetonitrile.

Tandem MS data were searched against NCBInr protein database using an in-house Mascot server. The search was performed using no enzyme, trypsin, trypsin+chymotrypsin separately and the results were combined. No taxonomy or modifications were specified. Peptide mass tolerance is ±1.0Da, MS/MS tolerance is ±0.5Da, and searched for all possible charge states (i.e. +1, +2, +3). The final result was manually inspected to remove all false positives.

3.3 Results and Discussion

3.3.1 MALDI-TOF Pattern Analysis of GOG Bank Serum Samples

In previous study (see **Chapter II**), we have demonstrated separating and enriching low molecular weight proteins/peptides in human serum using an optimized centrifugal ultrafiltration method. Three marker peaks (1260Da, 1465Da, 1545Da, respectively) were found and used to distinguish between cancer and normal samples. In this project, we developed a new organic precipitation method to extract LMW species from human serum. To validate the efficacy of our method and the markers we found before, we used MALDI-TOF pattern analysis coupled with this new method to test the same set of samples from Gynecologic Oncology Group tissue bank (18 normal, 20 patients).

A typical MALDI-TOF pattern comparison of the interested mass range (1200-1600Da) was shown in **Figure 3.1**. As could be seen, there are three marker

peaks which could be used to distinguish cancer patient and normal control. 1260Da and 1465Da peaks were the same peaks found using centrifugal ultrafiltration method in previous study. Interestingly, the 1545Da peak disappeared in most of the samples, and a new 1535Da peak could be detected. If we use peak area ratio 1465Da/1260Da>1 as marker for cancer-positive, 14 out 18 normal samples could be correctly classified (77.8% specificity), and 1535Da peak was not detectable in most of the normal samples. 100% specificity could be achieved if using 1466Da/1260Da>2 as marker. In case of sensitivity, 16 of 20 patient samples could be correctly identified using either 1465Da/1260Da>1 or 1465Da/1260Da>2 as distinguishing threshold.

There were 3 patient samples (G239, G252, G502) with no signal detected within 1200-1600Da range using centrifugal ultrafiltration method in previous study. No signal was detectable in these samples using organic precipitation method either. Based on the results above, we could conclude that the organic precipitation method we developed could achieve similar results as in previous study. The specificity is even better when using optimized cut-off threshold (1465Da/1260Da>2).

According to the SDS-PAGE result from Chertov group [Chertov et al., 2004], protein precipitation using 50% acetonitrile could not completely separate low molecular weight species from their carrier proteins; a large amount of LMW species were still detectable in the precipitated fractions on denaturing gel. This observation indicated that the rapid denaturation of carrier proteins may result in co-precipitation of some LMW species bound to them. And only free-phase and less tightly bound

species could be sufficiently extracted in the supernatant. This could explain why we got very similar proteomic pattern between organic solvent precipitation method and non-denaturing centrifugal ultrafiltration method previously developed (data not shown).

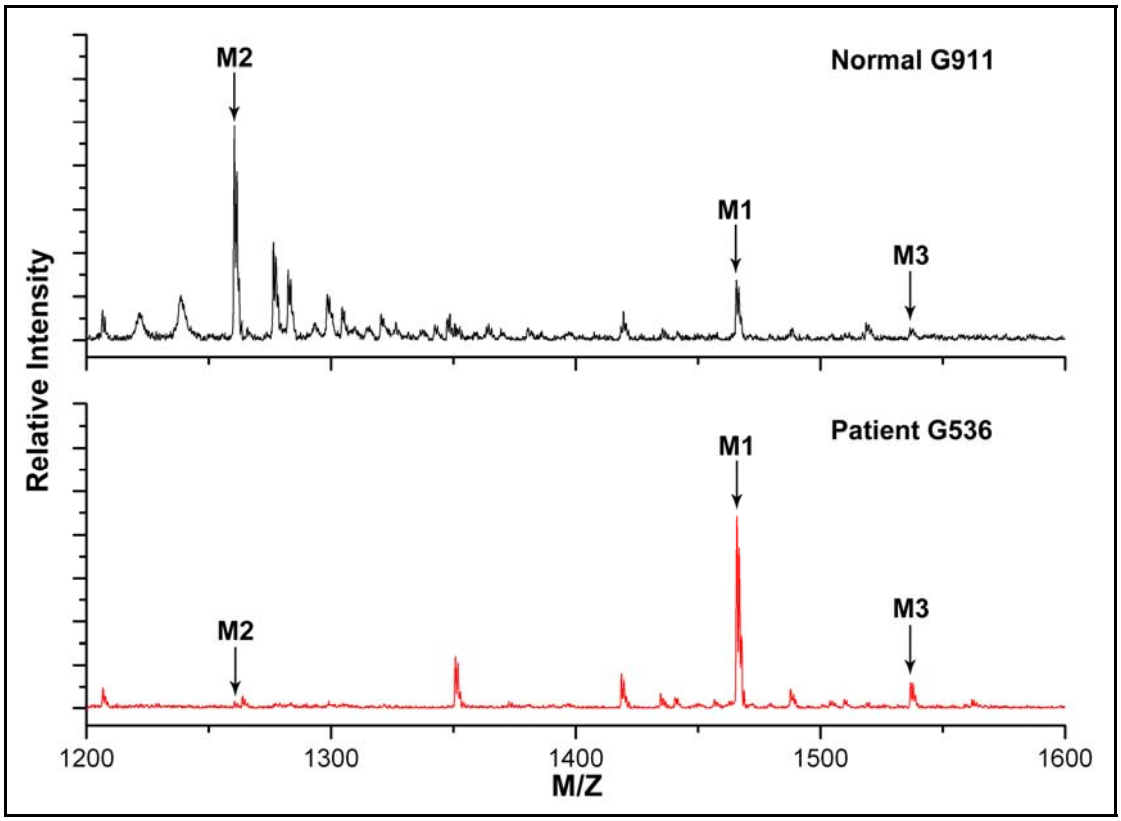
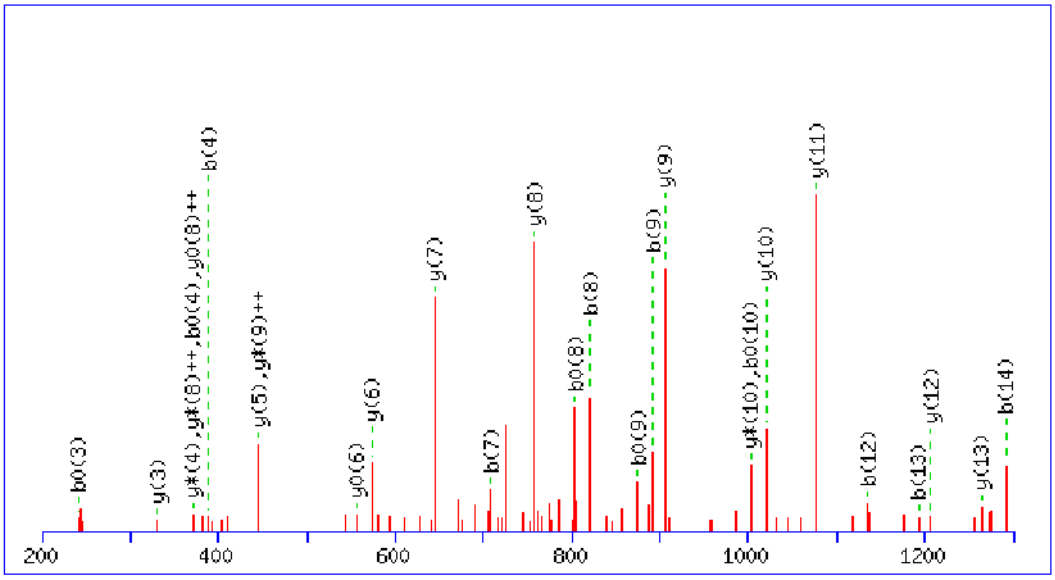


Figure 3.1 MALDI-TOF MS pattern analysis of serum samples from GOG bank.

3.3.2 Protein Identification Using LC/MS/MS and Database Searching

Since we could repeatedly detect several marker peaks using two different methods, the identity of those peaks has to be revealed. LC/MS/MS analysis and database searching were performed on tryptic digested normal and patient samples. The most important marker peak at 1465Da was identified as des-alanine-fibrinopeptide A (2-16, DSGEGDFLAEGGGVR, See **Figure 3.2**). And the result also indicated that intact fibrinopeptide A (1535Da) and its other fragments (1350Da, 1263Da, 1206Da) were corresponding to several major peaks detected within 1200-1600Da mass range in MALDI-TOF patterns. But none of these peaks had shown consistent correlation with ovarian cancer and therefore wasn't used as marker in our study. The intact fibrinopeptide A (1-16, ADSGEGDFLAEGGGVR, 1535Da) is much less abundant than its major fragment - 1465Da - in cancer patient samples and was not detectable in most of the normal samples (where 1465Da marker peak is also very weak). According to previous study [Profumo et al., 2005], des-alanine fibrinopeptide A (fpAY) should be a less abundant form compared with the original whole peptide after thrombin induced fibrinogen activation. But in our study, it is the major distinguishing peak and the relative quantity is much higher than the intact fibrinopeptide A and other fragments in benign and cancer patient samples. This may indicate some pathological change related to cancer development.



Mascot Search Results: Peptide View

Monoisotopic mass of neutral peptide Mr(calc): 1464.65
Ions Score: 85 Expect: 0.00013
Matches (Bold Red): 28/130 fragment ions using 31 most intense peaks

#	b	b^{++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	116.03	58.52	98.02	49.52	D							15
2	203.07	102.04	185.06	93.03	S	1350.63	675.82	1333.60	667.30	1332.62	666.81	14
3	260.09	130.55	**242.08**	121.54	G	**1263.60**	632.30	1246.57	623.79	1245.59	623.30	13
4	**389.13**	195.07	**371.12**	186.06	E	**1206.57**	603.79	1189.55	595.28	1188.56	594.79	12
5	446.15	223.58	428.14	214.57	G	**1077.53**	539.27	1060.51	530.76	1059.52	530.26	11
6	561.18	281.09	543.17	272.09	D	**1020.51**	510.76	**1003.48**	502.25	1002.50	501.75	10
7	**708.25**	354.63	690.24	345.62	F	**905.48**	453.25	888.46	**444.73**	887.47	444.24	9
8	**821.33**	411.17	**803.32**	402.16	L	**758.42**	379.71	741.39	**371.20**	740.40	**370.71**	8
9	**892.37**	446.69	**874.36**	437.68	A	**645.33**	323.17	628.30	314.66	627.32	314.16	7
10	1021.41	511.21	**1003.40**	502.20	E	**574.29**	287.65	557.27	279.14	**556.28**	278.65	6
11	1078.43	539.72	1060.42	530.71	G	**445.25**	223.13	428.23	214.62			5
12	**1135.45**	568.23	1117.44	559.23	G	388.23	194.62	**371.20**	186.11			4
13	**1192.48**	596.74	1174.46	587.74	G	**331.21**	166.11	314.18	157.59			3
14	**1291.54**	646.28	1273.53	637.27	V	274.19	137.60	257.16	129.08			2
15					R	175.12	88.06	158.09	79.55			1

Figure 3.2 LC/MS/MS analysis and database searching result for marker peak at 1465Da.

The identity of the 1260Da peak is still unknown even after LC/MS/MS analysis and database searching. Comparing MALDI-TOF patterns with or without deglycosylation indicated that the relative intensity of 1260Da peak decreased after deglycosylation in some samples (see **Figure 3.3**). This indicated that a N-linked oligosaccaride was removed after deglycosylation using PNGase F. Enzymatic deglycosylation treatment made the relative intensity of 1260Da peak more stable and reliable. It could serve as a good internal standard because it is not related to fibrinopeptide A and is less affected by the suppression effect generated from high levels of fibrinopeptide A fragments.

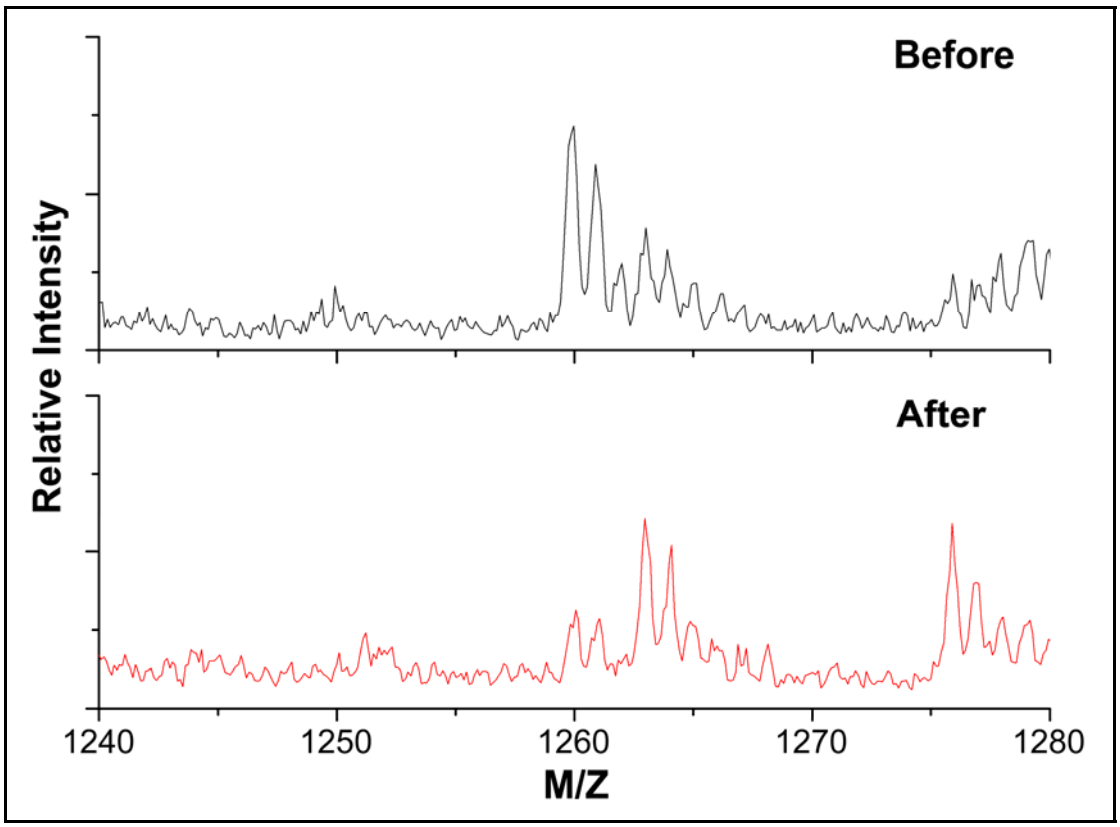


Figure 3.3 Relative intensity changes of 1260Da marker peak before and after enzymatic deglycosylation in some samples. It should be noted that the adjacent 1263Da peak is also a fragment of fibrinopeptide A.

3.3.3 MALDI-TOF Pattern Analysis of Other Sample Sets

Sample collection and preparation have proved to be essential for human serum proteomic research [Hsieh et al., 2006]. A standardized sampling procedure was suggested for proteome comparison and biomarker discovery. For samples collected years ago from different institutions, it is very hard to follow this recommendation. So, after validation of efficacy of the method and the identification of the marker peaks using the first GOG bank sample sets, we further tested the ovarian cancer marker (1465Da/1260Da >2) on other new sample sets.

The first sample set is from Cooperative Human Tissue Network – University of Pennsylvania Medical Center. This sample set consists of 19 normal serum samples. One of them was from male donor and therefore being excluded from this study. 15 out of 18 samples (~83%) didn't show any marker peak within the 1200-1600Da mass range using MALDI-TOF pattern analysis. Among the 3 samples which have detectable marker peaks, 2 of them can be classified as normal (1465Da/1260Da <2), only one sample showed false cancer-positive signal (1465Da/1260Da = 4.64). The discrepancy observed could be a result of different sample preparing procedure. But since low overall marker peak intensity is also an important observation in the first GOG tissue bank normal sample set (no marker peaks were detectable in one of those samples), we decided to add another distinguishing factor - no marker peaks detectable (or 1465Da/1260Da =1) - to our marker list for classifying normal samples. And with this 2 classification factors, we identified 35 out of 36 normal samples correctly.

The second sample set is from GOG tissue bank - Columbus Children Research Hospital. While the first set of GOG tissue bank samples we used were all collected in year 2003, this second sample set contains samples dated back to 1993 and most of them (41 out of 60) were collected before year 2000. This sample set contains 40 various carcinoma samples and 20 benign samples in 3 pathological categories. The MALDI-TOF pattern analysis resulted in a 97.5% (39 out of 40) sensitivity for carcinoma samples and 85% (17 out of 20) for benign samples when using 1465Da/1260Da >2 as marker. There was no significant difference between different types of benign and carcinoma samples. But the average level of marker value was notably elevated from benign to carcinoma. Two sample t-test also proved significant differences between normal, benign, and cancer sample sets (data not shown). Detailed statistical analysis was summarized in **Table 3.1** and **Table 3.2**.

Table 3.1 Statistical analysis of ovarian cancer marker: 1465Da/1260Da

	PA Normal	GOG Normal	Normal All	Benign	GOG Cancer 1	GOG Cancer 2	GOG Cancer All
Mean	1.15	0.60	0.87	13.74	56.64	49.95	52.18
SD* **	0.90	0.30	0.72	13.97	37.79	36.09	36.48
SE* **	0.21	0.07	0.12	3.12	8.45	5.71	4.71
Min	0.11	0.20	0.11	0.94	0.49	3.27	0.49
Max	4.64	1.10	4.64	57.59	100.00**	100.00**	100.00**
Median	1.00*	0.57	1.00*	8.65	49.28	40.74	41.90
N	18	18	36	20	20	40	60

* Marker ratio was manually set to "1" if no marker peaks detectable in normal samples

** Marker ratio was manually set to "100" if marker ratio is larger than 100 in cancer patient samples.

*** SD – Standard Deviation, SE – Standard Error

Table 3.2 Elevated marker level correlated to cancer development

	Normal	Benign	Carcinoma
1465Da/1260Da>2	1/36 (2.8%)	17/20 (85%)	55/60 (91.7%)
1465Da/1260Da>10	N/A	9/20 (45%)	49/60 (81.7%)
1465Da/1260Da>20	N/A	5/20 (25%)	40/60 (66.7%)

3.4 Conclusive Remarks

In this project, we have demonstrated a novel organic precipitation method coupled with enzymatic deglycosylation to extract cancer related markers from low molecular weight portion of human serum proteome. Several marker peaks were robustly detected and validated in different clinical sample sets. The identification of these markers was determined by tandem mass spectrometry coupled with database searching. Interestingly, most of the major peaks detected were derived from fibrinopeptide A – a product of thrombin induced fibrinogen activation.

