The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Establishment of a Fast and Accurate Proteomic Method for Body Fluid/Cell Type Identification

Author: Mechthild Prinz, Ph.D., Yingying Tang, M.D., Ph.D., Donald Siegel, Ph.D., Heyi Yang, Ph.D., Bo Zhou, Ph.D., Haiteng Deng, Ph.D.

Document No.: 236538

Date Received: November 2011

Award Number: 2008-DN-BX-K011

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Report Title - Establishment of a Fast and Accurate Proteomic Method for Body Fluid/Cell Type Identification

Award Number - NIJ Grant 2008-DN-BX-K011

Authors - Mechthild Prinz, Ph.D., Yingying Tang, M.D. Ph.D., Donald Siegel, Ph.D., Heyi Yang, Ph.D., Bo Zhou, Ph.D., Haiteng Deng, Ph.D.

Abstract - Body fluid identification plays an important role in forensic investigations both for directing the course of further inquiries, as well as in contributing to the evidence in a case. Yet current methods for body fluid detection vary widely in sensitivity, reproducibility, speed, cost and above all specificity – i.e. their ability to confirm that test results are conclusive, in addition many tests are founded on techniques that are decades old and have failed to keep pace with recent advances in science and technology. Two examples are the routinely used tests for blood (Kastle-Meyer) and saliva (amylase). Both assays give false positives for a variety of naturally occurring compounds, both rely on the subjective evaluation of qualitative results based on an analyst's best judgment, and both are more than 100 years old. Yet even the more quantitative and confirmatory immunoassays, like ELISA, suffer from the inherent limitations of antibody specificity (cross-reactivity), affinity (sensitivity), narrow working ranges (the Hook effect), and manufacturing production variability, all of which can contribute to varying results within and between testing laboratories. Additionally, there are many body fluids for which no tests are available, e.g. menstrual blood and vaginal fluid.

The medley of body fluid tests with their significant variability in results (confirmatory vs. presumptive, different sensitivities, etc.) stands in contrast to DNA STR testing, which uses a single methodology on a wide variety of sample types (cells, body fluids, latent fingerprints, hair, etc.) to produce consistent results held to common accepted standards. Part of the advantage inherent of DNA testing is that it focuses on a single biopolymer – DNA, making it relative easy to hone extraction, purification, testing and analysis to a common method. Yet most confirmatory body fluid tests also rely on the detection of a single type of biopolymer – proteins. And although protein chemistry is more variable and complex than that of DNA (there are twenty different amino acids each with distinct chemical properties compared with only four nucleotides for DNA), scientific and technical advances in protein chemistry and analysis now bring it within range of common methods and standards for forensic body fluid identification.

The overall objective of this project was to determine if a single, confirmatory methodology, employing mass spectrometry (**MS**) as the means of detection, could be developed for the

identification of five forensically important body fluids – blood, salvia, semen, menstrual blood and vaginal fluid. Toward that end specific aims were established to: 1) identify multiple protein markers for each body fluid, 2) determine the limits of detection for each body fluid, 3) establish a single extraction procedure that works well for all body fluids from a variety of substrates (cotton, polyester, condom, common beverage plastic and wood), 4) determine limits of detection for each body fluid in mixed samples, 5) evaluate aged mock forensic samples, and 6) establish a high throughput assay. Our results demonstrate the value of mass spectrometry as a routine tool for body fluid identification. Multiple markers were identified for blood, saliva and semen, and a common method developed for their extraction from all five substrates. Detection levels for all three body fluids were in the nanoliter range – lower than other methods currently in use. Body fluids mixtures could be distinguished, and aging for up to 20 months did not appreciably affect sample detection. Preliminary results suggest that all three body fluids are amenable to high throughput testing – we estimate that with the single robotic liquid handling device currently available in our laboratory (Biomek FX) we should be able to process 600-700 samples/week from extraction through identification. We also identified multiple markers in menstrual blood (five) and vaginal fluid (seven), however, because protein expression is known to vary in response to hormone levels, further investigation of these potential markers throughout the menstrual cycle is required.

Our results demonstrate that MS can be used as a single confirmatory body fluid assay that could replace the current variety of test methods and as such would simplify standardization for body fluid testing across laboratories.

Table of Contents

Abstract 1

Executive Summary 4

Introduction 14

Methods 19

Results 31

Discussion and Conclusions 55

References 58

Dissemination of Research Findings 61

Abbreviations

Alb - albumin
AMY1 - alpha-amylase 1
C6orf58 - chromosome 6 open reading frame 58
CHCA - alpha-cyano-4-hydroxycinnamic acid
CRISP1 - Cysteine-rich secretory protein 1
CST2 - cystatin-SA
ELISA - enzyme-linked immunosorbent assays
Hb - hemoglobin
HBA - alpha-hemoglobin
HBB - beta-hemoglobin
HBD 1 - human beta-defensins 1
LC - liquid chromatography
IEF - isoelectric focusing
IgG - immunoglobulin G
Igs - immunoglobulins
IPG - immobilized pH gradient
KLK3/PSA - kallikrein-related peptidase 3, orprostate-specific antigen
MALDI - matrix assisted laser desorption ionization
MMP - matrix metalloprotease
MS - mass spectrometry
MUC - mucin
PAEP - progestagen-associated endometrial protein
PAGE - polyacrylamide gel electrophoresis
PBGB - porphobilinogen deaminase
PETE - polyethylene terephthalate
SDS – sodium dodecyl sulfate
SEMG - semenogelin
SLC4A1- band 3 anion transport protein
SPTA1 - alpha-spectrin, erythrocytic 1
SPTB - beta-spectrin
TFA - trifluoroacetic acid
TOF - time-of-flight