Fibrinogen is the main protein of blood coagulation system. It is a large protein (MW 340 kDa) in human and it consists of two identical subunits that contain three polypeptide chains: α, β and γ. Fibrinopeptide A (1 – 16 amino acids) and B (1 – 17 amino acids) are released by thrombin from the N terminal parts of α- and β-chains, respectively. In this way fibrinogen is converted into fibrin, which forms a fibrin clot by polymerization [Maurer et al., 2000]. Pathological hemostasis activation often occurs in cancer patients [Wojtukiewicz et al., 2000]. And the presence of fibrin in and around tumors may be essential for tumor growth [Zacharski et al., 1990]. While there is no direct correlation between plasma FPA levels and plasma fibrinogen concentration, extravascular metabolism of fibrinogen and formation of fibrin in or around tumor may account for the leakage of FPA into the circulation [Wilner et al., 1978] and therefore responsible for abnormally elevated circulating FPA levels.

A correlation between elevated plasma level of FPA and ovarian cancer was reported previously [Gadducci et al., 1994]. Using competitive enzyme-linked

immunoassay (cELISA) to measure FPA levels, they have achieved very high specificity (100%) and fair sensitivity (64%) in distinguishing ovarian carcinoma patients from benign or controls. But they also concluded that there was no difference in FPA level between benign and controls. The antibody used in cELISA is able to capture all FPA related fragments as long as they still contain the recognition site. This will limit the differentiation power when there are metabolic fragments hanging around. In our study, the application of MALDI-TOF MS pattern analysis provided an accurate tool of measuring interested molecules. The real cancer related molecule could be well distinguished from others. Much higher sensitivity and the capability of distinguishing benign ovarian diseases from controls and cancer patients were two major benefits. The mechanism of high level des-alanine fibrinopeptide A and its correlation with ovarian cancer is still unknown. Orvisky and colleagues reported a down-regulated serum des-alanine FPA level in hepatocellular carcinoma [Orvisky et al., 2006]. This observation indicated that the level of circulating des-alanine fibrinopeptide A is cancer pathology specific and more research should be done to elucidate the mechanism.

As an additional procedure before MALDI-TOF pattern analysis, enzymatic deglycosylation proved to be critical for signal clean up and quantification. Peptide-N-glycosidase F (PNGase F) is one of the most widely used enzymes for the deglycosylation of glycoproteins. The enzyme releases asparagine-linked (N-linked) oligosaccharides from glycoproteins and glycopeptides, and the substrate could be as small as a tripeptide (Asn-Xaa-Ser/Thr, where X can be any amino acid, except Pro).

The removal of glycans was especially beneficial in normal samples which generate much less abundant signals and are readily to be affected by suppression effects (data not shown). The identity of 1260Da marker peak should be further studied, and it's highly probable to be a glycopeptide (see **Figure 3.3**).

As a conclusion, we have developed a novel organic solvent precipitation method for separating and enriching low molecular weight species from human serum. An enzymatic deglycosylation step was used before MALDI-TOF MS pattern analysis to improve data quality and marker quantification. Elevated circulating Des-alanine fibrinopeptide A level was found to have correlation with development of ovarian cancer. The whole procedure is simple and the results are exiting. Preoperative plasma may be a better sample system for validating markers found in this study. More clinical samples sets should be tested with standardized collecting and preserving protocols.

References

B-L Adam, Y. Qu, J.W. Davis, et al., “Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men”, Cancer Res., **62**:3609-3614 (2002)

J.N. Adkins, S.M. Varnum, K.J. Auberry, et al., “Toward a Human Blood Serum Proteome”, Molecular & Cellular Proteomics, **1**:947-955 (2002)

G. Alexe; S. Alexe; L.A. Liotta, et al., “Ovarian cancer detection by logical analysis of proteomic data”, Proteomics, **4**:766-783 (2004)

American Cancer Society, “Cancer facts and figures”, Altanta: American Cancer Society, (1996)

J. An, J-Y Sun, Q. Yuan, et al., “Proteomics analysis of differentially expressed metastasis-associated proteins in adenoid cystic carcinoma cell lines of human salivary gland”, Oral Oncology, **40**:400-408 (2004)

N.L. Anderson and N.G. Anderson, “The Human Plasma Proteome: History, Character, and Diagnostic Prospects”, Molecular & Cellular Proteomics, **1**:845-867 (2002)

J.M. Arthur, V. Thongboonkerd, J.A. Scherzer, et al., “Differential expression of proteins in renal cortex and medulla: A proteomic approach”, Kidney International, **62**:1314-1321 (2002)

L.L. Banez, P. Prasanna, L. Sun, et al., “Diagnostic potential of serum proteomic patterns in prostate cancer”, J. Urol., **170**:442-446 (2003)

R.E. Banks, M.J. Dunn, D.F. Hochstrasser, et al., “Proteomics: new perspectives, new biomedical opportunities”, Lancet **356**:1749-1756 (2000)

S. Beranova-Giorgianni, “Proteome analysis by two dimensional gel electrophoresis and mass spectrometry: strengths and limitations”, Trends in Analytical Chemistry, **22**, No. 5 (2003)

P. Chaurand, B.B. DaGue, R.S. Pearsall, et al., “Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry”, Proteomics, **1**:1320-1326 (2001)

S. Curry, “Beyond expansion: structural studies on the transport roles of human serum albumin”, Vox Sang, **83**, Suppl. 1:315-319 (2002)

E.P. Diamandis, “Mass Spectrometry as a Diagnostic and a Cancer Biomarker

Discovery Tool: Opportunities and Potential Limitations", Molecular & Cellular Proteomics, **3**:367-378 (2004)

D.B. Friedman, S. Hill, J.W. Keller, et al., "Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry", Proteomics, **4**:793-811 (2004)

A. Görg, C, Obermaier, G, Boguth, et al., "The current state of two-dimensional electrophoresis with immobilized pH gradients" Electrophoresis, **21**, No. 6:1037-1053 (2000)

D.J. Grossklaus, J.A. Smith, S.B. Shappell, et al., "The free/total prostate-specific antigen ratio is the best predictor of tumor involvement in the radical prostatectomy specimen among men with an elevated PSA", Urol. Oncol., **7(5)**:195-198 (2002)

A.E. Guppy and G.J. Rustin, "CA125 response: can it replace the traditional response criteria in ovarian cancer?", Onologist, **7(5)**:437-443 (2002)

S.P. Gygi, B. Rist, S.A. Gerber, et al., "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags", Nat. Biotechnol., **17**:994-999 (1999)

R.L. Hamler, K. Zhu, N.S. Buchanan, et al., "A two-dimensional liquid-phase

separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification", Proteomics, **4**:562-577 (2004)

P.A. Haynes, S.P. Gygi, D. Figeys, et al., "Proteome analysis: biological assay or data archive?", Electrophoresis **19**:1862–1871 (1998)

D.F. Hochstrasser, "Proteome in perspective", Clin Chem Lab Med, **36**:825–836 (1998)

T.W. Hutchens and T.T. Yip, "New desorption strategies for the mass spectrometric analysis of macromolecules", Rapid Commun. Mass Spectrom., **7**:576-580 (1993)

T.C. Hunter, N.L. Andon, A. Koller, et al., "The functional proteomics toolbox: methods and applications", Journal of Chromatography B, **782**:165-181 (2002)

D. Jiang, W. Ying, Y. Lu, et al., "Identification of metastasis-associated proteins by proteomic analysis and functional exploration of interleukin-18 in metastasis". Proteomics, **3**:724-737 (2003)

J. Klose and U. Kobalz, "Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome", Electrophoresis, **16(6)**:1034-1059 (1995)

K.R. Kozak, M.W. Amneus, S.M. Pusey, et al., “Identification of biomarkers for ovarian cancer using strong anion-exchange Proteinchips: Potential use in diagnosis and prognosis”, PNAS, **100**:12343-12348 (2003)

K. Kubota, K. Wakabayashi and T. Matsuoka, “Proteome analysis of secreted proteins during osteoclast differentiation using two different methods: 2D-GE and ICAT analysis with 2D HPLC”, Proteomics, **3**:616-626 (2003).

S. Lehrer, J. Roboz, H. Ding, et al., “Putative protein markers in the sera of men with prostatic neoplasms”, B.J.U.Intl., **92**:223-225 (2003)

Andrew J. Link, Jimmy Eng, David M. Schieltz et al., “Direct analysis of protein complexes using mass spectrometry”, Nature Biotechnol., **17**:676-682 (1999)

R.R.O. Loo, T.I. Stevenson, C. Mitchell, et al., “Mass spectrometry of proteins directly from polyacrylamide gels”, Anal Chem, **8**:1910–1917 (1996)

K.L. Meehan and M.D. Sadar, “Quantitative profiling of LNCaP prostate cancer cells using isotope-coded affinity tags and mass spectrometry”, Proteomics, **4**:1116-1134 (2004)

M.P. Molloy, S. Donohoe, E.E. Brzezinski, et al., “Large-scale evaluation of

quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling", Proteomics, **5**:1204–1208 (2005)

R.S. Negm, M. Verma and S. Srivastava, "The promise of biomarkers in cancer screening and detection", Trends in Molecular Medicine, **6(6)**:288-293 (2002)

W.F. Patton, "Detection technologies in proteome analysis", Journal of Chromatography B, **771**:3-31 (2002)

E.F. Petricoin III, A. Ardekani , B. Hitt, et al., "Use of proteomic patterns in serum to identify ovarian cancer", Lancet, **359**:572-577 (2002)

E.F. Petricoin III, D.K. Ornstein, C.P. Paweletz, et al., "Serum proteomic patterns for detection of prostate cancer", J. Natl. Cancer Inst., **94**:1576-1578 (2002)

E.F. Petricoin and L.A. Liotta, "SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer", Current Opinion in Biotechnology, **15**:24-30 (2004)

E. Petricoin, J. Wulfkuhle, V. Espina, et al., "Clinical Proteomics: Revolutionizing Disease Detection and Patient Tailoring Threapy", Journal of Proteome Research,

3:209-217 (2004)

Y. Qu, B-L Adam, Y. Yasui, et al., "Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminated prostate cancer from nonprostate patients", Clin. Chem., **48**:1835-1843 (2002)

T. Reynolds, "For proteomics research, a new race has begun", J Natl Cancer Inst, **94**:552-554 (2002)

A.K. Sato, D.J. Sexton, L.A. Morganelli, et al., "Development of mammalian serum albumin affinity purification media by peptide phage display", Biotechno. Prog., **18**:182-192 (2002)

B. Schulenberg, J.M. Beechem and W.F. Patton, "Mapping glycosylation changes related to cancer using the Multiplexed Proteomics technology: a protein differential display approach", Journal of Chromatography B, **793**:127-139 (2003)

Y. Shen, R. Zhao, S.J. Berger, et al., "High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics", Anal. Chem., **74**:4235-4249 (2002)

Y. Shen, J.M. Jacobs, D.G. Camp, et al., "Ultra-high-efficiency strong cation exchange

LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome", Anal. Chem., **76**:1134-1144 (2004)

A. Shevchenko, M. Wilm, O. Vorm, et al., "Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels" Analytical Chemistry, **68**:850- 858 (1996)

M.B. Smolka, H. Zhou, S. Purkayastha, et al., "Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis", Analytical Chemistry, **297**:25-31 (2001)

P.R. Srinivas, M. Verma, Y. Zhao, et al., "Proteomics for Cancer Biomarker Discovery", Clin. Chem., **48(8)**:1160-1169 (2002)

T. Stasyk and L.A. Huber. "Zooming in: Fractionation strategies in proteomics", Proteomics, **4**:3704–3716 (2004)

N. Tang, P. Tornatore and S.R. Weinberger, "Current developments in SELDI affinity technology", Mass Spectrometry Reviews, **23**:34-44 (2004)

W.A. Tao and R. Aebersold, "Advances in quantitative proteomics via stable isotope tagging and mass spectrometry", Current Opinion in Biotechnology, **14**:110-118

(2003)

R.S. Tirumalai, K.C. Chan, D-R.A. Prieto, et al., “Characterization of the low molecular weight human serum proteome”, Molecular & Cellular Proteomics, **2(10)**:1096-1103 (2003)

M. Verma, G.L. Wright, Jr, S.M. Hanash, et al., “Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers”, Ann N Y Acad Sci, **945**:103-115 (2001)

R. Vitorino, M.J.C. Lobo, A.J. Ferrer-Correira, et al., “Identification of human whole saliva protein components using proteomics”, Proteomics, **4**:1109-1115 (2004)

Y. Wagner, A. Sickmann, H.E. Meyer, et al., “Multidimensional Nano-HPLC for Analysis of Protein Complexes”, J. Am. Soc. Mass Spectrom., **14**:1003-1011 (2003)

M.P. Washburn, D. Wolters and J.R. Yates, III, “Large Scale analysis of the yeast proteome via multidimensional protein identification technology”, Nat. Biotechnol., **19**:242-247 (2001)

V.C. Wasinger, S.J. Cordwell, A. Cerpa-Potjak, et al., “Progress with gene-product mapping of the Mollicutes: mycoplasma genitalium” Electrophoresis, **16**:1090–1094

(1995)

M.R. Wilkins, J.C. Sanchez, A.A. Gooley, et al., “Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it”, Biotechnol Genet Eng Rev, **13**:19–50 (1996)

D.A. Wolters, M.P. Washburn and J.R. Yates, III, “An automated Multidimensional Protein Identification Technology for Shotgun Proteomics”, Anal. Chem., **73**:5683-5690 (2001)

J.D. Wulfkuhle, L.A. Liotta and E.F. Petricoin, “Proteomic Applications for the Early Detection of Cancer”, Nature Rev., Cancer, **3**: 267-275 (2003)

H. Zhang, W. Yan and R. Aebersold, “Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes”, Current Opinion in Chemical Biology, **8**:66-75 (2004)

H. Zhou, J.A. Ranish, J.D. Watts, et al., “Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry”, Nat. Biotechnol., **20**: 512-515 (2002)

PART II

TECHNOLOGY DEVELOPMENT FOR GENOMIC BIOMARKER ANALYSIS

CHAPTER IV

GENERAL INTRODUCTION

4.1 Single Nucleotide Polymorphisms (SNPs) Detection

4.1.1 What are Single Nucleotide Polymorphisms (SNPs)?

Single Nucleotide Polymorphisms (SNPs) are single base substitutions that occur at a specific position in a genome (**Figure 4.1.1a**). Single-base insertions and deletions are not strictly SNPs, but they are often included under this genetic definition (**Figure 4.1.1b**). In general, the rare (less frequent) allele has an abundance of 1% or greater in a population, otherwise it is referred to as point mutation [Brookes, 1999]. In diploid species as humans, SNPs are usually biallelic (a SNP has two alleles and three or four alleles are very rare). Averagely speaking, one SNP could be found every 300 to 1000 bases in humans. They are estimated to represent as much as 90% of all genetic variations [Cooper et al., 1985].

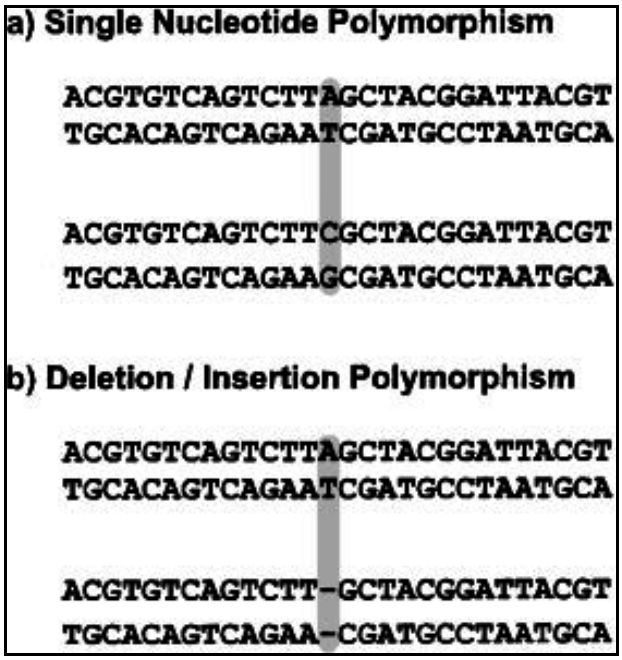


Figure 4.1.1 (a) The two alleles of a single nucleotide polymorphism; (b) a single nucleotide deletion/insertion.

1.42 million SNPs were reported [Sachidanandam et al., 2001] and now there are more than 6 million mapped human SNPs in public databases [Collins et al., 2003]. Only a small portion of located SNPs lie within coding regions (cSNPs) and an even smaller percentage will result in sequence changes in expressed proteins. However, non-coding SNPs can still influence mRNA stability and conformation, as well as the quantity and quality of gene expression product. To geneticists, it is highly valuable if a marker can be pinpointed directly in the genomic region of interest. An allele of a SNP can be an important genetic risk factor as it may increase susceptibility to certain diseases. Due to their biallelic nature, SNPs are fairly easy to detect in genotyping and data interpretation can be automated.

4.1.2 Genotyping SNPs Using MALDI-TOF Mass Spectrometry

Numerous approaches have been developed in genotyping SNPs. Criteria such as accuracy of SNP detection, sensitivity to identify SNPs using a small amount of template, throughput, and flexibility of the procedure were thoroughly explored. Traditional methods based on molecular biology techniques often lack essential features for clinical settings such as accuracy, automation, and throughput. In terms of robustness, accuracy, reproducibility, and success rates, mass spectrometry assays based on primer extension reactions surpassed most other methods in side by side comparison [Le Hellard et al., 2002].

All methods for SNP genotyping combine two fundamental steps: the generation of an allele-specific product, and the analysis thereafter. Currently, four

major methodologies are used for the generation of the allele-specific products: hybridization, ligation, cleavage, and primer extension. All of these methods for allele discrimination have been combined with MALDI-TOF analysis [Tost and Gut, 2005]. The most widely applied method for allele distinction in SNP genotyping and point mutation analysis is primer extension [Syvänen et al., 1990]. Several types of primer extension assays using MALDI-TOF MS as the detection platform have been developed. Differentiation of the alleles is basically achieved by three different strategies: single base primer extension (SBE), multiple base primer extension, and nucleotides depletion. Each extension product of a SNP is detected based on their mass differences. In single base primer extension, this is due to the addition of nucleotides that naturally differ in mass, and in multiple base primer extension, the addition of a different number of bases will distinct alleles.

For single base primer extension (SBE) reaction, the primer anneals immediately next to the polymorphic site and terminating dideoxynucleotides (ddNTPs) were used in extension reaction. Therefore, unlike in regular PCR reaction, the primer is only extended with a single nucleotide. This approach is applied in the PinPoint assay [Haff and Smirnov, 1997], and commercialized by Applied Biosystems as the Sequazyme^{TM} PinPoint SNP Typing Kit. The reaction procedure is outlined in **Figure 4.1.2**.

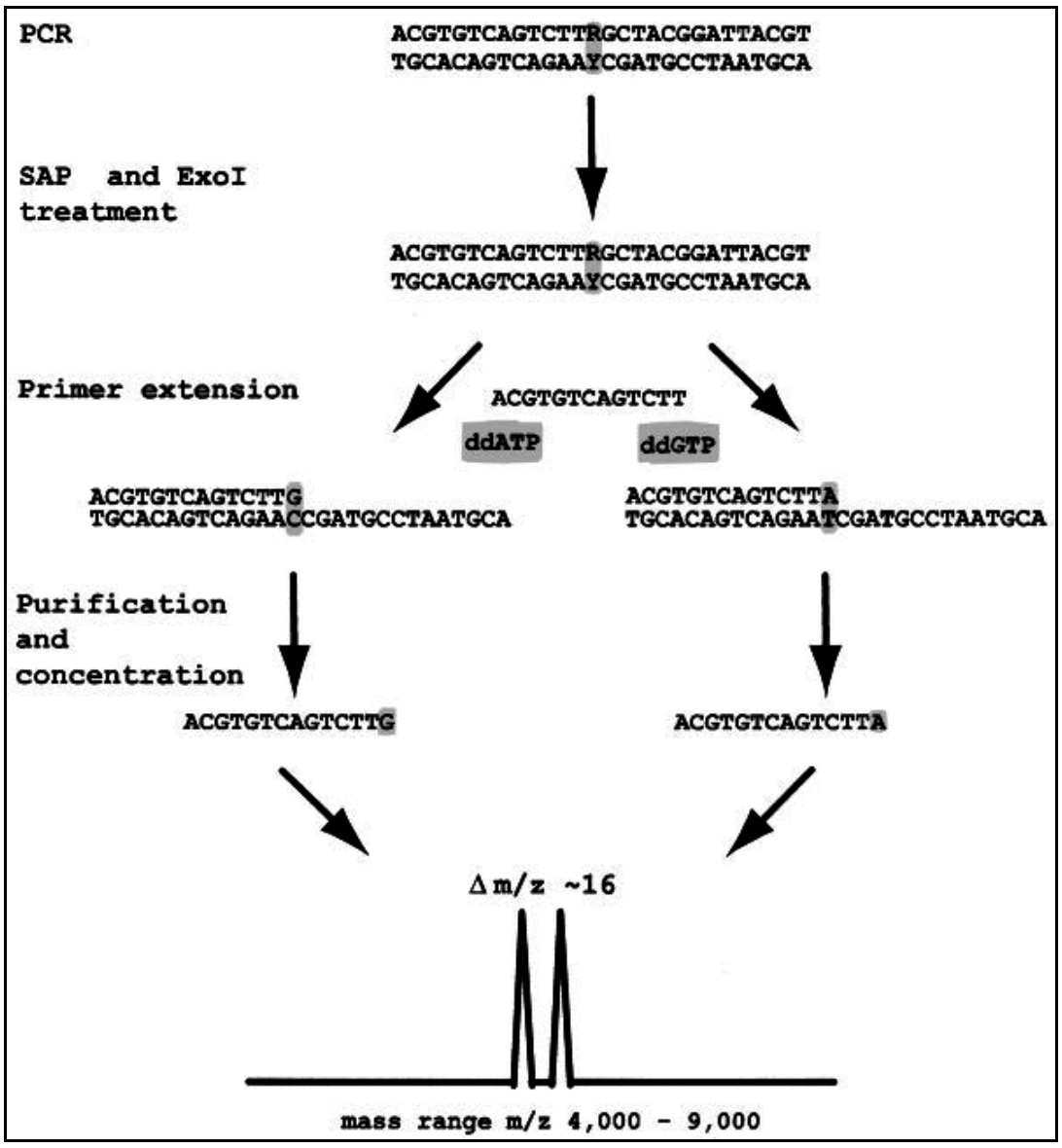


Figure 4.1.2 The reaction procedure of the PinPoint assay [Haff and Smirnov, 1997].

The GOOD assay was introduced by Sauer and colleagues [Sauer et al., 2000a,b]. Utilizing a chemical modification strategy – charge tagging [Gut et al., 1997], this assay doesn't require purification step prior to MALDI analysis. It reduces the extension products to a core-sequence of four to five nucleotides that contain the allele information by enzymatic digestion and could increase signal intensity to up to 100 fold.

Other methods like the solid phase capturable single base extension (SPC-SBE) assay [Kim et al., 2002] and the genoSNIP assay [Wenzel et al., 2003] are also based on single base primer extension principle. The SPC-SBE assay distinguishes itself from other assays by using biotinylated ddNTPs and only extended products are retained on target. Non-extended primers were eliminated and therefore increased the possibility of multiplexing. The genoSNIP assay is characterized by using photocleavable primers.

Since the smallest mass difference between two nucleotides (A and T) is only 9 Da, analysis of some SNPs may be difficult by SBE assays. To enhance the allele distinction in the mass spectrum, a mixture of deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) were used in multiple base primer extension principle. The primer anneals again immediately adjacent to a SNP site, one allele is extended by only one base (ddNTPs) while the second one by two or more (termination occurs at the first nucleotide in the template complementary to a ddNTP). This strategy could generate extension products with a mass difference of at least 300 Da. With enhanced allele discrimination, this concept is used in the primer oligo base extension (PROBE)

assay [Braun et al., 1997], the MassEXTEND assay [Leushner and Chiu, 2000] (Sequenom), and the very short extension (VSET) assay [Sun et al., 2000]. A schematic representation of PROBE assay is shown in **Figure 4.1.3**.

The MassEXTEND assay is an improved version of the original PROBE assay. It uses ion-exchange resin for sample conditioning, avoided biotinylated primers and immobilization as used in the PROBE assay. Extremely small volumes (nanoliter) of samples are transferred onto silicon target pre-spotted with 3-hydroxypicolinic acid (HPA) matrix. As the entire sample spot is volatilized with a few laser pulses, the need for searching sweet spots of the HPA matrix is avoided.

The very short extension (VSET) assay was designed to address some of the problems associated with both the original PinPoint and the PROBE assay. In contrast to the original PROBE assay, three ddNTP and one dNTP were used in primer extension reaction. As a result, for one allele, extension is one base and for the other are two bases, which separates the alleles by a mass of about 300 Da.

Nucleotide depletion genotyping assay (NUDGE) is essentially a variant of multiple base primer extension strategy. This method was reported by Decode Genetics recently [Blondal et al., 2003] and the primer is designed to anneal two bases upstream of the polymorphic site. The allele is discriminated in the presence of only three deoxynucleotides. The one that is complementary to one allele of the SNP is depleted. Therefore, termination occurs right next to the polymorphic position for one allele, while for the other allele, extension will not stop until the first nucleobase in the template complementary to the depleted dNTP.

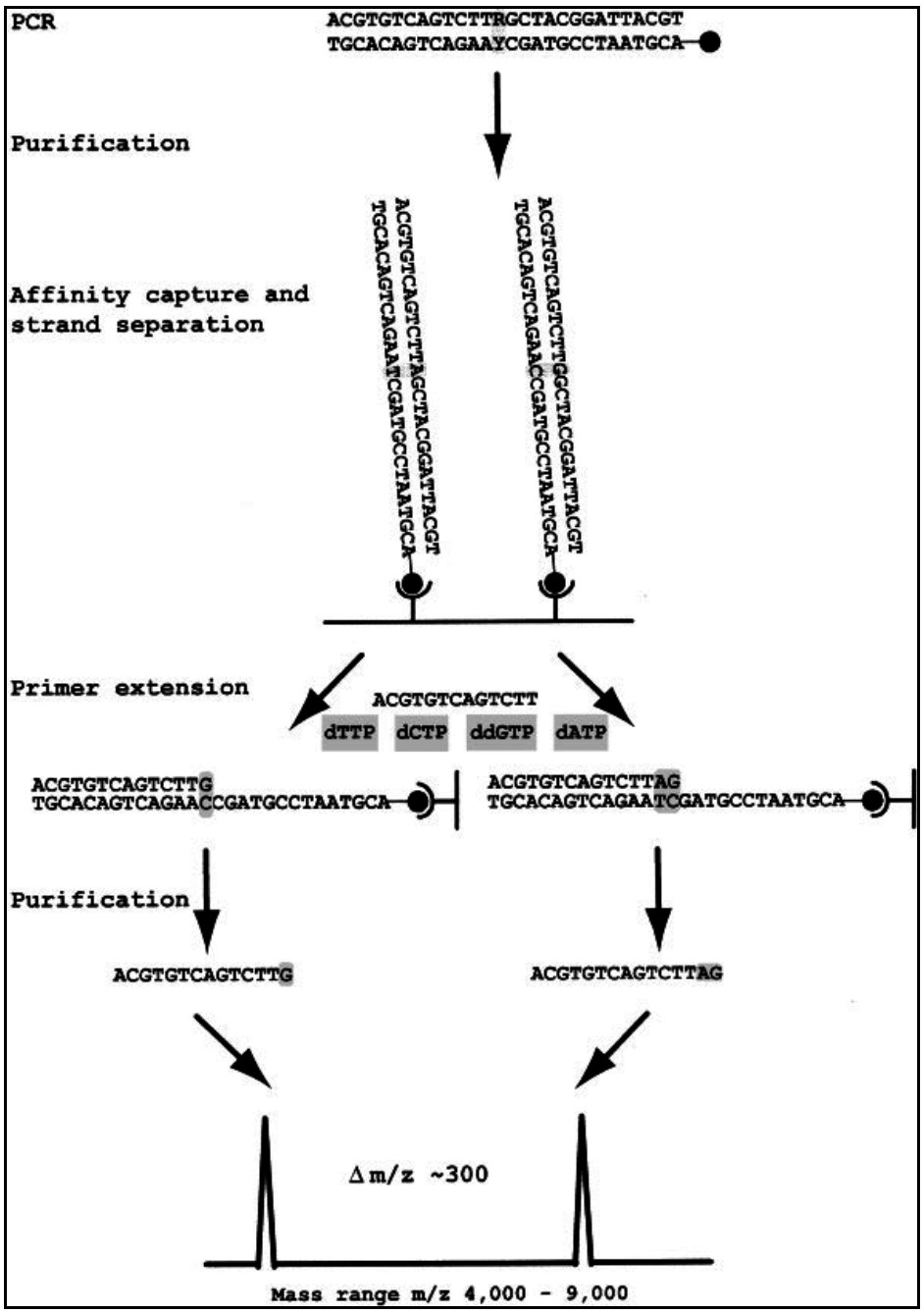


Figure 4.1.3 The PROBE assay [Braun, Little and Köster, 1997]. The biotin-label is shown as a black circle, which is captured in streptavidin-coated microtiter plates.

However, analysis of nucleic acids by MALDI mass spectrometry has some drawbacks. MALDI-TOF mass spectrometry is very sensitive to the presence of metal ions. Due to the multiple negatively charged sugar phosphate backbone of DNA molecule, it is susceptible to form adduct with common cations such as Na^+ and K^+. Peak broadening and reduction of resolution, sensitivity and accuracy are direct consequences of adduct formation. As cations such as sodium and potassium are highly abundant in buffers for molecular biology reactions, stringent purification procedures need to be applied. Approaches to overcome adduct formation are mostly based on ion-exchange principles. This could be done either on solid-phase [Tang et al., 1995] or through cation-exchange resin methods [Nordhoff et al., 1992]. The addition of ammonium containing organic modifiers to the matrix reduces heterogeneity of oligonucleotide ions caused by cation adducts. The ammonium is volatile and could be dissociated from the analyte molecules during laser pulse bombardment, and therefore leaving the analyte in free acid form. Diammonium citrate and tartrate are reported to be good choices serving this purpose [Pieles et al., 1993; Lecchi and Pannell, 1995].

The other drawback for MALDI-TOF mass spectrometry lies in the sample deposition method. The standard method for sample preparation is called dried-droplet method. In this method, a small drop (usually about 0.5μL) of sample solution is applied on a drop of prepared matrix solution (usually about 1μL) and allowing the mixture to dry. During the crystallization of matrix, analyte molecules were also redistributed and “sweet spots” were formed. The existence of

heterogeneous “sweet spots” will reduce the sensitivity and accuracy when applying MALDI-TOF in SNP analysis. Nanoliter liquid deposition technology was used to overcome “sweet spots” formation, but this will greatly limit the quantity of analyte loaded.

Choices of matrices, combination of sample conditioning prior to mass spectrometry analysis and sample preparation are all essential components of successful analysis of SNPs using MALDI-TOF mass spectrometry. Novel strategy is always needed and being continuously developed to improve the sensitivity, accuracy, and cost effectiveness of mass spectrometry based genotyping.

4.2 Microsatellite Instability Detection

4.2.1 Microsatellite Instability and Cancer

Microsatellites are one of the most abundant classes of intergenic repetitive sequences dispersed in the eukaryotic genome and contain repetitive units composed of one to five base pairs [Weber et al., 1989; Litt et al., 1989; Dib et al., 1996]. These sequences, as well as minisatellite sequences, are highly polymorphic in human populations and could be used as a marker for human identification or pedigree analyses [Hearne et al., 1992]. During the progress of the Human Genome Project, numerous microsatellite loci have been pinpointed on the human genome and are now being used for various genomic analyses.

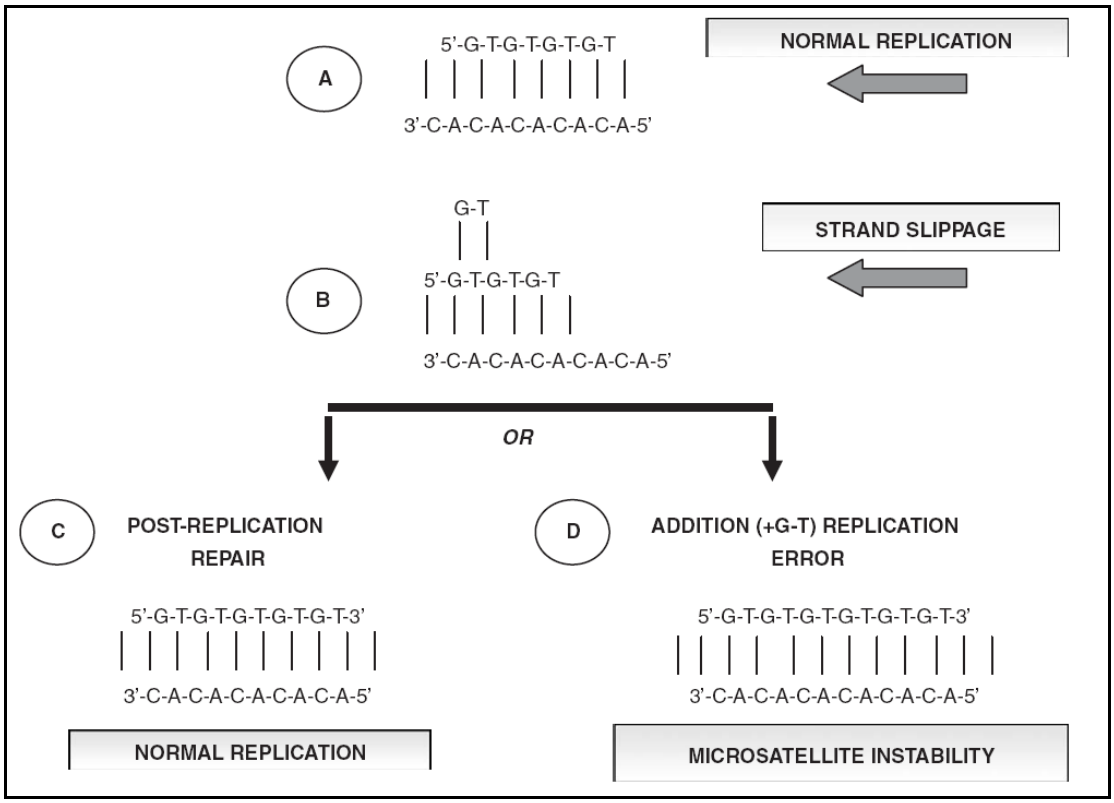


Figure 4.2.1 (A) Normal replication event of MS DNA; (B) strand slippage event initiating an abnormal replication; (C)effective repair leading to normal replication (A=C); (D) defective repair of B leading to microsatellite instability (A≠D). [Samara et al., 2006]

Compared with other regions on the genome, these repetitive sequences were thought to be stable within the life span of organisms. However, microsatellites are mainly located in non-coding regions (introns) of genome. Due to their repetitive characteristics, they are likely to be the site of insertion and deletion mutations during DNA replication, with an error rate of 1 in 100,000 base pairs copied. The majority of these errors could be detected and corrected by the inherent proof reading capabilities of DNA polymerases. A small proportion of mismatches that are missed by this mechanism are normally identified and corrected by a group of proteins named the mismatch repair (MMR) system [Fishel, 1999]. Dysfunction of the MMR system allows the accumulation of replication errors and leads to microsatellite instability [Strand et al., 1993], which could be defined as "a change of any length due to either insertion or deletion mutations of repeating units in a microsatellite" [Boland et al., 1998]. A representative scheme of MSI is shown in **Figure 4.2.1**.

Microsatellite alterations can be found in a number of tumors. There are two types of alterations: loss of heterozygosity (LOH), which can be detected in the majority of colorectal cancers (CRC), and microsatellite instability (MSI). Instability of a class of microsatellites with dinucleotide repeats was initially reported in colorectal cancer [Thibodeau et al. 1993; Ionov et al., 1993; Aaltonen et al., 1993] and in patients with hereditary non-polyposis colorectal cancer (HNPCC) [Peltomaki et al., 1993]. Because mutations in genes functioning in DNA mismatch repair were found to be inherited in pedigrees of HNPCC, microsatellite instability was regarded as an important phenotype of deficiency in DNA mismatch repair and as a marker of a high

risk for cancer. High-degree microsatellite instability (MSI-H) is a feature of most colorectal cancers arising as part of the hereditary non-polyposis colorectal cancer (HNPCC) as well as 15–20% of sporadic colorectal cancers [Haydon et al., 2002]. And it was reported that all colorectal cancers display some degree of microsatellite instability if enough markers are studied [Umar et al., 2004].

As a potential marker, MSI were also investigated in other cancers. The results are rather contradictory, though. For example, Mao and colleagues [Mao et al., 1994; 1996] observed high frequency microsatellite instability in transitional cell carcinoma (TCC) of bladder with both tumor DNA and DNA extracted from urine. But other researches demonstrated much lower frequency, or lack of correlation in TCC pathological data with MSI [Catto et al., 2004].

Siah and colleagues reported low detection rate of MSI in early-onset breast cancer patients [Siah et al.; 2000]. Two of the most recommended MSI marker, BAT-25 and BAT-26, failed to show any MSI in tumor DNA. While Wild and Murata all reported successful detection of MSI with a relatively higher detection rate [Wild et al., 2000; Murata et al., 2002].