Executive Summary

1. Statement of the Problem – Body fluid identification plays an important role in forensic investigations in both directing the course of further inquiries, as well as in contributing to the evidence in a case. Yet the science and technology behind body fluid identification has not kept pace with the forensic DNA revolution for individual identification. And while an individual can be identified from the DNA in their body fluids, DNA testing for body fluid identification through cytosine methylation is still in its infancy with significant challenges both in terms of identifying body fluid specific markers as well as techniques that are not destructive to DNA itself (see Literature Review below). However, because body fluids have evolved to perform different functions, they contain within them different proteins, or different levels and combinations of proteins that give each body fluid a unique protein signature that can be used to distinguish one body fluid from another. Hemoglobin in blood and amylase in saliva are typical examples. Yet the techniques routinely used to identify different body fluid proteins vary dramatically in sensitivity, reproducibility, speed, cost and above all specificity – i.e. their ability to confirm that test results are conclusive. Neither the hemoglobin (Kastle-Meyer) nor amylase tests are confirmatory; both give false positives for a variety of naturally occurring compounds, both rely on the subjective evaluation of qualitative results based on an analyst's best judgment, and both are more than 100 years old. Even the more quantitative and confirmatory immunoassays like ELISA, suffer from the inherent limitations of antibody specificity (cross-reactivity - therefore false positives), variable antigen affinity (therefore low sensitivity and/or false negatives), tight antigen-antibody concentration ratios (possible false negatives – the Hook effect) and manufacturing production variability (therefore loss of assay reproducibility), as well as high costs and long assay times. Additionally, there are many body fluids for which no tests are available, e.g. menstrual blood and vaginal fluid.

Unlike DNA testing, which relies on a single, dominant technology for testing all samples regardless of their source, there is no single uniform methodology to simultaneously evaluate an unknown forensic sample for all possible body fluids. Consequently, each body fluid test must be performed separately, consuming time and evidence, as well as requiring multiple instruments and laboratory personnel trained in the nuances of each method. Right now, most body fluid tests are qualitative, relying on the subjective evaluation of analysts – e.g. the clearing of an opaque field around a suspected amylase sample in an agarose gel. In short, a single reproducible, quantitative test that can identify all body fluids simultaneously is lacking.

What would constitute an ideal body fluid assay? One able to detect body fluid fingerprint proteins accurately, reproducibly, fast, with high sensitivity, inexpensively and above all, confirmatory. Peptide identification by mass spectrometry (**MS**) offers all these advantages.

MS identifies peptides by determining their unique amino acid sequences; consequently body fluid marker peptide identification by MS is a confirmatory assay. Multiple markers for each body fluid increase the confidence of correct identification. The ability to identify multiple body fluid markers in a unknown sample allows for the detection of mixtures. Most importantly, the MS assay does not need to be tuned or changed for each body fluid; rather, it is unbiased and identifies whatever is present in a sample. This means that a single assay, using the same extraction, purification and detections methods can be performed for all body fluids, eliminating multiple techniques, saving time and money, and giving the same confirmatory results for all samples. Here we report our progress in developing such an assay.

2. Purpose - The overall goal of this work was to establish an accurate, sensitive, reproducible, robust and confirmatory assay for the identification of five forensically important body fluids (blood, saliva, semen, menstrual blood and vaginal fluid) using a single methodology. Because body fluids can be distinguished by their protein content, experiments were designed to identify unique protein markers for each body fluid, as well as common extraction, purification and detection techniques. As part of the overall goals, experiments were also designed to measure assay sensitivity, quantify markers, distinguish mixtures, and test mock forensic samples. The rationale behind these experiments is described below in *Research Design*; our results are presented in *Findings and Conclusions*.

3. Research Design - As described above, the goal of this work was to establish a single, confirmatory method for the identification of five different body fluids based on the specific protein markers present in each, and their detection by mass spectrometry. The experimental design for this project was extensive, beginning with body fluid marker selection through methods for marker extraction, detection as well as overall assay evaluation. Consequently, this section will be divided into five subsections aimed at addressing the organizational rationale of each major component. However, as this project is based on protein detection by mass spectrometry, it will begin with a short description of mass spectrometry as a detection tool and the criteria used for marker peptide selection.

3.1 ASSAY PRINCIPLES – How mass spectrometry can be used for body fluid detection and how body fluid protein makers are selected.

● PEPTIDES & MASS SPECTROMETRY – Proteins are made up of long strings of amino acids. There are twenty different amino acids in the protein alphabet that can be linked to one another in a nearly infinite number of combinations giving each protein a unique sequence. Consequently, proteins can be distinguished from one another if their amino acid sequence is known. One way of determining a protein's sequence is by digesting it into smaller peptide fragments and measuring their masses. As each peptide fragment is composed of a unique set of amino acids, and as amino acid has a unique mass, the peptide's mass can be used to determine its sequence. Confirmation of a peptide's sequence can be determined by breaking it into smaller pieces and measuring their masses. While this may sound like a complex and time-consuming assay, in fact measurements are made in a matter of seconds and sequences determined equally as fast by comparing results against computer databases. The instrument used to measure peptide fragment masses is a mass spectrometer. Because a mass spectrometer can measure any peptide's mass, this single technique can determine whether an unknown sample is blood, saliva, semen or something else without need of any other reagents or methods. It is confirmatory, and we have shown, more sensitive than other body fluid detection techniques currently in use.

● SELECTION OF BODY FLUID SPECIFIC PEPTIDE MARKERS – Body fluids are composed of hundreds of proteins not all of which are unique to any particular body fluid. Immunoglobulins, for example, are found abundantly in blood, saliva and vaginal fluid. Additionally, proteins, even body fluid specific proteins, are made up of numerous peptides, not all of which are equally useful as markers as some are more easily detected than other. The goal of this work was to identify, first through the literature, then through empirical experimentation, ideal body fluid protein markers. An ideal marker should have the following characteristics: i) it should be unique or highly enriched in the body fluid of interest (hemoglobin in blood is a good example), ii) it should be abundant, as this will increase the sensitivity of the assay as more abundant proteins can be detected in smaller sample volumes, and iii) the peptides from the chosen protein marker should be easily detected by mass spectrometry – different peptides have different ionization efficiencies and therefore different levels of detection.

As stated, initial marker selections were based on information in the literature, however, following experiments performed in the lab some of these markers changed. This was particularly true for menstrual blood and vaginal fluid for which original marker identification in

the literature was based on the presence of mRNAs for the marker proteins, and for which we were unable to find any proteins.

3.2 EXTRACTION & PURIFICATION – Here experiments were designed to identify a single buffer that could extract and solubilize all body fluids from a variety of different substrates. The major components evaluated were: buffer types, ionic strength, ionic and non-ionic detergents, and chaotropic salts. Five different buffers were tested against mock forensic samples dried onto five different substrates (cotton, polyester, condom, wood and common plastic beverage container (polyethylene terephthalate)). Water was used as control.

3.3 ASSAY SENSITIVITY – Experiments to test assay sensitivity were designed to determine the smallest amount of each body fluid that could be detected. This was tested under two different scenarios. In one scenario body fluid proteins were purified by a variety of techniques prior to detection by MALDI MS; this method was the most sensitive. The second set of experiments was designed to determine detection limits from unpurified samples, which, if sufficiently sensitive, would allow for significantly faster sample throughput. Short descriptions of each method follow, detailed narratives and results are found in the main body of this report.

- Protein Purification Method (PAGE-LC-MALDI, IEF-LC-MALDI, or LC-MALDI): Three different purification techniques were tried based on different separation principles: isoelectric focusing (**IEF**) which separates proteins based on their electric charge, polyacrylamide gel electrophoresis (**PAGE**) which separates proteins based on their size, and reverse phase liquid chromatography (**LC**) which separates peptides (digested proteins) based on their hydrophobicity. Separation of proteins by either IEF or PAGE was followed by protein digestion and LC. Direct LC of digested samples without prior separation was also tried.
- Direct Detection of Unpurified Samples (direct spotting MALDI or MALDI): In the direct detection method, extracted body fluid samples were not purified but digested and markers directly measured by MS.

3.4 MIXTURE DISCRIMINATION – One goal of this project was to determine the limits of detection of individual body fluid markers in mixed samples. Mixture experiments were carried out by combining different body fluids at varying ratios, purifying proteins by LC, and measuring marker peptides.

3.5 AGED FORENSIC SAMPLES – As forensic samples age, their biological content can degrade. Here experiments were designed to determine the effects of aging on the MS assays. Mock aged samples were prepared by spotting known quantities of body fluids on five different substrates (cotton, polyester, condom, wood and common plastic beverage container (polyethylene terephthalate)) and allowing them to age for varying lengths of time up to 20

months. Two methods for detecting the effects of aging were performed. The first used SDS PAGE gels to evaluate overall quality by comparing the total proteins extracted from aged samples to those extracted from freshly prepared samples. In the second method, body fluid specific protein markers were evaluated. In this method, aged samples and freshly prepared controls were extracted, digested, separated by LC and body fluid specific peptides compared for degradation by MS.

3.6 MENSTRUAL BLOOD AND VAGINAL FLUID - As described above, the original markers selected for menstrual blood and vaginal fluid were based on published reports of the presence of mRNAs for these proteins in both body fluids. Extensive analysis by PAGE, IEF, LC as well as other techniques (see *Findings* below) revealed that none of the proteins represented by these mRNAs were detectable in either body fluid. Consequently, experiments were undertaken to identify new protein markers. Two different methods were used.

i) Vaginal Fluid Markers: Because of growing medical interest in the vaginal fluid proteome for both reproductive health issues and cancer diagnosis and therapy, an increasing number of publications and online databases have begun to establish a library of vaginal fluid proteins. Like most body fluids, the vast majority of vaginal fluid proteins are not unique. However, by comparing our database of vaginal fluid protein with those identified by others, and by cross checking the combined vaginal fluid proteome with the proteomes of other body fluids, we were able to identify several candidate markers.

ii) Menstrual Blood Markers: Unlike vaginal fluid, there are no publications or databases for menstrual blood proteins. Consequently, we undertook a complete proteome analysis of menstrual blood using multiple techniques including PAGE, IEF and LC. Here too we identified several candidate markers. We are currently preparing a manuscript describing the menstrual blood proteome.