The discrepancy of results could be greatly related to the experimental design. Choices of markers, sources and handling of clinical samples, researchers' standard of MSI categorization and MSI detection methods can all be distinguishing factors among studies. More work should be done to develop accurate and objective assay systems so that the clinicopathologic features of microsatellite instability could be revealed.

4.2.2 Microsatellite Instability Detection

4.2.2.1 Markers for Detecting Microsatellite Instability

In 1997, an International Collaborative Group announced guidelines for the identification and assessment of microsatellite instability [Boland et al., 1998]. They classified tumors into microsatellite stable (MSS) and two groups of microsatellite unstable cancers based on their degree of instability: high (MSI-H) or low (MSI-L). BAT-25 and BAT-26, two mononucleotide microsatellite markers, and three dinucleotide markers (D2S123, D5S346 and D17S250) were chosen as a panel to be studied for determining MSI status. Tumors displaying instability in two or more markers were designated MSI-H, instability in only one marker was designated MSI-L and if no markers were affected the tumor was designated MSS. There is now a growing consensus that the pathological phenotype of MSI-H is characterized particularly by instability in the mononucleotide markers [Umar et al., 2004]. Colorectal cancers with MSI-H demonstrated unique biological behavior while MSI-L cancers are less well characterized and are often grouped with cancers that are categorized as MSS [Gonzalez-Garcia et al., 2000].

More and more markers were tested in different research approaches. It was generally proved that the detection efficiency could be improved using expanded marker panels. And besides the mono- and dinucleotide markers originally defined on the NCI workshop, markers with longer sequence instability (tetranucleotide repeats, for example) demonstrated some advantages in cancer detection [Catto et al., 2003]. The hMLH1 promoter region was reported to be highly methylated in about 80% of

microsatellite instability MSI (+) colorectal cancers, but in none of the MSI (-) colorectal cancers [Masato Maekawa et al., 1999]. The correlation between hMLH1 gene hypermethylation and MSI was also reported in T cell lymphoma [Scarisbrick et al., 2003]. These findings could be a good lead to the pathological features behind MSI.

4.2.2.2 Methods for Microsatellite Instability Detection

Usually, a set of microsatellite markers is amplified by PCR followed by gel or capillary electrophoresis to separate PCR amplicons. Various detection method have been applied to illustrate the markers amplified, such as autoradiography [Thibodeau et al, 1993], silver staining [Schlegel et al, 1996], or fluorescence techniques [Gyapay et al, 1996]. The representative results using different methods are shown in **Figure 4.2.2 and Figure 4.2.3**.

Autoradiography method is using radiolabeled (^{32}P) oligonucleotide primers during polymerase chain reaction (PCR). The amplified products are electrophoresed on a sequencing gel and visualized using X-ray developed films. Silver stain is an alternative band visualization method for polyacrylamide gel electrophoresis (PAGE) without involving radioactive isotopes. Both methods suffer the known disadvantages of PAGE, such as migration error, low sensitivity and cumbersome procedures. Newly developed fluorescence based sequencing method could overcome all problems mentioned above [Oda, et al., 1997].

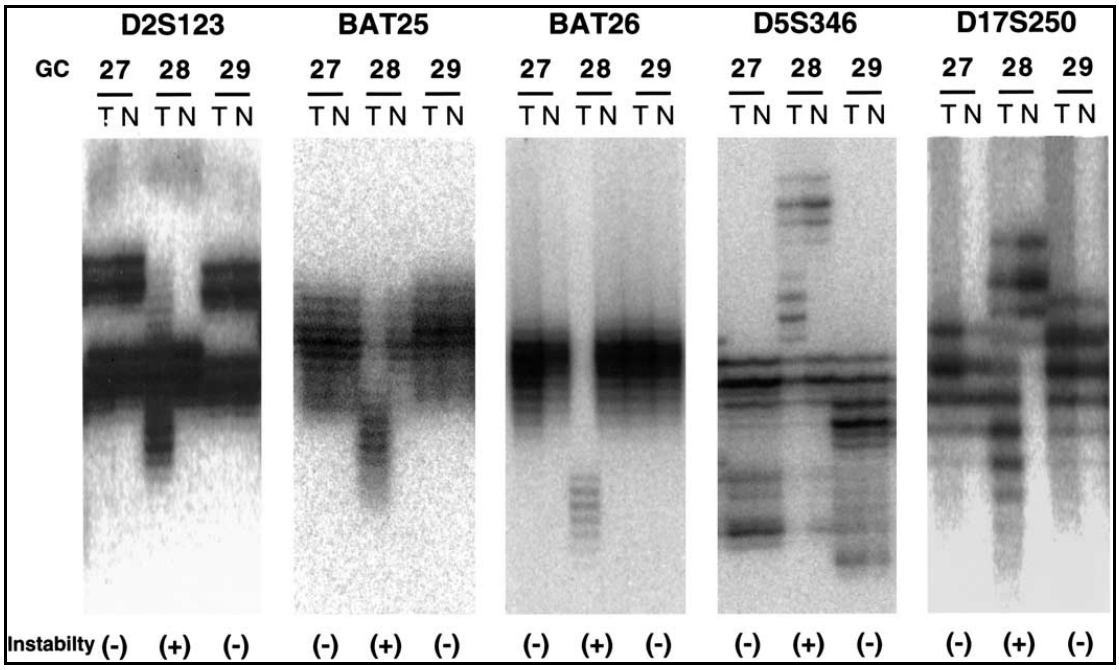


Figure 4.2.2 A representative example (case #28) of microsatellite instability in gastric cancers (radioactive electrophoresis method). T: tumor samples; and N: corresponding normal samples. Microsatellite markers are shown on the top of the gels. Bands of abnormal size are detected in all tumor lanes of case #28 [Yamada et al., 2002].

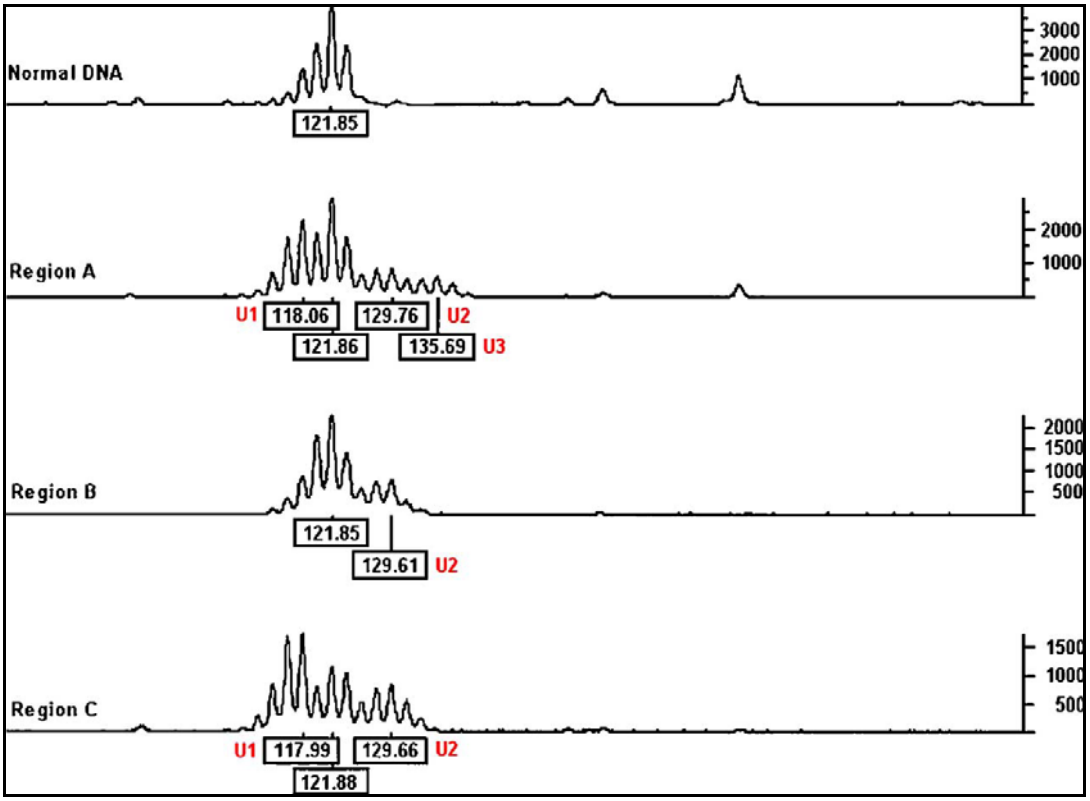


Figure 4.2.3 Variation of MSI profile among different regions in MSI tumors. Amplification of D2S123 microsatellite in 3 different regions of case 109. This is a homozygote marker with 1 normal allele of 121 bp. Three unstable alleles are amplified in region A, only 1 in region B, and 2 in region C [Danjoux et al., 2006].

The separation power of capillary electrophoresis combined with high throughput automated instrumental design could make fluorescence based sequencing a standard routine for MSI analysis.

Other than the widely adapted strategies, researchers have applied newly developed technologies in MSI analysis and cancer detection. Dietmaier and colleagues have established a technique to detect MSI by LightCycler Real-time PCR and melting point analysis [Dietmaier et al., 2001]. Amplification of microsatellites by real-time PCR is followed by melting point analysis to display alterations in the length of repetitive sequences. According to their results, amplification and melting point determination of BAT26 and BAT25 was achieved in 129/162 (80%) and 123/162 (76%) formalin-fixed and paraffin-embedded tissue samples, respectively. Moreover, MSI could be detected only in MSI-high tumors using both BAT25 and BAT26 markers. This new technique allows MSI detection within 1 hour and provides an alternative way for fast, high-throughput MSI analysis.

Denaturing high-performance liquid chromatography (DHPLC) was compared side by side with fluorescence capillary electrophoresis (FCE) in MSI detection [Deschoolmeester et al., 2006]. Developed by a Korean group in 2003 [Kim et al., 2003], this method was proved to be as effective as FCE on quasimonomorphic BAT markers, but hard to interpret for the detection of dinucleotide markers.

Clinicopathologic features of microsatellite instability positive cases have remained unclear. Introducing automated sequencer for microsatellite assay was a breakthrough for MSI analysis and cancer detection. The use of dual

fluorescence-labeled primers and an automated sequencer is recommended for accurate detection. However, to improve MSI detection using highly complicated clinical samples other than malignant tissues, novel sample preparation and PCR strategy has to be developed.

CHAPTER V

MATRIX-ASSISTED LASER DESORPTION/IONIZATION TIME-OF-FLIGHT ANALYSIS OF LOW-CONCENTRATION OF OLIGONUCLEOTIDES AND MINI-SEQUENCING PRODUCTS

5.1 Introduction

The genotyping method based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is one of the most promising approaches to large scale genotyping of single nucleotide polymorphisms (SNPs) [Braun et al., 1997a,b; Haff et al., 1997a,b; Hihhins et al., 1997; Little et al., 1997a,b,c; Fei et al., 1998; Fu et al., 1998; Ross et al., 1998; Chen et al., 1999; Griffin et al., 1999; Li et al., 1999; Sun et al., 2000], as MALDI-TOF is fast, provides higher accuracy in molecular weight data, and requires no fluorescent or radioactive labeling. However, these advantages are somewhat not match well to mini-sequencing, the

most widely used sequencing reaction employed in genotyping. This is evident by the fact that the overall detection limit of a MALDI-TOF based genotyping method is significantly poorer than that of fluorescent-based detection, though MALDI-TOF itself is extremely sensitive. The poor detection limit leads to a significant increase in genotyping costs and difficulty in designing multiplexed assays.

This mismatch is due to the fact that, unlike other methods, MALDI utilizes a very small portion of sequencing products for data acquisition. In MALDI, each laser shot targets a spot of ~100μm^2. In general the MALDI sensitivity is, to a large extent, dictated by the amounts of the analyte present on the targeted spot, rather than total quantities loaded. In other words, the more analytes that are present on the targeted spot, the higher the sensitivity will be. Conventional methods use a metal sample substrate. The aqueous solution of 3-HPA/DNA tends to spread to a large area on the metal surface as a result of strong interactions between metal and polar solvents. For example 0.5μL of aqueous solution can spread to an area of 1.5 mm in diameter. The formation of a large-area sample inevitably reduces the oligonucleotide quantity in a targeted spot, thus reducing sensitivity. Clearly, an ideal sample preparation method should deposit most, if not all, of the mini-sequencing products onto a target, and more importantly it should produce the smallest possible sample area so that the deposited molecules can be maximally assessed by the same laser shot. Other essential requirements for a good sample preparation method include good sample homogeneity for fast automatic data acquisition, and good sample-to-sample and spot-to-spot reproducibility for quantitative analysis.

Recently, we reported a shrinkage strategy in which a hydrophobic surface was used as the sample substrate [Hung et al., 1998]. The hydrophobic nature of the sample substrate forces the aqueous drop to shrink to a small area thereby concentrating analyte molecules on target. More recently, Nordhoff and coworkers extended this method by adding a small hydrophilic spot, which acts as an anchor that, if touched, holds the sample droplet and directs crystallization onto it [Schuerenberg et al., 2000]. Two sample loading procedures have been investigated. The first one shrinks the matrix solution first, followed by adding the DNA solution onto the top of the dried matrix film [Hung et al., 1998; Huang et al., 1999]. The second procedure shrinked the aqueous solution containing both matrix and analyte [Schuerenberg et al., 2000]. These methods work well when applied t analysis of high-concentration oligonucleotides, but did not work well with low-concentration oligonucleotides, at least in our hand, in terms of sensitivity and reproducibility. In addition, both procedures involve the shrinkage of the solution containing the matrix on a hydrophobic surface, but it is difficult to control the shrinkage of the aqueous 3-HPA solution. It was observed that if too much matrix was loaded, much larger sample areas were formed. If too little matrix was loaded, the sample quality is poorer.

In this work, we developed a new procedure in which a few μL of the oligonucleotide solution are first deposited on the target, followed by addition of a small volume (0.1μL) of saturated 3-HPA solution. It is seen that this procedure is superior to existing procedures, and can substantially improve the MALDI sensitivity and sample homogeneity. The application of this procedure on low-concentration

oligonucleotides and mini-sequencing products showed that the detection limit and data robustness were greatly improved.

5.2 Methods and Materials

Oligonucleotides were purchased from MWG Biotech Inc.(High Point, NC, USA). 3-Hydroxypicolinic acid (3-HPA), diammonium citrate, and NH_4^+ ion-exchange beads were obtained from Aldrich (Milwarkee, WI, USA). All samples were used without further purification. The saturated 3-HPA matrix solution was prepared in 50 mM diammonium citrate solution containing 25% (v/v) of acetonitrile. Stock DNA solutions were prepared in water. The concentration of the DNA solution was determined by UV absorption measurements. The DNA mixture solutions were prepared prior to loading them on the sample tip.

All mass spectra were collected using OmniFlex™ MALDI-TOF mass spectrometer (Bruker, Bilerica, MA, USA) operated in the linear, negative mode, delayed extraction has been turned on. OmniFlex™ uses a 337-nm pulsed nitrogen laser which has a duration of 3ns and a full power of 175μJ. Our experiment was typically carried out at about 78% of full laser power with 600 ns delay extraction and 9.3kV ion focus voltage. This laser power was slightly higher than that used when samples were prepared by conventional methods. Sampling rate is 2Hz. The sample target was a 400μm, 6 x 6 anchorchip™ abtained from Brucker. If not specified, otherwise, 2 μL of oligonucleotide solution were first deposited on the anchorchip™, followed by dispensing 0.1μL of the saturated 3-HPA matrix solution. All sample

spots were searched manually to locate the “sweet” spots.

5.3 Results and Discussion

5.3.1 Sample Preparation

The first series of experiments was performed to determine the best sample preparation conditions. The sample target used was a 400μm anchoring chip which is coated with a thin film of Teflon-like hydrophobic material and carries a 400μm hydrophilic center. The selection of anchoring chips over a pure hydrophobic center is based on the observation that the sample solution tends to form a dried spot on the hydrophilic surface, allowing the use of an automatic liquid handling system to precisely deposit a small volume of matrix onto the top of a dried sample spot without visualization.

The key to this procedure is to confine all deposited samples to the hydrophilic surface. It was seen that the addition of a small amount of diammonium citrate to the oligonucleotide solution could facilitate the shrinkage of a aqueous solution on an anchoring chip and reduce the salt effects. The diammonium citrate solution was normally treated with NH_4^+ ion-exchange beads before being added to the oligonucleotide solution. In general, the sample solution contained 0.001M of diammonium citrate if the oligonucleotide concentration was 0.01 μM or higher. When lower concentration samples were analyzed, the amount of diammonium citrate should accordingly be reduced. In this work, we manually dispensed both oligonucleotide and matrix solutions to the sample target. Therefore, care was taken to

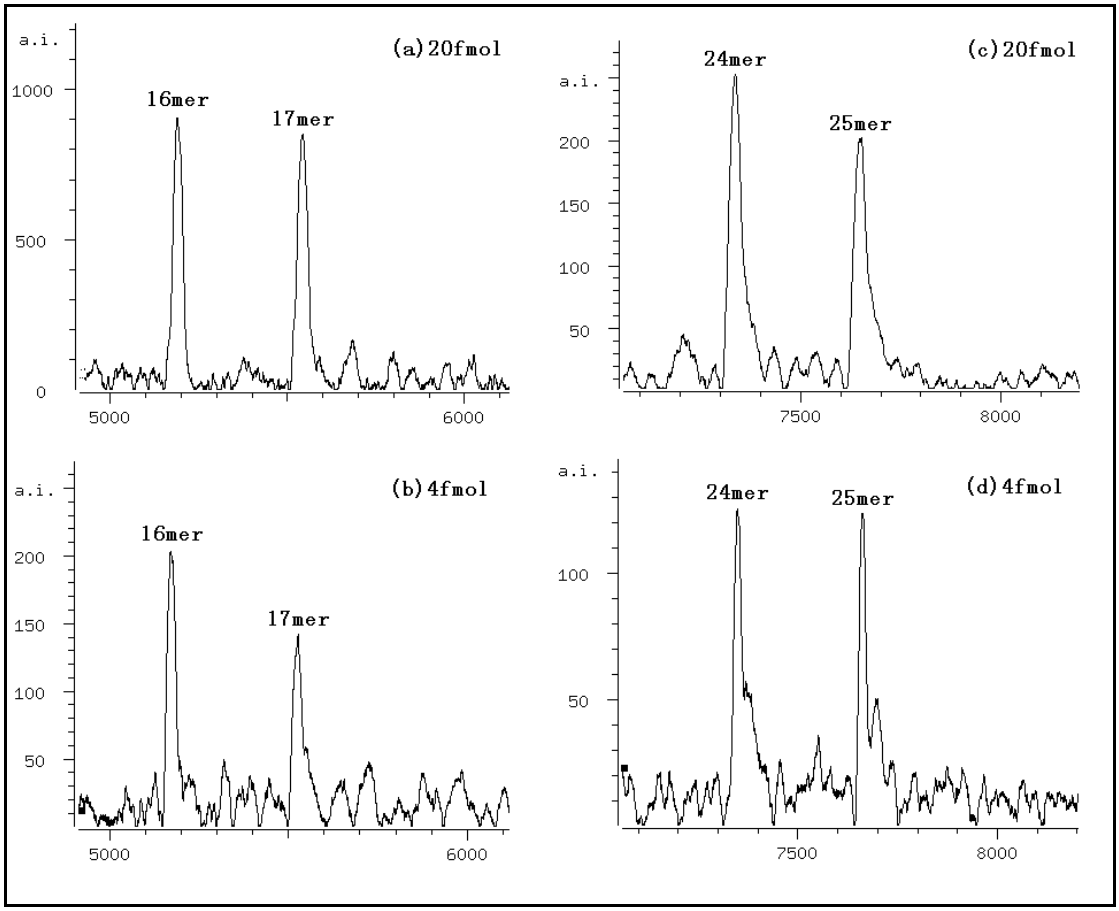
deposit the sample solution onto the center of the anchoring spot. Otherwise, the sample solution may not be confined to the hydrophilic center. In the same way, great care was taken to deposit a small drop of the matrix solution onto the top of the shrunk DNA spot. We found that the result was poorer if the matrix was not deposited on the top of the sample spot. An automated liquid handling system should be used when a large number of samples are prepared by this method, as this will eliminate the potential human error during sample preparation.

We investigated the matrix concentration effect using the matrix solution at three different concentrations. In all three cases, 0.1 μL of the 3-HPA solution was loaded. We observed that the sample sizes created from these three solutions were similar, but the more diluted the 3-HPA solution was, the thinner the 3-HPA film was formed. The diluted solution tended to form a 3-HPA film containing many thin needle-like threads. In contrast, the saturated solution produced many thin threads surrounded by small crystals across the entire sample. Interestingly, it was the saturated 3-HPA solution that gave the best results in terms of sensitivity, homogeneity, and reproducibility. We also investigated the effect of CH3CN and observed the similar results when the CH_3CN concentration was 25-50%. Higher concentration of the CH_3CN will interfere with the shrinking process. The results indicated that the careful selection of CH_3CN combined with saturated 3-HPA yields better matrix/sample co-crystals and more uniform samples (less gaps between crystal threads), improving the sampling efficiency. Therefore, the saturated 3-HPA solution prepared in 25% of CH_3CN was used throughout this experiment. In general, the

sample size formed by this procedure is about 0.4-0.6 mm.

5.3.2 Sensitivity and Homogeneity

Figure 5.1 shows typical MALDI-TOF mass spectra of equal molar oligonucleotide mixtures containing either 16/17 or 24/25 mers. The samples were prepared by loading 2 μL of the mixture solution, followed by the addition of 0.1 μL) of the saturated 3-HPA solution to the dried oligonucleotide spot. The concentration of each oligonucliotide mixture was ~0.01 μM (**Figure 5.1**(a) and **5.1**(c)) and ~0.002 μM (**Figure 5.1**(b) and **5.1**(d)), respectively. As seen from **Figure 5.1**, all four oligonucleotides were detectable at concentrations of 0.01 and 0.002 μM. However, at 0.01 μM, little efforts were needed to search for 'sweet' spots, while some searching efforts were required if the concentration was reduced to 0.002μM. A similar sensitivity was obtained with many other oligonucleotides including one 30 mer. It should be noted that no attempt was made to determine the uppermost sensitivity of this procedure as we were more interested in determining the lowest DNA concentration at which we could produce homogenous and reproducible samples, since these factors are equally important to genotyping of SNPs. Therefore, we evaluated this new procedure at the oligonucleotide concentration of 0.01 μM or higher in the rest of this work.



Figrue 5.1 MALDI-TOF mass spectra of equal molar oligonucleotide mixtures of 16/17- and 24/25mers obtained at concentrations of 0.01μM ((a) and (c)) and 0.002μM ((b) and (d)).

It was observed that the sample size could be further be further reduced, as the sample size was dictated by the volume of matrix. At present, it is difficult for us to load less than 0.1 μL of solutions with a standard syringe or pipetter. However, a piezoelectric microdispenser can deposit as little as 10 nL of a solution on target [Little et al., 1997; Miliotis et al., 2000]. It is anticipated that the use of a piezoelectric microdispenser might further improve the sensitivity of this new procedure.

The ability of rapid screening of a large number of samples is essential to high-throughput genotyping. A series of experiments with a mixture of 16- and 17 mers at 0.01 μM was conducted to evaluate the application of this procedure to a high-throughput operation. First, we examined the signal duration on a given spot to see whether a good spectrum can be generated just from a single targeted spot. **Figure 5.2** displays three spectra collected from the first 30 shots, second 30 shots, and third 30 shots from the same targeted spot, respectively. In most of the cases, strong ion signals were still observable even after 70 laser shots if the laser power was kept at the level that was slightly higher than threshold. Importantly, it was seen that good signal-to-noise (S/N) peaks could be robustly generated just from a single spot. The good signal duration and quality allow acquisition of a good spectrum from one spot instead of multiple spots, thereby avoiding the need to search for additional spots. It should be noted that, in some cases, there were no oligonucleotide ion signals or signals were much weaker during the first few laser shots. However, the signals actually increased significantly after a few laser shots. This suggests that oligonucleotides distributed from the top surface to the interior part of the 3-HPA film

and that long signal duration may result from the formation of a thicker 3-HPA matrix film.

The sample homogeneity was further examined by preparing samples in an identical manner, followed by collecting one good spectrum from each of 36 wells. In this experiment, we intended to determine how many sample spots should be searched randomly to find a 'sweet' spot. An S/N ratio of 5 was used to accept or reject the spectra. If the S/N ratio was equal to or better than 5, no more searching was performed. Otherwise, more spots were searched until a good spectrum was collected. In this experiment, the 'sweet' spots were searched manually. We found that all 36 samples produced the similar high-quality spectra and that the average number of the spots searched was 2-3. For example, **Figure 5.3** displays the typical spectra obtained from sample targets 3, 7, 16, 24, 28 and 35, respectively. Moreover, we observed that high-quality spectra could be produced across the entire sample. This result clearly indicated that the samples prepared by this new procedure are highly homogenous and therefore it is suitable for high-throughput DNA analysis.

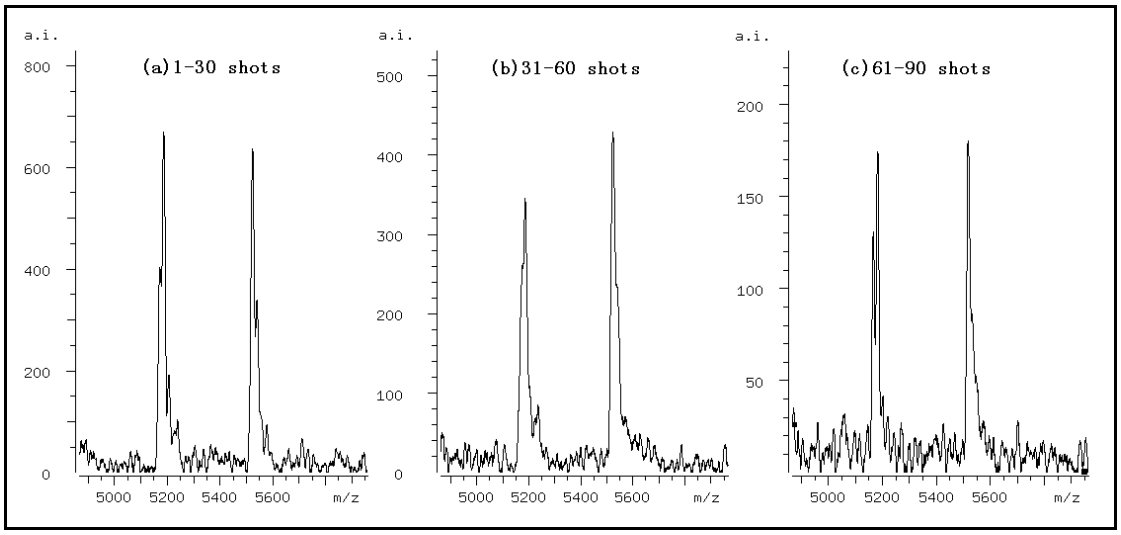


Figure 5.2 Typical MALDI-TOF mass spectra obtained from the first 30 shots, second 30 shots, and third 30 shots.

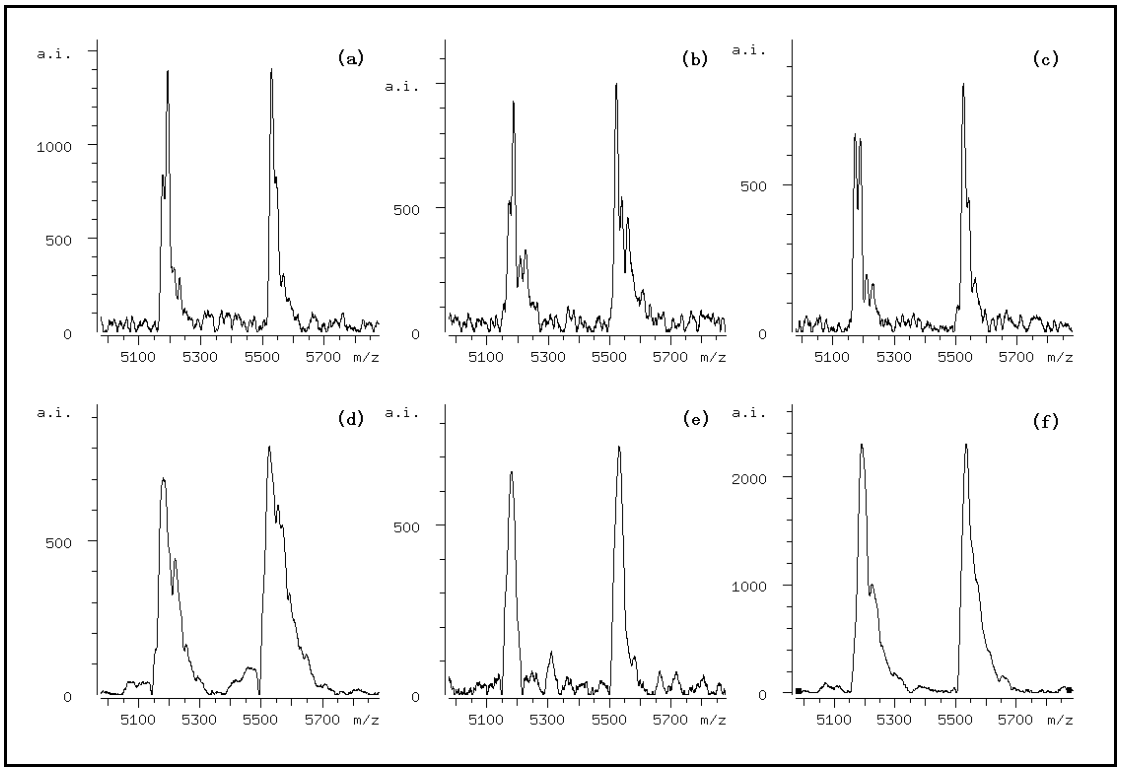


Figure 5.3 Representative MALDI-TOF spectra obtained from 36 sample spots prepared in an identical manner. The spectra shown were obtained from spots 3, 7, 16, 24, 28, and 35.

5.3.4 Reproducibility and Quantitative Analysis

One other issue related to genotyping of SNPs is the determination of the frequency of each of the two alleles when a pooling sample of many people is analyzed. Therefore, we also evaluated this procedure for its applicability in quantitative analysis of two oligonucleotides which are different by one base. The objective of this evaluation was to determine how many people could potentially be pooled together for an accurate determination of the allele frequencies. In this experiment, six samples were prepared in an identical manner, and five spectra were collected from five different positions from each of the six samples. The sample mixture used contained 0.04 μM of 16- and 17 mers. **Table 5.1** lists the signal intensity ratios of the 16- and 17mer peaks from each of the 30 spectra. These peak intensity ratios were calculated based on the peak aera including the area of the salt adduct peaks. It was seen that, on average, only one out of 10 spots produced an unusually large or small peak ratio. Based on this observation, we used the following sampling procedure to improve the data reproducibility from spot to spot. All five spectra collected from the same sample were compared with each other. If any spectra produced a peak ratio larger or smaller than 30% of the average ratio, those spectra were discarded and only the remaining spectra were used to calculate the average peak ratio for that sample. **Table 5.1** also lists the average peak ratios of each sample after we discarded two peaks (point 2 of sample 1 and point 5 of sample 2) having unusually large or small peak ratios. It was seen that the largest peak ratio difference among the six samples could be reduced to a level of less than 10%. Therefore, we

expect that when this method is used to measure the allele frequency of a pooling sample of more than five individuals, the error of the resulting frequency will be about 10%. However, if the pooling sample contains five individuals or less, this method can yield the precise allele frequency, as the smallest genotype frequency change is 33% from the 4/6 to 5/5 if the pooling sample contains five individuals.

Sample	Peak Ratio (16mer/17mer)					Average Peak Ratio with All Points	Adjusted Average Peak Ratio
	Point 1	Point 2	Point 3	Point 4	Point 5		
1	0.862	1.266	0.935	0.962	0.935	0.992	0.9235
2	0.99	1.064	0.901	0.885	0.676	0.9032	0.96
3	1.124	0.862	1.099	1.136	0.82	1.0082	1.0082
4	0.901	1.01	0.917	0.962	0.855	0.929	0.929
5	0.794	1.163	0.87	0.901	0.847	0.915	0.915
6	1.064	0.794	0.962	1.064	1.087	0.9942	0.9942
SD*						0.0462	0.0391

Table 5.1 Experimentally determined peak ratios and calculated average peak ratios of 16/17mers. * SD = Standard Deviation

5.3.5 Analysis of Mini-sequencing Products

This work was carried out with the primary aim of improving the overall detection sensitivity of the MALDI-TOF based genotyping methods. Therefore, we finally evaluated application of this method to the detection of mini-sequencing products. Mini-sequencing is a short sequencing reaction which only extends the sequencing primer by a few bases. The SNP site genotyped in this work was SNP888 [Martin et al., 2000]. Two people were genotyped. One is Homozygous, while the other is C/T Heterozygous. In brief, genomic DNA was amplified by PCR, followed by purification with a size-exclusion column. The purified PCR products were then subjected to VSET mini-sequencing [Sun et al., 2000]. A sequencing primer is annealed to the target alleles immediately upstream of the SNP site, which is extended by one of the ddNTPs from one allele, then it is extended by the addition of one dNTP, and then terminated by a ddNTP from another allele. As a result, two extension products which differ by one base are produced from a heterozygous individual, while only one extension product is yielded from homozygous person. 50 or 100 fmol of the extension primer were used for mini-sequencing. A total of four identical mini-sequencing reactions were prepared for each concentration. Mini-sequencing products were subjected to purification with Ziptip^{TM} (Millipore, MA, USA). The Oligonucleotides were eluted from the tip with 1μL of an aqueous solution containing 20% of CH3CN. The mini-sequencing products were treated with the same tip three times to improve the oligonucleotide recovery efficiency. Thereafter, all recovered DNA solutions were combined and deposited on a sample well, followed by the

addition of matrix.

Figure 5.4(a)-**5.4**(d) showed the results of genotyping of these two individuals using 100 and 50 fmol of the extension primer, respectively. **Figure 5.4**(a) and **5.4**(b) were obtained from genotyping the C/T heterozygous individual, while **Figure 5.4**(c) and **5.4**(d) were obtained from genotyping the homozygous individual. The peaks labeled P# (# =1 or 2) correspond to the extended primers. It was seen that the genotypes of all samples were correctly identified. In general, the signal quality and sample homogeneity from these mini-sequencing products were similar to those of the samples produced with 20 fmol of pure oligonucleotides. This suggests that the new procedure could indeed be used to analyze mini-sequencing products in a sensitive and accurate manner. The sensitivity achieved with this new procedure is about 10 times better than that achieved with currently used sample preparation methods, and is comparable with the sensitivity of fluorescence-based detection. Because of the lower amounts of extension primers used in mini-sequencing, less PCR products and DNA polymerase are required for mini-sequencing.