4. Findings and Conclusions – This section describes the progress we have made in achieving the goals of this grant, i.e. establishing an assay for the simultaneous identification of different body fluids using a single, confirmatory method. Our original goals were to 1) identify multiple protein markers for each of five body fluids (blood, saliva, semen, menstrual blood and vaginal fluid); 2) determine the limits of detection for each body fluid, 3) establish a single extraction procedure that works well for all body fluids, 4) determine limits of detection for each body fluid in mixed samples, 5) evaluate aged mock forensic samples, and 6) establish a high throughput assay. Progress made toward achieving these goals is discussed below for blood, semen, and saliva. Progress toward identifying new markers for menstrual blood and vaginal

fluid follow in section 4.7. Section 4.8 summarizes findings and presents conclusions. Details of experimental design and outcomes can be found in the METHODS and RESULTS sections in the main body of this report.

4.1 IDENTIFYING MARKER PEPTIDES FOR BLOOD, SALIVA & SEMEN – Our original goal was to identify three proteins that could be used as markers for each body fluid. Criteria for markers were that they had to be unique to or highly enriched in each body fluid, and they should be abundant so as to improve detection sensitivity – the more abundant a marker the smaller the amount of sample that could be detected. Using samples collected from volunteers, proteins for each body fluid were extracted and purified by either PAGE or IEF. Purified proteins were digested, peptides separated by LC and detected by MALDI TOF/TOF mass spectrometry. As may be seen in the table below five marker proteins were identified in blood, four in saliva and six in semen. All met our criteria for body fluid markers.

Body Fluid	Identified Biomarker Protein
Blood	Hemoglobin alpha (HBA) Hemoglobin beta (HBB) Beta-spectrin (SPTB) Alpha-spectrin (SPTA1) Band 3 anion transport protein (SLC4A1)
Saliva	Alpha-amylase 1 (AMY1) Cystatin SA (CST2) Uncharacterized protein C6orf58 (C6oor58) Histatin-1 (HTN1)
Semen	Semenogelin-2 (SEMG2) Semenogelin-1 (SEMG1) Mucin 6 (MUC6) Prostate specific antigen (PSA) Progestagen-associated endometrial protein (PAEP) Cysteine-rich secretory protein 1 (CRISP1)

In order to evaluate the detection efficiency of peptide markers and perform quantitative analyses on extraction loss, mixture ratios, and other parameters, we developed and tested a panel of isotopically labeled peptide markers.

4.2 DETERMINING ASSAY SENSITIVITY – Experiments were performed to determine the smallest volumes of different body fluids that could be detected. Samples were solubilized, serially

diluted, digested and peptides purified by LC prior to MALDI MS. Our results (Figure 4 in the RESULTS section of this report) demonstrate that the LC-MALDI method has equal or greater sensitivity than chemical, immunochemical, enzymatic and mRNA methods. Our marker proteins (Table 2, RESULTS) were easily identified from 2 ng of blood and semen, and 10 ng of saliva, corresponding to 0.006 nl blood, 0.1 nl of semen and 4 nl of saliva.

4.3 ESTABLISHING A SINGLE EXTRACTION PROCEDURE FOR ALL BODY FLUIDS – Experiments designed to identify a common buffer that could be used to extract the maximum amount of marker proteins from all body fluids deposited on different substrates (cotton, polyester, wood, condom and common plastic beverage container) were evaluated for buffer type, ionic strength, ionic and non-ionic detergents and chaotropic salts. Results showed that ionic detergents (SDS) and chaotropic salts (urea/thiourea) were most effective for protein extraction. There use, however, is dependent on which downstream protein purification method is used. A single buffer that showed uniform extraction results for all body fluids from all substrates, regardless of the downstream purification method, was RIPA buffer, which contains modest amounts salts, ionic and non-ionic detergents (Figure 3, RESULTS), and was the buffer of choice for both pure and mixed samples.

4.4 DETECTING BODY FLUIDS IN MIXED SAMPLES: Mixtures of body fluids dramatically increase the complexity of the proteins present in a sample and simultaneously dilute target proteins. In addition, because some body fluids can be detected in much smaller volumes than others, for example blood can be detected in minute quantities, 0.0003 μl, while saliva requires 0.5 μl, the amount of blood that can be detected in saliva is much different from the amount of saliva that can be detected in blood. To determine the levels at which body fluids can be identified in mixtures, we prepared varying ratios of different body fluids (e.g. blood/saliva 100/1 and saliva/blood 100/1) and tested them by LC MALDI MS. Limits of detection can be seen in the table below; for example, blood can be detected in a saliva/blood ration of 1/1000, whereas saliva can only be detected in a blood/saliva ration of 1/1. Full experimental details can be found in Table 5 in the RESULTS section of this report.

Body Fluid Mixtures	Ratios (Volume) at Which Both Body Fluids Can be Detected
Blood/Saliva	1 / 1
Saliva/Blood	1000 / 1
Blood/Semen	5 / 1

Semen/Blood	200 / 1
Saliva/Semen	500 / 1
Semen/Saliva	10 / 1

4.5 EVALUATION OF PROTEIN DEGRADATION IN AGED MOCK SAMPLES: Experiments designed to determine the effects of aging on body fluids were described above in section 3.5 above. In short, the samples that were tested were aliquots of blood, saliva and semen that had been spotted on cotton and allowed to age for up to 20 months. To determine the effect of aging, two methods were used. The first evaluated total proteins by SDS PAGE. Gels were stained, scanned and bands quantified using ImageJ software from NIH. The second method evaluated specific marker proteins by sample extraction, digestion, LC and MALDI TOF/TOF. Equal amounts of freshly prepared body fluid samples were run alongside aged samples as positive controls. As may be seen in Figure 5 and Figure 6 in the RESULTS Section of this report some protein degradation occurred in all body fluids. However, aged body fluids could still be identified by MALDI TOF/TOF at similar sensitivity levels as freshly prepared controls.

4.6 ESTABLISH A HIGH THROUGHPUT ASSAY – Maximum achievable MALDI TOF/TOF assay sensitivity was obtained using the most thorough conditions – protein and peptide purifications followed by MS detection of multiple peptide markers for each body fluid. While these conditions are extremely sensitive and identify multiple marker peptides for each body fluid, they require overnight protein and peptide purifications. To determine if body fluid identification could be done faster and without sacrificing sensitivity, preliminary experiments were performed on unpurified samples that were digested and directly analyzed by MALDI TOF/TOF. To determine the limits of detection for this method, serial dilutions of unpurified digested blood, semen and saliva samples were directly assayed by MALDI MS at following total protein levels: 100 ng, 50 ng, 20 ng, 10 ng, 5 ng, and 2 ng. Peptides for semen marker semenogelin-2, blood markers hemoglobin A and B, and the saliva marker α-amylase 1 could be detected at the 5 ng level. Thus, while levels of detection were close to those achieved by the 2D protein/peptide purification method, the numbers of peptides identified were significantly less. However, we estimate that this method could be developed to process approximately 600-700 samples per week, at reagent/consumable costs of < $1.00/sample. Work toward achieving this goal is currently underway.

4.7 MENSTRUAL BLOOD & VAGINAL FLUID – As described above, the original markers we planned to use to identify menstrual blood and vaginal fluid were based on reports in the literature of the presence of their mRNAs in these body fluids. However, extensive evaluation by IEF, PAGE, and LC were unable to identify any of the protein products of these mRNAs. Additional experiments were tried in an attempt to enrich these proteins, including: centrifugal fractionation, hemoglobin, albumin and immunoglobulin depletion, and protein deglycosylation followed by IEF and LC. Here too, none of the marker proteins could be found. Consequently, we performed complete proteome analysis of both menstrual blood and vaginal fluid to identify candidate markers that could be used to establish assays for both body fluids. A total of 1,067 proteins were identified in menstrual blood and 687 in vaginal fluid. These proteins were then cross-referenced against protein libraries for blood, saliva and semen, as well as several other body fluids, and candidate markers selected on the basis of their uniqueness to, or enrichment in menstrual blood or vagina fluid. Five potential markers were identified in menstrual blood and seven in vaginal fluid, see table below. Experiments are planned to confirm these data from samples from multiple individuals, over the entire course of their menstrual cycles and representing diverse ages and ethnicities - as all these factors can influence protein expression levels (see Section IV *Conclusions/Implications for Future Research* in the main body of this report). It should be noted that to our knowledge, ours is the first menstrual blood proteome analysis to be performed, and a manuscript is in preparation for publication.

Body Fluid	Protein Name
Menstrual Blood	Hemoglobin subunit zeta (HBZ)
	15-hydroxyprostaglandin dehydrogenase [NAD+] (HPGD)
	Progestagen-associated endometrial protein (PAEP)
	Mucin-5B (MUC5B)
	Proteasome subunit beta type-5 (PSMB5)
Vaginal Fluid	small proline-rich protein 3 (SPRR3)
	Dimethylaniline monooxygenase 3 (FMO3)
	Transmembrane protease, serine 11B (TMPRSS1B)
	Cartilage intermediate layer protein 1 (CILP1)
	Cartilage intermediate layer protein 2 (CILP2)
	Beta-crystallin S (CRYGS)

	Isoform 1 of A-kinase anchor protein 13 (AKAP13)

4.8 CONCLUSIONS: We have developed a highly sensitive, confirmatory protein-based body fluid assay that can detect blood, semen and saliva simultaneously using a single, common method. Levels of detection, based on volume, are in the nanoliter range for all three body fluids. Samples aged for periods up to twenty months show little evidence of protein degradation, and can be detected with similar sensitivity to freshly prepared samples. In addition, even partially degraded marker peptides can be used for body fluid identification. Identification of body fluids in mixed samples is dependent on the protein content of the body fluids under investigation, and the ratio of the mixture. Here, for example, we demonstrated that blood can be detected in saliva when the sample volume ratios is 1:1000 respectively, while, saliva can be detected in blood only at a ratio of 1:1. Significantly improved sensitivities could be achieved by introducing protein purification methods upstream of detection.

For daily routine testing of individual body fluids, experiments were initiated to develop a high throughput assay that can detect one or more body fluid markers without any prior purification techniques. Initial results are promising, and we estimate this method could process between 600-700 samples/week at an estimated reagent/consumable cost of <$1.00/sample, making this confirmatory method competitive with current, and often non-confirmatory, methods. Finally, progress was made toward establishing mass spectrometry assays for menstrual blood and vaginal fluid. Five candidate protein markers were identified for menstrual blood and seven for vaginal fluid. Work on a high throughput assay, as well as menstrual blood and vaginal fluid assays are being pursued with our new grant.