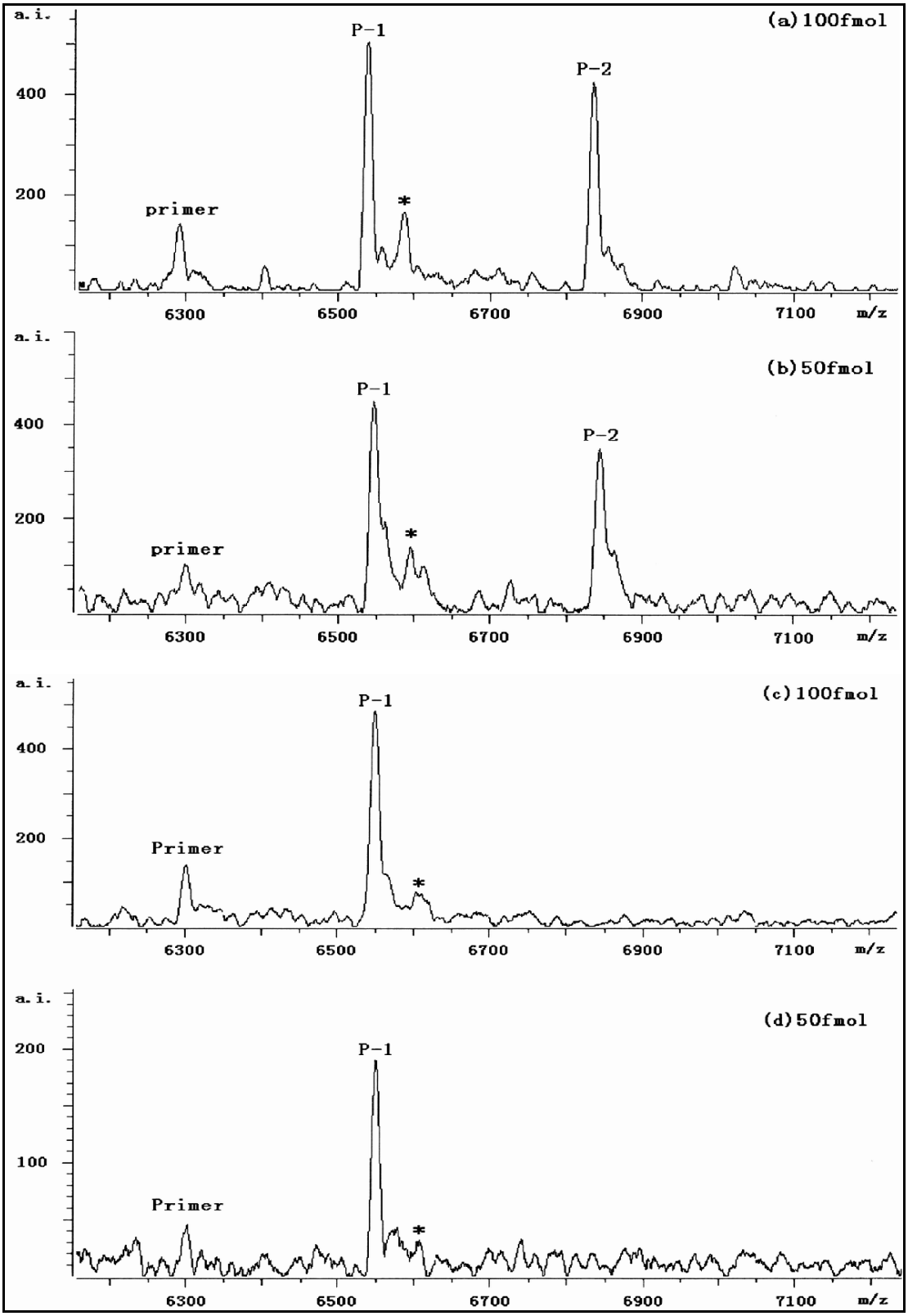


Figure 5.4 MALDI-TOF mass spectra obtained from genotyping two individuals ((a) and (b) C/T heterozygous; (c) and (d) C homozygous) using 100 and 50 fmol of the extension primers, respectively. The peaks labeled with * may be due to the products arising from the impurityies present in the mini-sequencing primers

5.4 Conclusions

A new procedure for preparing MALDI samples on a hydrophobic surface was developed for the MALDI-TOF analysis of low-concentration oligonucleotides. This new procedure was simple, robust, and superior to the conventional sample preparation methods. It was shown that this procedure allowed the robust detection of as low as 0.01μM of oligonucleotides ranging from 15- to 30 mer. Moreover, the samples prepared by this procedure were more homogenous and therefore substantially reduced the need to search for 'sweet' spots. The increased shot-to-shot and sample-to-sample reproducibility also make it possible to perform high-throughput and quantitative analyses of DNA. It was shown that this procedure could reduce sample-to-sample peak ratio variation to the level of 10%, therefore allowing the accurate determination of the allele frequency of a pooling sample of at least five people. Importantly, we demonstrated that this procedure could improve the overall detection sensitivity of the MALDI-TOF-based genotyping method by about 10-fold, evident by the robust detection of SNP888 using as little as 50 fmol of the extension primer. This detection sensitivity is comparable with that of the fluorescence-based detection method.

CHAPTER VI

DETECTION OF MRS ALTERATIONS IN A LARGE BACKGROUND OF NORMAL DNA FOR SCREENING OF MSI-H CANCERS

6.1 Introduction

High-frequency microsatellite instability (MSI-H) is a form of genetic instability observed in virtually all tumors from tumors from patients with hereditary nonpolyposis colorectal cancer (HNPCC) and in a subset of various sporadic cancers. Its hallmark is extensive instability in simple repeat nucleotide sequences, including mononucleotide repeat sequences (MRS) and microsatellites [Boland et al., 1998]. MSI-H is most common in colorectal, gastric, and endometrial cancers [Markowitz et al., 1995; Halling et al., 1999; MacDonald et al., 2000; Alexander et al., 2001]. Colorectal cancer is the second deadliest cancer in the United States, and studies identified that about 15% of colorectal cancers are MSI-H, suggesting an annual

incidence in the United States of 20,000 to 26,000 MSI-H colorectal cancers each year [Alexander et al., 2001].

Detection of colorectal cancer by the DNA analysis from stools is a promising screening method [Ahlquist et al., 2000 and 2002; Dong et al., 2001; Traverso et al., 2002a,b]. MRS markers are excellent indicators of MSI-H [Myeroff et al., 1995; Hoang et al., 1997; Grady et al., 1998; Zhou et al., 1998; Lacopetta et al., 1999; Loukola et al., 1999; Jin et al., 2001; Loukola et al., 2001]. For example, BAT26 alterations were found in > 90% of MSI-H colon tumors, and 85% of MSI-H colon cancers harbor a mutation in transforming growth factor-β-RII (A)10 (TGF-βRII (A)10) [Myeroff et al., 1995; Hoang et al., 1997; Zhou et al., 1998; Loukola et al., 2001]. Studies show that alterations of these markers are not present in normal tissues, suggesting that they are tumor specific [Grady et al., 1998; Lacopetta et al., 1999; Loukola et al., 1999; Jin et al., 2001]. Stools contain only a small percentage of mutated cells. Thus, molecular assays need to be highly sensitive such that altered DNA can be readily detected in a large background of wild-type DNA. Highly sensitive PCR-based methods are available for the point mutation detection [Orum et al., 1993; Jacobson et al., 1994; Lehman et al., 1996; Peter et al., 1997; Zirvi et al., 1999; Sun et al., 2002], but they are not suited for MRS analysis, as PCR itself can produce substantial slippage errors when a locus containing MRS is amplified. In other words, PCR can artificially yield mutant alleles even in the absence of mutant DNA, leading to false-positives.

Ahlquist et al. reported on the use of a modified primer extension method for

the detection of BAT26 mutations in stool [Ahlquist et al., 2000 and 2002; Dong et al., 2001]. However, primer extension methods may not work well for the detection of mutations in short MRS markers, such as TGF-β-RII (A)10, because polymerase slippage PCR products generated from the wild-type DNA can substantially overlap with the common (A)9 mutant. Recently, Vogelstein et al. developed an elegant digital PCR method [Vogelstein et al., 1999] and successfully employed this approach to detect mutant BAT26 alleles in fecal DNA [Traverso et al., 2002]. Digital PCR is sensitive, but the method requires analysis of a large number of subdivided samples each containing a single DNA molecule to screen each patient.

We herein report an alternative detection strategy referred as probe clampling primer extension-PCR (PCPE-PCR) for the detection of MRS alternations in a large background of wild-type DNA. We demonstrated that PCPE-PCR can detect both mutated BAT26 and TGF-β-RII (A)10 markers in the presence of 500-fold excess of normal DNA, and that as few as three copies of mutant DNA could be detected.

6.2 Methods and Materials

Normal DNA was purchased from Sigma (St. Louis, MO). Mutant TGF-β-RII (A)10 DNA was extracted from cell line of HCL116, and mutant BAT26 DNA was collected from cell line of HEC1A. The low-abundance mutant DNA samples were created by mixing mutant DNA with normal DNA. The abundance and number of mutant DNA in the created samples were estimated based on the number of mutant DNA in the original samples and dilution factors. The stool DNA samples

were extracted from stools of MSI-H colorectal cancer patients [Traverso et al., 2002]. In this work, a peptide nucleic acid probe of 5'-GGCTTTTTTTTTCCT-3' (Applied Biosystems, Foster City, CA) was used for TGF-β RII (A)10 assay, and an oligo probe (5'-GGTAAAAAAAAAAAAAAAAAAAAAAAGGG-3') was used for BAT26 assay. The oligo probe was phosphorothioted at the first five positions of 5' and 3' ends to minimize the probe cleavage by DNA polymerases and was also 3'-end phosphorylated to prevent the probe itself from undergoing primer extension.

For TGF-β RII (A)10, PCPE was carried out in 25μL using 3 μmol/L of the peptide nucleic acid probe, 0.01μmol/L of the extension primer (5'-Biotin-TGCACTCATCAGAGCTACAGG-3'), 0.1mmol/L of each deoxynucleotide triphosphates, 2mmol/L of $MgCl_2$, 1x AmpliTaq Gold PCR buffer, and 0.5 unit of AmpliTaq Gold DNA polymerase (Applied Biosystems). After denaturation at 95℃ for 10 minutes, PCPE was done for 25 to 50 cycles consisting of 30 seconds at 95℃, 120 seconds at 58℃, 60 seconds at 54℃, and 60 seconds at 72℃, with a final extension of 5 minutes at 72℃. After PCPE, the formed single-strand DNA fragments were captured using streptavidin-coated magnetic beads (Dynal Biotech, Lake Success, NY). Twentyfive microliters of extension products were mixed with equal volume magnetic beads in B&W buffer (10mmol/L Tris-HCl (pH 7.5), 1mmol/L EDTA, 2.0mol/L NaCl), and incubated at room temperature for 1 to 3 hours. Thereafter, the supernatants were removed followed by washing the beads with 200 μL of 0.1mol/L NaOH for 5 minutes, and twice by water.

A similar procedure was used for BAT26 except for that PCPE was done in

25 to 50 cycles consisting of 30 seconds at 95℃, 120 seconds at 68℃, 60 seconds at 62℃, and 60 seconds at 72℃, with a final extension of 5 minutes at 72℃. The extension primer is 5'-Biotin-TGCAGTTTCATCACTGTCTGC-3'.

The purified beads containing the single-strand DNA fragments served as templates of fluorescence-based PCR. The PCR mixture included 1x PCR buffer, 0.2mmol/L of each deoxynucleotide triphosphates, 2mmol/L $MgCl_2$, 0.1 μmol/L of the forward and reverse primers, and 0.5 unit of TaqGold DNA polymerase, respectively. After denaturation at 95℃ for 10 minutes, PCR (25μL) was done for 42 cycles consisting of 30 seconds at 95℃, 30 seconds at 54℃, and 30 seconds at 72℃, with a final extension of 5 minutes at 72℃ for TGF-β RII (A)10 and BAT26. The fluorescence-labeled PCR products were size analyzed using a CEQ8000 sequencer (Beckman Coulter, Fullerton California). The forward and reverse primers used were 5'-GAAGATGCTTCTCCAA-3' and 5'-dye(D4)-ATCAGAGCTACAGGAACAC-3' for TGF-β-RII (A)10, 5'-dye(D4)-ATTGGATATTGCAGCAGTC-3' and 5'-AACCAATCAACATTTTTAACCC-3' for BAT 26, respectively.

Fecal DNA samples from individuals having MSI-H colon cancers were the kind gift of Dr. Bert Vogelstein (John Hopkins University) and corresponded to samples previously employed in studies published by Dr. Vogelstein's laboratory.

6.3 Results and Discussion

6.3.1 Principle of PCPE-PCR

A PCPE-PCR assay consists of five steps. First, DNA is extracted from a

clinical sample, such as stools. Second, the DNA samples containing both mutant and wild-type alleles are subject to PCPE that preferentially produces extension products of mutant DNA. Third, the extension products are extracted, leading to enrichment of mutant DNA. Fourth, PCR is done using the extracted extension products as templates. Finally, PCR products are analyzed using DNA fragment analysis or other methods to reveal the mutations. **Figure 6.1** shows a schematic representation of detecting mutant DNA (A9) in a large background of normal DNA (A)10 by PCPE-PCR.

The second and third steps are key to PCPE-PCR. As shown in **Figure 6.1**, the PCPE system has an extension primer and a blocking probe, both of which are complementary to the same strand of the wild-type sequence. The blocking probe is complementary to a wild-type sequence containing MRS and designed in such a way that probe-DNA duplexes are more stable than extension primer-DNA duplexes when DNA is wild-type but less stable than extension primer-DNA duplexes if DNA is mutated. As a result, this blocking probe can tightly bond to wild-type DNA, preventing read-through by polymerases (elongation arrest); thus, shorter extension products that do not contain the MRS sequence are formed (note that they will not be amplified by the following PCR as one of the PCR primers will not cover them). In contrast, when DNA is mutated, this blocking probe would no longer bind to the targeted mutation site; thereby, primer elongation is initiated by polymerases, and long extension products containing the altered MRS sequence are formed. In this method, we use solid-phase extraction to extract the extension products. A biotin is incorporated to the extension primer; therefore, the formed extension products are

consequently biotinylated as well. The extension products are isolated using a biotin-streptavidin affinity extraction procedure, leading to enrichment of mutant DNA. Thereafter, we use the extracted extension products as PCR templates, and the mutation is revealed by analyzing the PCR products.

In this study, we used the TGF-β-RII (A)10 and BAT26 markers as examples to show the detection of mutations in both short and long MRS markers by this new technology, respectively.

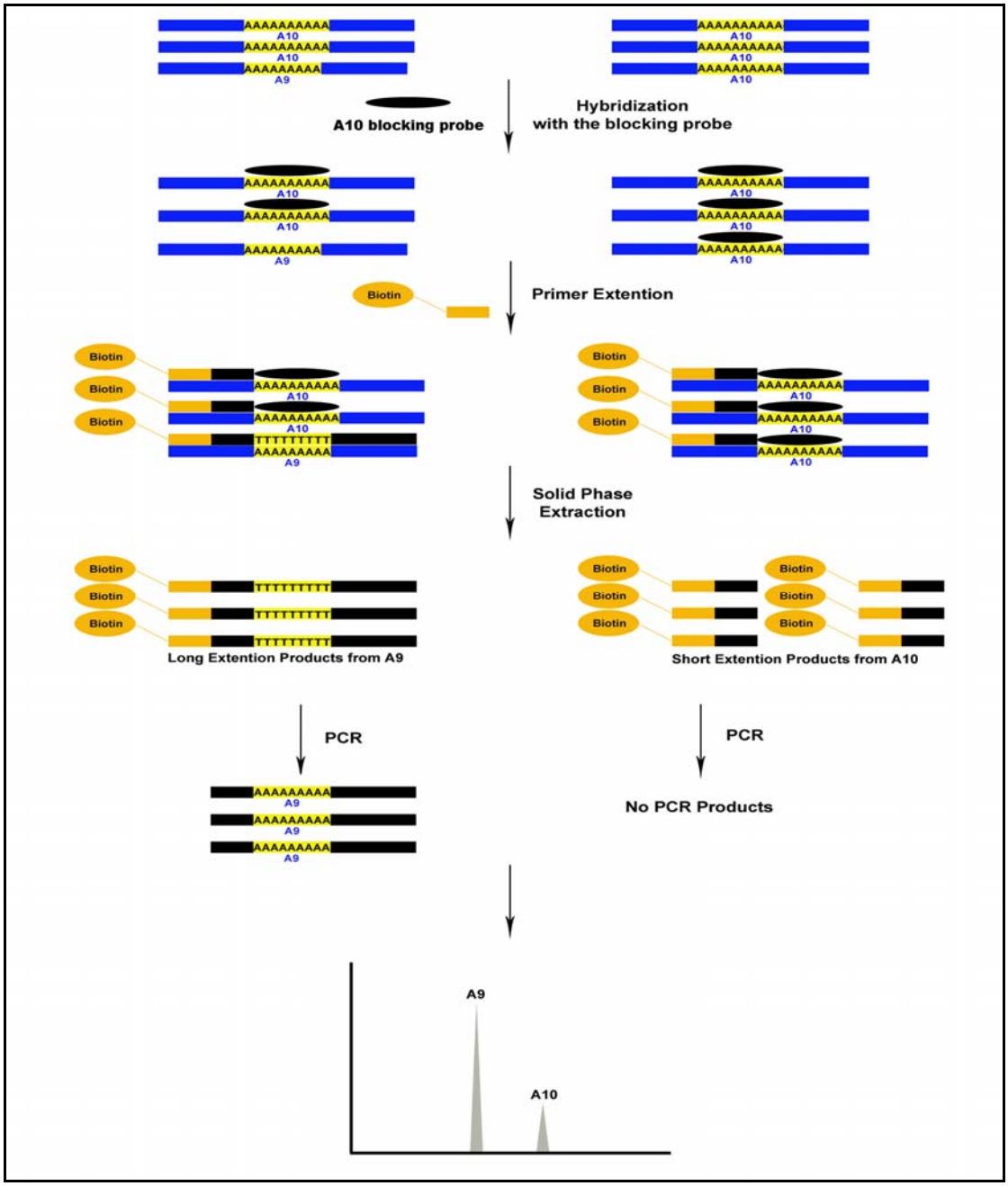


Figure 6.1 Schematic illustration of the PCPE-PCR principle of detecting mutant DNA (A9) in the presence of a large background of normal DNA (A10).

6.3.2 Detection of Mutations in TGF-β RII (A) 10

We first used PCPE-PCR to detect mutations of short MRS marker, which contain 12 or less repeats, and they are often altered by one or two bases. Because of PCR slippage, it can be challenging to detect these mutations in the presence of excessive wild-type alleles. For example, 10% to 30% of the PCR products amplified from the normal TGF-β-RII (A)10 allele can be shortened A9 alleles, making it challenging for conventional methods to detect the legitimate presence of the mutant A9 allele if the abundance of mutant DNA is low. However, the detection of this marker was readily achieved by PCPE-PCR.

Figure 6.2A shows the PCR product distribution obtained using primer extension but without blocking probe. The sample contained only wild-type DNA. It was seen that this distribution was identical to that generated from PCR without primer extension, establishing the fact that the sequence of the DNA molecule produced from primer extension truly reflects that of genomic DNA. The A9 peak seen in **Figure 6.2A** was due to PCR slippage, and its intensity was generally <50% of the A10 intensity. Because of this fact, we can use the peak intensity ratio of A9 to A10 to determine whether mutant DNA is present. It was found that a ratio of 0.8 is a good threshold to robustly distinguish the legitimate presence of the A9 allele from the artificial production of this allele by PCR slippage. In other words, we called the legitimate presence of the mutant A9 allele in the sample if the ratio is 0.8. Otherwise, the absence of mutant DNA would be called.