In summary, our results demonstrate that mass spectrometry is a robust, sensitive and above all confirmatory assay that can identify in a single test whatever body fluid is present in an unknown sample, thus eliminating the multiple methodologies currently needed and reducing the assay time, cost and the amount of sample consumed.

5. Implications for Policy and Practice – Current methods used for body fluid identification encompass a wide range of technologies including chemical, biochemical, immunochemical enzymatic, spectrophotometric and others. Some tests take advantage of modern scientific advances such as monoclonal antibodies, while others use chemical technologies that are a many decades old (e.g. Kastle-Meyer, Takayama), some are confirmatory, others, presumptive, some sensitive, other not, some rapid, other require overnight incubations. This assortment of

techniques presents practical laboratory as well as policy problems, as courts and juries come to expect test results that meet the high standards set by DNA testing.

The practical laboratory problems are many; some tests require biological activity, meaning that sample degradation in the field or in storage limit their usefulness. The need to perform multiple tests on an unknown sample to find out which body fluid is present is both time and sample consuming and requires multiple and often expensive instrumentation in the lab, as well as trained personnel to operate them. Finally, with progress that has been made in the field of proteomics, it is not necessary to settle for presumptive body fluid testing. The demonstration that a single methodology, mass spectrometry, can rapidly, and cost effectively, identify all body fluids in a sample simultaneously and confirmatively will have implications for both the policy and practice of forensic science. The 2009 National Academy's report on forensic science has identified the need to bring all fields of forensics to the level of quality of DNA testing, and policies for common standards and practices for body fluid testing are inevitable. Because mass spectrometry identifies the amino acid sequences of body fluid specific peptides, it offers a high level of certainty for the results. Also, as the use of MS becomes more prevalent, its methods more streamlined and user friendly, sample extraction and preparations "kits" may become available and in our opinion mass spectrometry may become, in practice, the gold standard of body fluid testing.

MAIN BODY OF THE FINAL TECHNICAL REPORT

I. INTRODUCTION

A. STATEMENT OF PROBLEM – Body fluid identification plays an important role in forensic investigations, yet the methods used for body fluid detection have for the most part failed to keep pace with scientific and technological advances. As a consequence, current testing is a diverse mixture of techniques that vary dramatically in sensitivity, reproducibility, speed, cost and above all specificity – i.e. their ability to confirm that test results are conclusive. For example, neither the hemoglobin (Kastle-Meyer) nor amylase tests are confirmatory; both give false positives for a variety of naturally occurring compounds, both rely on the subjective evaluation of qualitative results based on an analyst's best judgment, both are more than 100 years old, and both are in common use today. Even more quantitative and confirmatory immunoassays, like ELISA, suffer from the inherent limitations of antibody specificity, affinity, narrow working ranges (i.e. antigen-antibody concentration ratios – the Hook effect), and manufacturing production variability, as well as high costs and long assay times. Additionally,

there are many body fluids for which no tests are available, e.g. menstrual blood and vaginal fluid.

Unlike DNA testing, which relies on a single, dominant technology for testing all samples regardless of their source, there is no single uniform methodology to simultaneously evaluate an unknown forensic sample for all possible body fluids. Consequently, each body fluid test must be performed separately, consuming time and evidence, as well as requiring multiple instruments and laboratory personnel trained in the nuances of each method. In addition, most body fluid tests are qualitative, relying on the subjective evaluation of analysts – e.g. the clearing of an opaque field around a suspected amylase sample in an agarose gel, or the relative appearance of a band of an immuno stick. In short, a single reproducible, quantitative test that can identify all body fluids simultaneously is lacking.

What would constitute an ideal body fluid assay? A single method able to simultaneously detect all body fluids accurately, reproducibly, rapidly, with high sensitivity, at low cost and above all, it must by confirmatory. The unique protein composition of body fluids offers the ideal set of markers, and mass spectrometry (**MS**) offers the ideal technology.

MS identifies peptides by determining their unique amino acid sequences; consequently body fluid marker peptide identification by MS is a confirmatory assay. Multiple markers for each body fluid are available and thus increase the confidence of correct identification. The detection of different body fluid markers in an unknown sample allows for the detection of mixed samples. Most importantly, the MS assay does not need to be tuned or changed for each body fluid; rather, it is unbiased and identifies whatever proteins are present in an unknown sample.

Adoption of MS for body fluid detection would result in a single, uniform assay, using the same extraction, purification and detections methods for all body fluids, eliminating multiple techniques, saving time, money and sample, and giving the same confirmatory results for all body fluids. Above all, it would allow for the establishment of common, uniform standards throughout the forensic community in a manner similar to what is now in place for DNA testing. Here we report our progress in developing such an assay.

B. LITERATURE REVIEW AND CITATIONS – Many of the standard techniques currently used for body fluid identification are labor-intensive, technologically diverse and costly in terms of money, time and sample. In addition, these techniques often rely on chemical or enzymatic reactions that decline as samples age and can be mimicked by contaminating organic or inorganic materials. Finally, many of these tests are presumptive rather than confirmatory. A review of current methodologies for body fluid identification follows.

• ***COLORIMETRIC / CRYSTAL FORMATION ASSAYS -*** Identification of blood is typically carried out by catalytic colorimetric assays (e.g. phenolphthalein, benzidine) or crystal formation tests (Teichmann, Takayama). While the colorimetric assays (presumptive) can be fairly sensitive, they lack specificity. False positives can result from the presence of chemical oxidants and catalysts (including rust), as well as from vegetable and animal matter (e.g. apple, potato, blackberry, saliva, mucus, and others, Saferstein 1982-1993). Crystal tests (confirmatory) require, by today's standards, relatively large amounts of material (0.1 mg hemoglobin), and crystal formation can decline with age of the sample (Takayama), or when samples are exposed to substances that cause hematin to lose iron (Teichmann). A negative result does not confirm the absence of blood. Species cannot be determined by this method.

• ***ENZYMATIC ASSAYS -*** Enzymatic assays are commonly used to identify semen and saliva. The semen assay relies on acid phosphatase, while the saliva assay relies on amylase. Both assays are presumptive, and suffer from limitations inherent on relying on enzyme activities that decline over time and can be accelerated by environmental conditions that denature or degrade proteins (e.g. heat, chemicals, drying, etc., Ende 1961; Kohn 1986). In addition, alkaline phosphatase activity (semen) is also found in other body fluids, notably vaginal fluid, and consequently is not definitive evidence for semen even if found. The absence of amylase activity does not preclude the presence of saliva. Neither test can determine species.

• ***MICROSCOPIC OBSERVATIONS -*** While microscopic identification of sperm cells is confirmatory, stains and body fluid do not always reveal their presence. This can be due to low or no sperm numbers in the ejaculate because of hereditary or pathological conditions or elective surgery, as well as the difficulty of recovering intact sperm from fibers. Thus, failure to identify semen by microscopic examination is not conclusive for its absence (Saferstein 1982-1993).

• ***IMMUNOCHEMICAL ASSAYS -*** Enzyme-linked immunosorbent assays (**ELISA**) for specific body fluid markers (e.g. PSA in semen) address some of the potential problems outlined above since they do not require body fluid enzymes be active for detection and are confirmatory. However, limitations do exist; antibodies must be highly specific with strong binding affinities and antigenic sites must be present in appropriate conformations (i.e. not denatured or degraded). Because antibodies are products by living organisms, there can be significant batch-to-batch variations in binding affinities which requires assays to be recalibrated after each new lot is validated. False positives are not unknown, and for reliable results material adjacent to a stain should be tested as "environmental epitopes" and non-specific binding of primary antibodies to such materials can occur. In addition, as with all of the above assays, ELISA testing is done on one body fluid at a time. That is, a forensic sample of unknown body fluid composition must be divided and

tests carried out separately for each. In 2009 the New York City Office of Chief Medical Examiner processed a total of 68,159 samples for body fluid identification (11,756 Kastle-Meyer presumptive blood tests, 6,961 amylase presumptive saliva tests, and 12,035 p30 presumptive and 12,719 sperm searches or AP tests). This type of consecutive testing is time and sample consuming, and expensive. Finally, while species identification can be performed by ELISA, it requires production of individual antibodies for each species tested with all aforementioned limitation for each antibody. Importantly, there is no current, routine test for vaginal fluid and menstrual blood.

• ***MESSENGER RNA ASSAYS*** **-** In view of these limitations, body fluid mRNA expression profiling methods have been developed (Juusola and Ballantyne 2005; Nussbaumer, Gharehbaghi-Schnell et al. 2006; Juusola and Ballantyne 2007; Bauer and Patzelt 2008, Haas and Ba¨r et al. 2008; Haas and Kratzer et al. 2009; Fleming and Harbison 2009; Hanson and Ballantyne et al. 2009; Zubakov and Kayser et al. 2010). Juusola and Ballantyne, for example, recently described a multiplex RT-PCR assay which can detect multiple body fluid-specific mRNAs and is optimized for detection of blood, saliva, semen, vaginal secretions and menstrual blood from single or mixed stains.

There are, however, several potential drawbacks to this method that may limit its usefulness. Primary among these is the stability of RNA and the ubiquitous nature of RNases which digest it. RNases are generally small, stable molecules that function at neutral pH, and can be difficult to inactivate (Gilman 2002; Brown, Mackey et al. 2004). RNases are not only present within cells, where they function in cellular metabolism, but are part of the body's defense mechanism and are secreted into such body fluids as tears, blood, saliva and perspiration (Blank, Dekker et al. 1981, Ambion Technical Bulletin 178). They are also produced by microorganisms. Studies have shown that RNA stability varies between tissues (Inoue, Kimura et al. 2002; Nussbaumer, Gharehbaghi-Schnell et al. 2006) and is dependent on storage conditions (Tanner, Berk et al. 2002). Consequently, digestion of mRNA by RNases within the cell, from their presence in body fluids, and from possible contamination during evidence collection through analysis from perspiration (fingertips) and/or microorganisms, has the potential to limit the usefulness of this method. Even partial degradation of a sample can detrimental to the assay, as the entire region of amplification from the 5' primer to the 3' primer must be intact for the assay to be effective. In addition, common polymorphisms at primer binding sites can reduce or eliminate amplification.

Another potential problem is that not all body fluid "specific" markers are unique to a body fluid, although they may be predominantly expressed there (see e.g. tissue/body fluid distribution of kallikreins (Shaw and Diamandis 2007)). This is particularly a problem for PCR

assays which can amplify extremely small amounts of nucleic acids, and therefore potentially detect signals from contaminating tissues/body fluids. Consequently, quantitative RT-PCR assays using several markers for each body fluid will need to be established in order to construct body fluid "fingerprints." This will require more samples and will increase the complexity of the multiplex assay system.