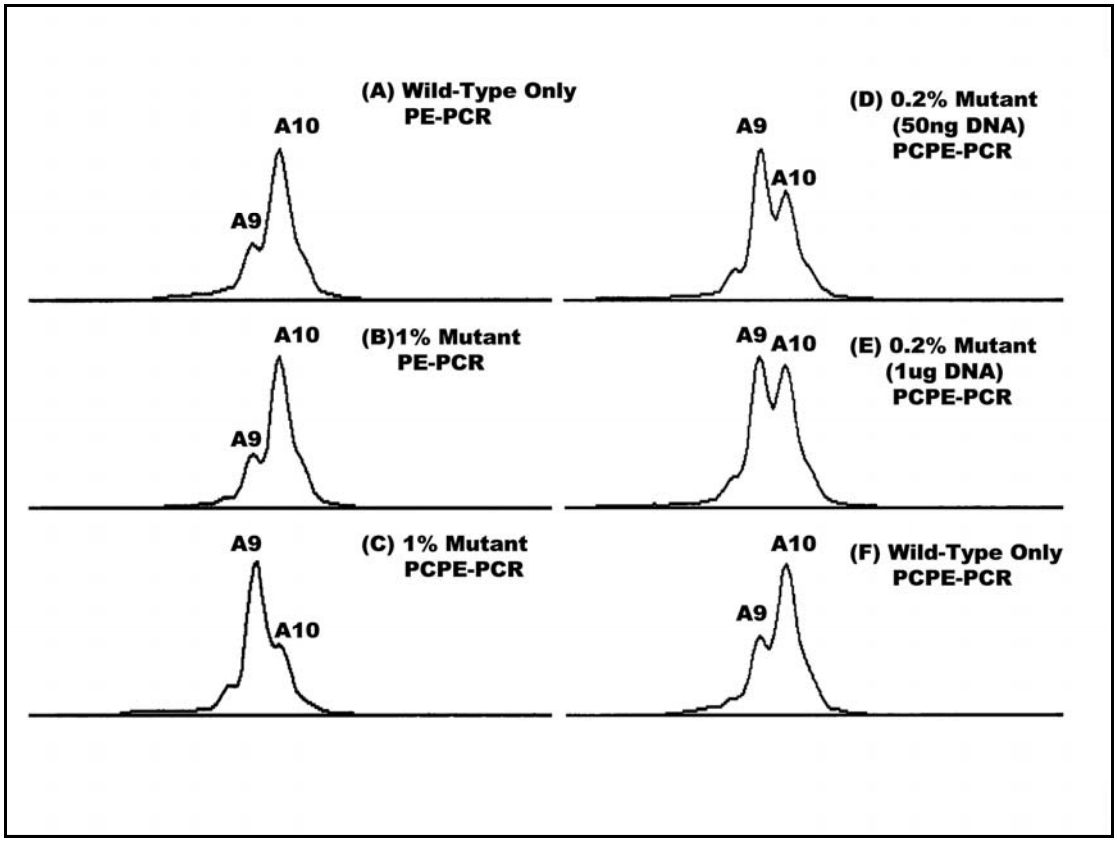


Figure 6.2 TGF-β-RII (A)10 spectra obtained from different conditions. It is noted that: (a) PE-PCR denotes the use of primer extension (without the blocking probe) followed by PCR; (b) the percentage indicates the abundance of mutant DNA in the sample; (c) the peak A9 corresponds to mutant DNA, whereas the peak A10 corresponds to wild-type DNA.

Figure 6.2B shows the result of detecting 1% (0.5 ng) of mutant DNA in 50 ng of normal DNA without the blocking probe, where no mutant DNA was observed. In contrast, **Figure 6.2C** displays the result obtained from the same sample but with the blocking probe, where the A9 peak became stronger than the A10 peak, indicating the legitimate presence of mutant DNA in the sample assayed. **Figure 6.2D** displays the result of detecting 0.1 ng (0.2%) of mutant DNA in 50 ng of normal DNA, where the A9 peak was still stronger, establishing that PCPE-PCR could detect 0.2% of mutant TGF-h-RII (A)10. The amount of human DNA extracted from a clinical specimen may vary from patient to patient; thus, the assay must have a good dynamic range. **Figure 6.2E** displays the detection of mutant DNA (0.2%) in the presence of 1 μg of normal DNA, indicating that PCPE-PCR works well in a large dynamic range. It should be noted that the specificity of this assay is good as shown by assaying pure wild-type DNA (**Figure 6.2F**), where the A9 peak was <50% of the A10 peak, correctly indicating the absence of mutant DNA. .

6.3.3 Detection of Mutations in BAT26

We next used PCPE-PCR to detect mutations in large MRS markers. These makers generally contain 20 repeats or more and typically contract by multiple bases. BAT26, one of the most wildly used markers for MSI-H colon cancer, was examined. BAT26 typically contracts by 10 bases or more in MSI-H colorectal cancer but contracts less in adenoma. The data shown here were based on the detection of mutant BAT26 from cell line HEC1A, in which one allele contains about 14As (large

deletion), whereas another has 20As (small deletion). The use of this mutant cell line allowed us to evaluate the performance of PCPE-PCR in the detection of both small and large deletions under the same assaying condition. **Figure 6.3A** and **6.3B** shows the PCR product distribution of both pure normal and mutant BAT26 DNA, respectively. The numbers shown in these spectra specify the size of the corresponding PCR products. The peaks clustering around the positions of 86, 80, and 74 corresponded to the PCR products arising from wild-type BAT26, small deletion in BAT26, and large deletion in BAT26, respectively. The formation of multiple peaks resulted from PCR slippage. As seen from **Figure 6.3A**, when there was no mutant DNA, the PCR products of wild-type DNA cluster around peak 86, and the peak intensity of smaller PCR products generally decrease with a decrease in DNA size. For example, peak 80 is weaker than peak 81. However, when mutant DNA was present, the PCR product distribution changed substantially as shown in **Figure 6.3B**. In general, the PCR products of BAT26 with large deletion do not substantially overlap with those of normal DNA, allowing the determination of mutant DNA based on the DNA fragment size. For example, the appearance of peaks around position 74 is indicative of large deletion in BAT26. Although PCR products of BAT26 with small deletion do overlap with those of wild-type DNA, the peak intensity pattern becomes altered due to overlapping. For example, peak 79 is stronger than peak 81 in **Figure 3B**. Thus, we were able to determine whether small deletion in BAT26 was legitimately present based on the resulting changes in the peak intensity pattern.

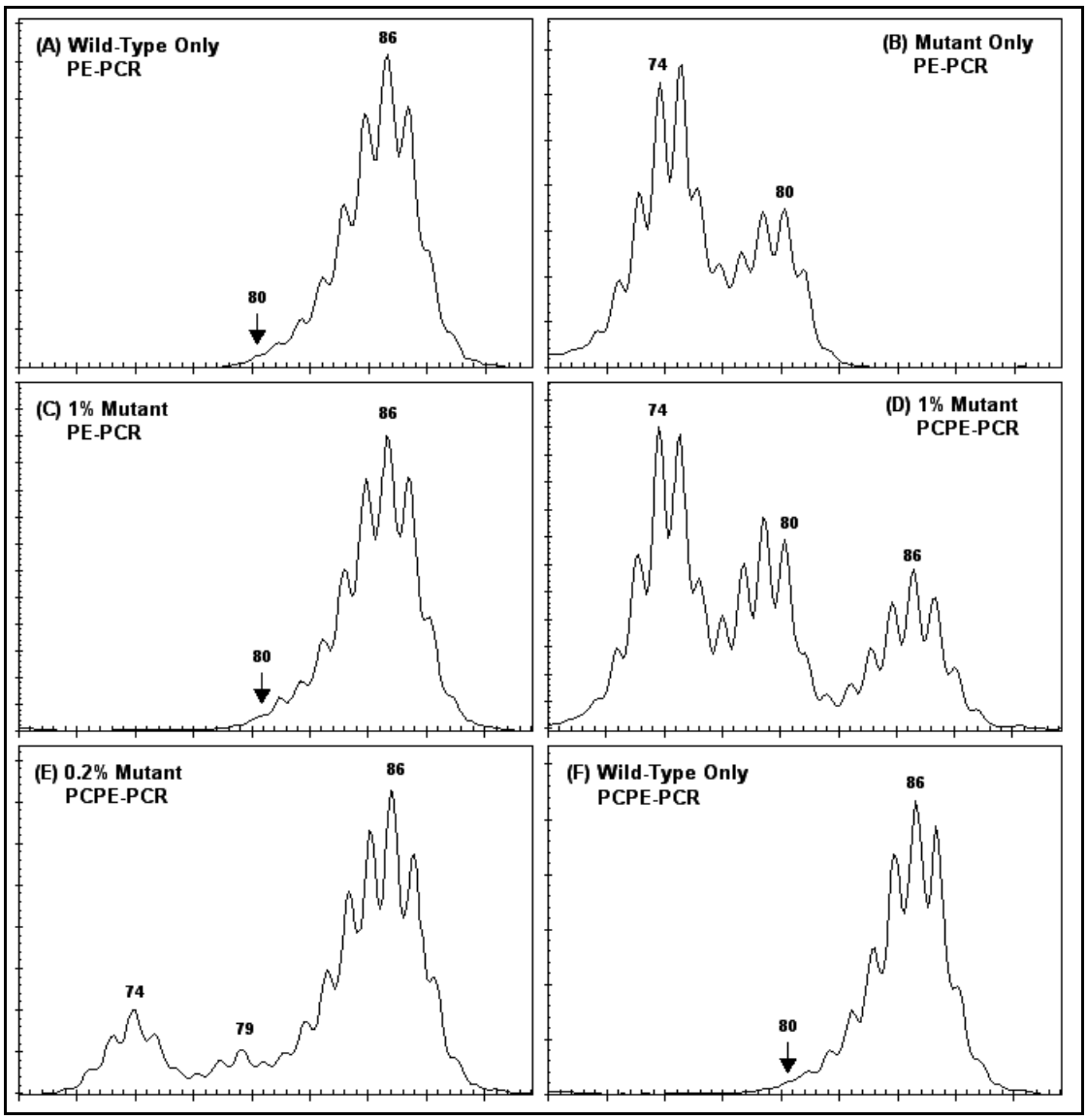


Figure 6.3 BAT26 spectra obtained from different conditions and samples. It is noted that: (a) PE-PCR denotes the use of primer extension (without the blocking probe) followed by PCR; (b) the percentage indicates the abundance of mutant DNA in the sample; and (c) the numbers of 86, 80, 79, and 74 specify the size of the corresponding PCR products.

Figure 6.3C shows the detection of 1% (0.5 ng) of mutant DNA in 50 ng of wild-type DNA without the blocking probe, where no mutant DNA was observed. Figure 3D displays the result obtained form the same sample but with the blocking probe, where new peaks appeared around the peak position of 74, indicating the presence of large deletion in BAT26. In addition, peaks 79 and 80 became stronger than peak 81, suggesting the presence of small deletion in BAT26. **Figure 6.3E** displays the result of detecting 0.1 ng of mutant DNA (0.2%) in 50 ng of normal DNA, where the peaks clustering around 74 were seen, revealing the presence of large deletion in BAT26. Moreover, peak 79 was higher than peak 81, indicating the presence of small deletion in BAT26. **Figure 6.3F** was obtained using pure wild-type DNA, where neither new peaks nor a change in the peak intensity pattern were observed, indicating the absence of mutant DNA. It is noted that the above experiments were repeated at least twice, and the consistent results were obtained.

Several studies established that the abundance of mutant DNA in stools was generally >0.5% [Dong et al., 2001; Traverso et al., 2002a,b]; thus, an assay with such a sensitivity should be adequate to detect mutant BAT26 from stools [Traverso et al., 2002a,b]. Although the BAT26 assay is generally more sensitive to detecting large deletion in BAT 26, it is capable of detecting 0.2% of mutant BAT26 that was deleted by six bases, suggesting that the PCPE-PCR–based BAT26 assay could detect both small and large deletion in BAT26 in a sensitive manner. This feature is important to screening as the size of deletion in BAT26 varies among patients.

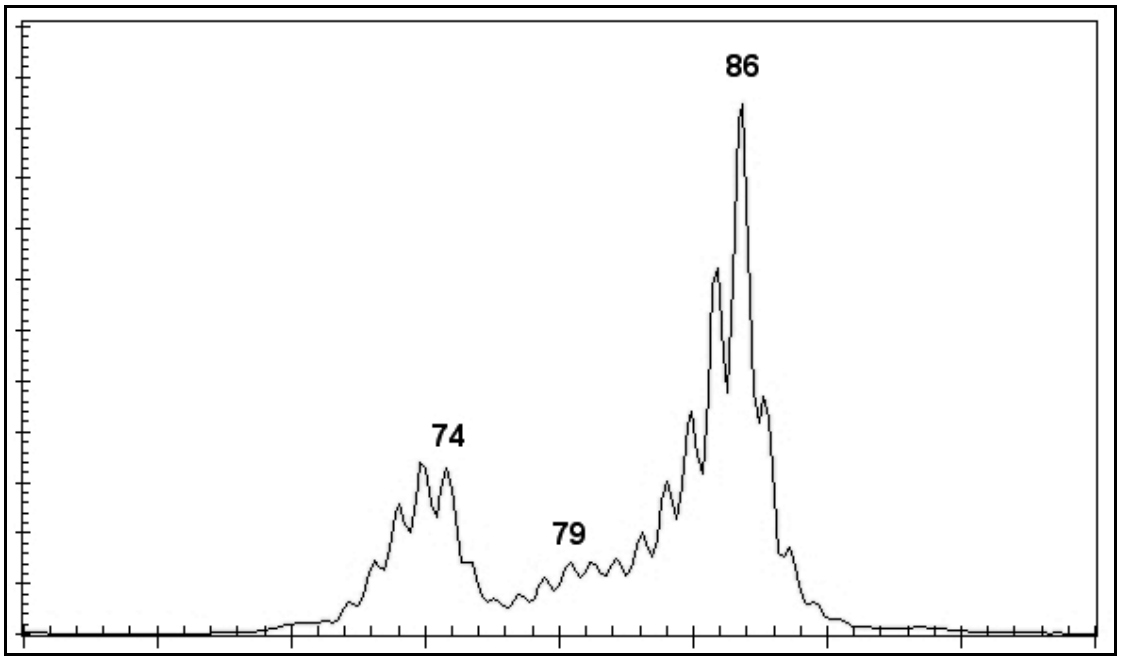


Figure 6.4 Fragment analysis spectrum obtained from the PCPE-PCR analysis of three copies of both large deletion and small deletion BAT26 in the presence of 600 copies wild-type DNA.

Because mutant DNA extracted from stools could be scanty [Dong et al., 2001; Traverso et al., 2002a,b], we also determined the lowest copy of the mutant BAT26 molecules that could be detected in the presence of a 200-fold excessive amount of wild-type DNA. As seen from **Figure 6.4**, three copies of both 12- and 6-base deleted BAT26 were detected in the presence of 600 copies of normal DNA using our BAT26 assay. It was reported that at least 4,000 copies of human DNA could be extracted from 10 g of stools, yielding at least 20 copies of mutant DNA [Dong et al., 2001; Traverso et al., 2002a,b]. Thus, the BAT26 assay is sufficiently sensitive for fecal DNA testing.

6.3.4 Detection of Mutations from Fecal DNA

We finally tested the BAT26 assay in a set of six fecal DNA samples extracted from stool samples collected from MSI-H colorectal cancer patients to determine whether this assay can detect MSI-H colorectal tumors. Stool DNA was extracted and purified as described previously [Traverso et al., 2002]. Because the samples were originally extracted for digital PCR, the total number of human DNA molecules in them is low. **Figure 6.5** shows the result of detecting the BAT26 mutations from one of the stool DNA samples using PCPE-PCR, where the total amount of stool DNA used was about 120 copies. The peaks clustering around the peak position 72 indicated the presence of mutated BAT26, thus revealing that this patient has an MSI-H colorectal tumor. Mutated BAT26 molecules were similarly successfully detected from three of the other assayed fecal DNA samples. However,

two of the six fecal DNA samples from MSI-H patients tested negative for BAT-26 because of failure of observing PCR products. During the course of this work, we found that the assay would become less robust when mutant DNA molecules were less than three copies. This is because robust isolation of extension products became challenging if fewer biotinylated molecules were formed. Because the total number of human DNA in the samples are 120 copies, we suspect that in these two negative samples, the copy number of mutant DNA may be <3, which is the minimum level required for PCPE-PCR to detect the mutated BAT26 molecules.

The test is clearly very sensitive; however, the specificity has yet to be completely determined with fecal DNA samples and will need to be further studied in a large number of samples from patients.

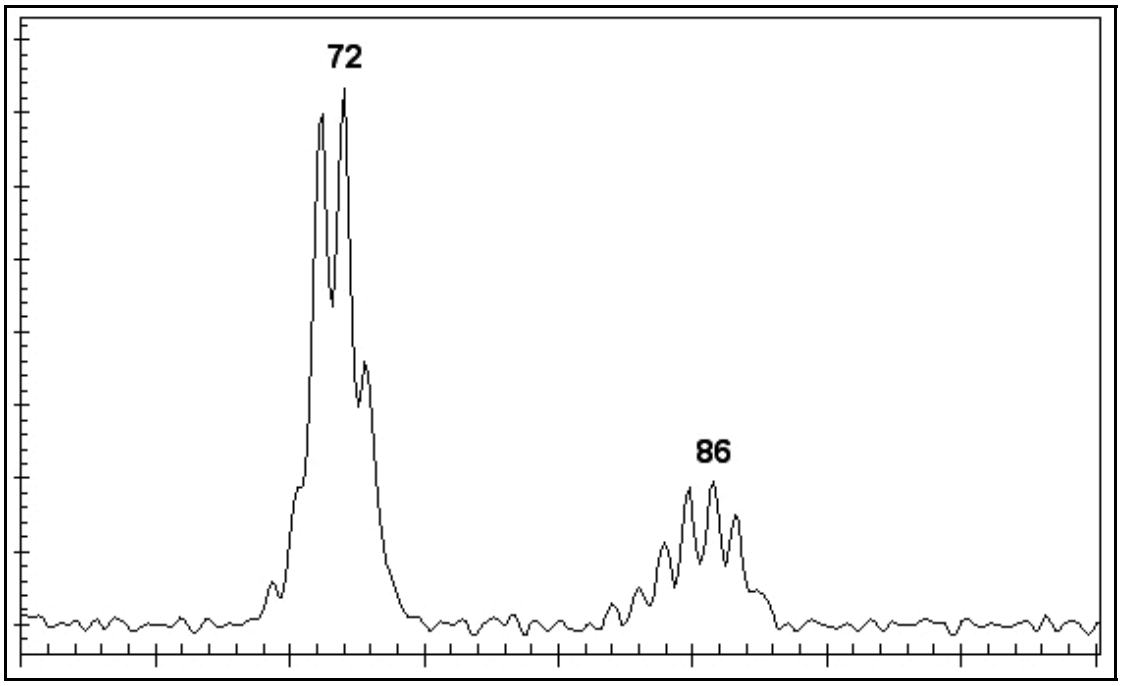


Figure 6.5 Fragment analysis spectrum obtained from the PCPE-PCR analysis of the BAT26 mutation in a stool DNA sample collected from a MSI-H colorectal cancer patient. It is noted that the peaks around 72 correspond to mutated BAT26, whereas the peaks around 86 correspond to wild-type DNA.

6.4 Conclusion

PCPE-PCR offers several unique features. First, PCPE-PCR is sensitive. Second, the same PCPE-PCR assay can be used to detect both large and small deletions. Third, the PCPE-PCR assay uses primer extension as pre-amplification to increase template copies, allowing the detection of low copies of mutant DNA. This feature is particularly useful in analysis of fecal-derived DNA, where mutant DNA may be scarce. Finally, PCPE-PCR is simple and amendable to a cost-effective and high-throughput operation.

Clearly, the primary application of the PCPE-PCR assay is for noninvasive screening of MSI-H cancers. In the case of colorectal cancer, a combination of sigmoidoscopy and stool analysis has been suggested to be a potentially cost-effective alternative to colonoscopy [Traverso et al., 2002]. Sigmoidoscopy inspects the distal colon, whereas PCPE-PCR detects MSI-H tumors, which often occur in the proximal colon. Hence, PCPE-PCR can be a part of a group of methods for screening of colorectal cancers. It should be noted that the PCPE-PCR strategy reported here is a general strategy and thus can also be used to design assays for the detection of other important MRS markers, including BAT25 and short MRS in BAX [Lacopetta et al., 1999]. If this method is used to detect a marker such as BAT 26, which is polymorphic, it is necessary to test germ line DNA to distinguish the legitimate mutations from polymorphism. In addition, PCPE-PCR can detect other DNA mutations. In the case of point mutations, our preliminary work indicated that PCPE-PCR could detect p53 mutations in a large background of normal DNA. Hence,

PCPE-PCR could potentially serve as a generally applicable method for the detection of various DNA alterations in a large background of normal DNA.

Reference

L.A. Aaltonen, P. Peltomaki, F.S. Leach, et al., “Clues to the pathogenesis of familial colorectal cancer”, Science, **260**:812-816 (1993)

D.A. Ahlquist, J.E. Skoletsky, K.A. Boynton, et al., “Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel”, Gastroenterology, **119**:1219-1227 (2000)

D.A. Ahlquist and A.P. Shuber, “Stool screening for colorectal cancer: evolution from occult blood to molecular markers”, Clin. Chim. Acta., **315**:157-168 (2002)

J. Alexander, T. Watanabe, T.T. Wu, et al., “Histopathological identification of colon cancer with microsatellite instability”, Am. J. Pathol., **158**:527-535 (2001)

T. Blondal, B.G. Waage, S.V. Smarason, et al., “A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms”, Nucleic Acids Res., **31**:e155 (2003)

C.R. Boland, S.N. Thibodeau, S.R. Hamilton, et al., “A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability

in colorectal cancer", Cancer Res., **58**:5248-5257 (1998)

A. Braun, D.P. Little and H. Koster, "Detecting CFTR gene mutations using primer oligo base extension and mass spectrometry", Clin. Chem., **43**:1151−1158 (1997)

A. Braun, D.P. Little, D. Reuter, et al., "Improved analysis of microsatellites using mass spectrometry", Genomics, **46**:18−23 (1997)

A.J. Brookes, "The essence of SNPs", Gene, **234**:177-186 (1999)

J.W.F. Catto, G. Xinarianos, J.L. Burton, et al., "Differential expression of hMLH1 and hMSH2 is related to bladder cancer grade, stage and prognosis, but not microsatellite instability", Int J Cancer, **105(4)**:484-490 (2003)

J.W.F. Catto, M. Meuth and F.C. Hamdy, "Genetic instability and transitional cell carcinoma of the bladder", BJU International, **93(1)**:19-24 (2004)

X. Chen, Z. Fei, L.M. Smith, et al., "Stable-isotope-assisted MALDI-TOF mass spectrometry for accurate determination of nucleotide compositions of PCR products", Anal. Chem., **71**:3118-3128 (1999)

F.S. Collins, M. Morgan and A. Patrinos, "The human genome project: lessons from

large-scale biology", Science, **300**:286-290 (2003)

D.N. Cooper, B.A. Smith, H.J. Cooke, et al., "An estimate of unique DNA sequence heterozygosity in the human genome", Hum. Genet., **69**:201-205 (1985)

M. Danjoux, R. Guimbaud, T.A. Saati, et al., "Contribution of microdissection for the detection of microsatellite instability in colorectal cancer", Human Pathology, **37**: 361-368 (2006)

V. Deschoolmeester, M. Baay, and W. Wuyts, "Comparison of Three Commonly Used PCR-Based Techniques to Analyze MSI Status in Sporadic Colorectal Cancer", Journal of Clinical Laboratory Analysis, **20**:52-61 (2006)

C. Dib, S. Faure, C. Fizames, et al., "A comprehensive genetic map of the human genome based on 5264 microsatellites", Nature, **380**:152-154 (1996)

W. Dietmaier and F. Hofstädter, "Detection of Microsatellite Instability by Real Time PCR and Hybridization Probe Melting Point Analysis", Lab Invest., **81**:1453-1456 (2001)

S.M. Dong, G. Traverso, C. Johnson, et al., "Detecting colorectal cancer in stool with the use of multiple genetic targets", J. Natl. Cancer Inst., **93**:858-865 (2001)

Z. Fei, T. Ono and L.M. Smith, “MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs”, Nucleic Acids Research, **26(11)**:2827–2828 (1998)

R. Fishel, “Signaling mismatch repair in cancer”, Nat Med., **5**:1239-1241 (1999)

D. Fu, K. Tang, A. Braun, et al., “Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry”, Nat. Biotechnol., **16(4)**:381-384 (1998)

I. Gonzalez-Garcia, V. Moreno, M. Navarro, et al., “Standardized approach for microsatellite instability detection in colorectal carcinomas”, J Natl Cancer Inst., **92**:544-549 (2000)

W.M. Grady, A. Rajput, L. Myeroff, et al., “Mutation of the type II transforming growth factor-h receptor is coincident with the transformation of human colon adenomas to malignant carcinomas”, Cancer Res. **58**:3101-3104 (1998)

T.J. Griffin, J.G. Hall, J.R. Prudent, et al., “Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry”, Proc. Natl. Acad. Sci., **96**:6301-6306 (1999)

I.G. Gut, W.A. Jeffery, D.J.C. Pappin, et al., “Analysis of DNA by ‘Charge Tagging’

and matrix-assisted laser desorption/ionization mass spectrometry", Rapid Commun. Mass Sprectrom., **11**: 43-50 (1997)

G. Gyapay, F. Ginot, S. Nguyen, et al., "Genotyping Procedures in Linkage Mapping", Methods, **9**:91-97 (1996)

L.A. Haff and I.P. Smirnov, "Single-Nucleotide Polymorphism Identification Assays Using a Thermostable DNA Polymerase and Delayed Extraction MALDI-TOF Mass Spectrometry", Genome Res., **7**:378-388 (1997)

L.A. Haff and I.P. Smirnov, "Multiplex genotyping of PCR products with MassTag-labeled primers", Nucleic Acids Res., **25**:3749-3750 (1997)

K.C. Halling, J. Harper, C.A. Moskaluk, et al., "Origin of microsatellite instability in gastric cancer", Am. J. Pathol., **155**:205-211 (1999)

A.M. Haydon and J.R. Jass, "Emerging pathways in colorectal-cancer development", Lancet Oncol., **3**:83-88 (2002)

C.M. Hearne, S. Ghosh and J.A. Todd, "Microsatellites for linkage analysis of genetic traits", Trends Genet., **8**:288-294 (1992)

G..S. Higgins, D.P. Little, H. Koster, “Competitive oligonucleotide single base extension combined with mass spectrometric detection for mutation screening”, Biotechniques, **23**:710−714 (1997)

J.M. Hoang, P.H. Cottu, B. Thuille, et al., “BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines”, Cancer Res., **57**:300-303 (1997)

K.C. Hung, H. Rashidzadeh, Y. Wang, et al., “Use of Paraffin Wax Film in MALDI-TOF Analysis of DNA”, Anal. Chem., **70**:3088-3093 (1998)

Y. Ionov, M.A. Peinado, S. Malkhosyan, et al., “Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis”, Nature, **363**:558-561 (1993)

D.R. Jacobson and N.E. Mills, “A highly sensitive assay for mutation ras genes and its application to the study of presentation and relapse genotypes in acute leukemia”, Oncogene, **9**:553-563 (1994)

Z. Jin, G. Tamura, M. Satoh, et al., “Absence of BAT-26 instability in gastric intestinal metaplasia”, Pathol. Int., **51**:473-475 (2001)

I.J. Kim, Y. Shin, H.C. Kang, et al., “Robust microsatellite instability (MSI) analysis by denaturing high-performance liquid chromatography (DHPLC)”, J Hum Genet.,

48:525-530 (2003)

S. Kim, J.R. Edwards, L. Deng, et al., "Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry", Nucleic Acids Res., **30**:e85 (2002)

B.J. Lacopetta, R. Soong, A.K. House, et al., "Gastric carcinomas with microsatellite instability: clinical features and mutations to the TGF-b type II receptor, IGF-II receptor and BAX genes", J. Pathol. **187**:428-432 (1999)

S. Le Hellard, S.J. Ballereau, P.M. Visscher, et al., "SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis", Nucleic Acids Res., **30**:e74 (2002)

P. Lecchi and L.K. Pannell, "6-Aza-2-thio-thymine: a matrix for MALDI spectra of oligonucleotides", J. Am. Soc. Mass Spectrum., **6**:1276-1277 (1995)

T.A. Lehman, F. Scott, M. Seddon, et al., "Detection of K-ras oncogene mutations by polymerase chain reaction based ligase chain reaction", Anal. Biochem., **239**:153-159 (1996)

J. Leushner and N.H. Chiu, "Automated mass spectrometry: a revolutionary technology for clinical diagnostics", Mol. Diagn., **5**:341-348 (2000)

J. Li, J.M. Butler, Y. Tan, et al., “Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry”, Electrophoresis, **20**:1258-1265 (1999)

M. Litt and J.A. Luty, “A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene”, Am. J. Hum. Genet., **44**:397-401 (1989)

D.P. Little, A. Braun, B. Darnhofer-Demar, et al., “Apolipoprotein E polymorphism identification using temperature cycled primer oligo base exten-sion and mass spectrometry”, Eur. J. Clin. Chem. Clin. Biochem., **35**:545−548 (1997)

D.P. Little, A. Braun, B. Darnhofer-Demar, et al., “Rapid detection of RET proto-oncogene codon 634 mutations using mass spectrometry”, J. Mol. Med., **75**:745−750 (1997)

D.P. Little, A. Braun, M.J. O’Donnell, et al., “Mass spectrometry from miniaturized arrays for full comparative DNA analysis”, Nat. Med., **3**: 1413-1416 (1997)

D.P. Little, T.J. Cornish, M.J. O’Donnell, et al., “MALDI on a Chip: Analysis of Arrays of Low-Femtomole to Subfemtomole Quantities of Synthetic Oligonucleotides

and DNA Diagnostic Products Dispensed by a Piezoelectric Pipet", Anal. Chem., **69**:4540-4546 (1997)

A. Loukola, R. Salovaara, P. Kristo, et al., "Microsatellite instability in adenomas as amarker for hereditary nonpolyposis colorectal cancer", Am. J. Pathol., **155**: 1849-1853 (1999)

A. Loukola, K. Eklin, P. Laiho, et al., "Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC)", Cancer Res., **61**:4545-4549 (2001)

N.D. MacDonald, H.B. Salvesen, A. Ryan, et al., "Frequency and prognostic impact of microsatellite instability in a large population-based study of endometrial carcinomas", Cancer Res., **60**:1750-1752 (2000)

M. Maekawa, K. Sugano, H. Kashiwabara, et al., "DNA Methylation Analysis Using Bisulfite Treatment and PCR–Single-Strand Conformation Polymorphism in Colorectal Cancer Showing Microsatellite Instability", Biochemical and Biophysical Research Communications, **262**:671-676 (1999)

L. Mao, D.J. Lee, M.S. Tockman et al., "Microsatellite alterations as clonal markers for the detection of human cancer", Proc Natl Acad Sci., **91**:9871-9875 (1994)

L. Mao, M.P. Schoenberg, M. Scicchitano et al., “Molecular detection of primary bladder cancer by microsatellite analysis”, Science, **271**:659-662 (1996)

S.J. Markowitz, J. Wang, L. Myeroff, et al., “Inactivation of the type II TGF-receptor in colon cancer cells with microsatellite instability”, Science, **268**:1336-1338 (1995)

E.R. Martin, E. Lai, J. Gilbert, et al., “SNPing Away at Complex Diseases: Analysis of Single-Nucleotide Polymorphisms around APOE in Alzheimer Disease” Am. J. Hum. Genet., **67**:383-394 (2000)

T. Miliotis, S. Kjellstrom, P. Onnerfjord, et al., “Protein identification platform utilizing micro dispensing technology interfaced to matrix-assisted laser desorption ionization time-of-flight mass spectrometry”, Chromatogr. A, **886**:99-110 (2000)

H. Murata, N.H. Khattar, Y. Kang, et al., “Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability”, Oncogene, **21(37)**:5696-5703 (2002)

L. Myeroff, R. Parsons, S.J. Kim, et al., “A transforming growth factor h receptor type II gene mutation common in colon and gastric but rare in endometrial cancers with microsatellite instability”, Cancer Res., **55**:5545-5547 (1995)

P. Nollau and C. Wagenera, “Methods for detection of point mutations: performance and quality assessment, on behalf of the IFCC Scientific Division and Committee on Molecular BiologyTechniques”, Clin. Chem., **43**:1114-1128 (1997)

E. Nordhoff, A. Ingendoh, R. Cramer, et al., “Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared”, Rapid Commun. Mass Spectrom., **6**:771-776 (1992)

S. Oda, E. Oki, Y. Maehara, et al., “Precise assessment of microsatellite instability using high resolution fluorescent microsatellite analysis”, Nucleic Acids Res., **25**:3415-3420 (1997)

H. Orum, P.E. Nielsen, M. Egholm, et al., “Single base pairmutation analysis by PNA directed PCR clamping”, Nucleic Acids Res., **21**:5332-5336 (1993)

P. Peltomaki, R.A. Lothe, L.A. Aaltonen, et al., “Microsatellite instability is associated with tumors that characterize the hereditary nonpolyposis colorectal carcinoma syndrome”, Cancer Res., **53**:5853-5855 (1993)

U. Pieles, W. Zürcher, M. Schär, et al., “Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides”, Nucleic Acids Res., **21**:3191-3196 (1993)

M.A. Rodriguez-Bigas, C.R Boland, S.R. Hamilton, et al., “A national cancer institute workshop on hereditary nonpoliposis colorectal cancer institute workshop on hereditary nonpoliposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines”, J Natl Cancer Inst., **89**:1758-1762 (1997)

P. Ross, L. Hall, I.P. Smirnov, et al., “High level multiplex genotyping by MALDI-TOF mass spectrometry”, Nat Biotechnol., **16**:1347–1351 (1998)

R. Sachidanandam, D. Weissman, S.C. Schmidt, et al., “A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms”, Nature, **409**:928-933 (2001)

K. Samara, M. Zervou, N.M. Siafakas, et al., “Microsatellite DNA instability in benign lung diseases”, Respiratory Medicine, **100**:202-211 (2006)

S. Sauer, D. Lechner, K. Berlin, et al., “A novel procedure for efficient genotyping of single nucleotide polymorphisms”, Nucleic Acids Res., **28**:e13 (2000)

S. Sauer, D. Lechner, K. Berlin, et al., “Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay”, Nucleic Acids Res., **28**:e100 (2000)

J.J. Scarisbrick, T.J. Mitchell, E. Calonje, et al., “Microsatellite Instability Is

Associated with Hypermethylation of the hMLH1 Gene and Reduced Gene Expression in Mycosis Fungoides", Journal of Investigative Dermatology, **121**: 894-901 (2003)

J. Schlegel, T. Vogt, K. Munkel, et al., "DNA fingerprinting of mammalian cell lines using nonradioactive arbitrarily primed PCR (AP-PCR)", Biotechniques, **20**:178-180 (1996)

M. Schuerenberg, C. Luebbert, H. Eickhoff, et al., "Prestructured MALDI-MS Sample Supports", Anal. Chem., **72**:3436-3442 (2000)

S.P. Siah, D.M. Quinn, G.D. Bennett, et al., "Microsatellite instability markers in breast cancer: A review and study showing MSI was not detected at 'BAT 25' and 'BAT 26' microsatellite markers in early-onset breast cancer", Breast Cancer Research and Treatment, **60**:135-142 (2000)

M. Strand, T.A. Prolla, R.M. Liskay, et al., "Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair", Nature, **365**:274-276 (1993)

X. Sun, H. Ding, K. Hung, et al. "A new MALDI-TOF based mini sequencing assay for genotyping of SNPS", Nucleic Acids Res., **28**: e68 (2000)

X. Sun, K. Hung, L. Wu, et al., “Detection of tumor mutations in the presence of excess amounts of normal DNA”, Nat. Biotechnol. **20**:186-189 (2002)

A.C. Syvänen, K. Aalto-Setala, L. Harju, et al., “A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E”, Genomics, **8**:684-692 (1990)

K. Tang, D.J. Fu, S. Kotter, et al., “Matrix-assisted laser desorption/ionization mass spectrometry of immobilized duplex DNA probes”, Nucleic Acids Res., **23**: 3126-3131 (1995)

S.N. Thibodeau, G. Bren, and D. Schaid, “Microsatellite instability in cancer of the proximal colon”, Science, **260**:816-819 (1993)

J. Tost and I.G. Gut, “Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications”, Clinical Biochemistry, **38**:335-350 (2005)

G. Traverso, A. Shuber, L. Olsson, et al., “Detection of proximal colorectal cancers through analysis of faecal DNA”, Lancet, **359**:403-404 (2002)

G. Traverso, A. Shuber, B. Levin, et al., “Detection of APC mutations in fecal DNA from patients with colorectal tumors”, N. Engl. J. Med., **346**:311-320 (2002)

A. Umar, C.R. Boland, J. Terdiman, et al., “Revised bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability”, J Natl Cancer Inst., **96**:261-267 (2004)

B. Vogelstein and W. Kinzler, “Digital PCR”, Proc. Natl. Acad. Sci. USA, **96**:9236-9241 (1999)

J.L. Weber and P.E. May, “Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction”, Am. J. Hum. Genet., **44**:388-396 (1989).

T. Wenzel, T. Elssner, K. Fahr, et al., “Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides”, Nucleotides, Nucleotides and Nucleic Acids, **22**:1579-1581 (2003)

P. Wild, R. Knuechel, W. Dietmaier, et al., “Laser microdissection and microsatellite analyses of breast cancer reveal a high degree of tumor heterogeneity”, Pathobiology, **68(4–5)**:180-190 (2000)

T. Yamada, T. Koyama, S. Ohwada, et al., “Frameshift mutations in the MBD4/MED1 gene in primary gastric cancer with high-frequency microsatellite instability”, Cancer Letters, **181**:115-120 (2002)

X.P. Zhou, J.M. Hoang, Y.J. Li, et al., “Determination of the replication error phenotype in human tumors without the requirement formatching normal DNA by analysis of mononucleotide repeat microsatellites”, Genes Chromosomes Cancer, **21**:101-107 (1998)

M. Zirvi, T. Nakayama, G. Newman, et al., “Ligasebased detection ofmononucleotide repeat sequences”, Nucleic Acids Res., **27**:e40-40 (1999)