Importantly, mRNA body fluid assays sensitivity appears to be nearly one order of magnitude lower than what we have attained by MALDI mass spectrometry for blood, saliva and semen (Haas et al. 2008; Haas and Kratzer et al. 2009) see Figure 4.

- ***DNA METHYLATION ANALYSIS FOR FORENSIC TISSUE IDENTIFICATION -*** Recent evidence from the literature suggests that DNA methylation patterns may be different in different tissues (Eckhardt, F. et al. 2006, Hong, S-J. et al. 2009, Ghosh, S. et al. 2010, Bustos, C.D. et al., 2009) and that these differences may be used to establish a forensic DNA methylation assay for tissue identification (Frumkin, D., et al, 2010epub). The advantages of such an assay would be that it would uses amplification and detection technologies already available in DNA forensic labs, that DNA is less labile than RNA (which has also been suggested for tissue identification, see above), and that it would consume only a small portion of a sample being processed for STR typing. While there is promise for this method, there are also many hurtles that need to be overcome.

The principle of the test - that it is possible to identify unique methylation patterns for each tissue and body fluid - will require significant testing in very large numbers of individuals. This is because it is already well established that methylation patterns differ with: age (Gronniger, E. et al. 2010), gender (El-Maarri, O. et al. 2007), ethnicity (Adkins, R.M. et al. 2011epub), health and disease (Egger, G., et al. 2004, Andraos. C. et al. 2011), as well as environmental conditions (Vercelli, D. 2004) and diet (Choi, S-W. et al. 2010). Methylation patterns can also differ between the same tissues in different individuals as well as within the same tissue of a single individual (Egger, G., et al. 2010). Consequently, it will be necessary to identify "invariable" tissue/body fluid specific methylation sites so that a unique tissue/body fluid specific pattern can be routinely recognized across all the variables described above.

In addition, of the three technologies commonly in use for methylation detection: endonuclease-based, affinity-based and bisulfite-based, the latter is the current gold standard. However, methylcytosine conversion (i.e. C to U) does not currently reach 100%, and DNA degradation during bisulfite treatment can be as great as 90%. Such extensive degradation can be problematic when DNA input in limited (Grunau C, et al. 2001, Ehrich M, et al. 2007).

Consequently, for this technique to be used for routine forensic testing, conversion and recovery rates will need to be improved.

• ***BODY FLUID ANALYSIS BY MASS SPECTROMETRY* –** Mass spectrometry (**MS**) is an accurate, reliable and sensitive method for simultaneously determining the identity of all proteins in a sample without a priori knowledge of their source. This means that no special extraction or preparation conditions, or reagents (e.g. antibodies or primers) are required to identify different body fluids and, since all identifying information is inherently present in a sample (known or unknown), the mass spectrometer determines body fluid identity by determining which proteins are present.

The use of MS for body fluid analysis is well established. Because of the clinical importance of evaluating easily obtainable body fluids for diagnosis, treatment and evaluating the effects of drugs on cellular and bodily function, the proteomes of blood (Hu, Loo et al. 2006; Pasini, E.M., et al. 2006; Yang and Huang 2007, Anderson, Polanski et al. 2004; Liu and Clemmer, et al. 2007), saliva (Hu, Xie et al. 2005; Xie, Rhodus et al. 2005; Guo, Rudnick et al. 2006), semen (Pilch and Mann 2006) and vaginal fluid (Venkataraman N, AL. Cole, et al. 2005; Dasari, Pereira et al. 2007; Pereira, L., A.P. Reddy, et al. 2007; Quinzio, D. MK., K.Oliva, et al 2007; Shaw, Smith et al. 2007; Tang, De Seta et al. 2007; Klein, LL, KR, Jonscher, et al. 2008; Zegels and Van Ostade, et al. 2009) have been extensively studied, and the unique proteins that constitute each body fluid are known.

Mass spectrometry is particularly well suited for body fluid identification for forensic purposes for several reasons: 1) It does not rely on biological activity of selected markers, but on protein sequence, making it confirmatory and raising body fluid testing to the level of DNA analysis. 2) Proteins are less subject to degradation than RNA, and partial protein fragments can often be used for identification. 3) Our results showed that MS assay sensitivity for blood, saliva and semen is currently better than mRNA methods (see Figure 4). 4) All samples are prepared in the same way and subjected to one MS analysis. Since no other tests are necessary there is a large saving in sample, time and costs. 5) By choosing a *group* of unique or enriched protein markers for each body fluid, a body fluid specific “protein fingerprint” is established giving greater confidence in results. 6) Body fluid specific protein fingerprints make detecting mixtures easier, as confounding results, such as finding PSA in a blood sample, are eliminated since all fingerprint proteins would need to present to make that determination. 7) Species information is inherently coded in proteins and can be determined simultaneously with body fluid identification. 8) Recent data suggest that Y-chromosome specific proteins are

present in blood (Ofran et al. 2010) suggesting that gender identification can also be determined.

C. STATEMENT OF HYPOTHESIS / RATIONALE FOR THE RESEARCH – Current methodologies used for body fluid identification are diverse in terms of technology, and vary greatly in terms of the quality of their results. Many tests in common use are presumptive and common forensic standards, similar to those used for DNA STR typing, are lacking. The establishment of a single, confirmatory method capable of testing all body fluids at high levels of sensitivity would simplify the current medley of techniques, replace the uncertainty of presumptive tests, and facilitate standardization. The advent of mass spectrometry for biomolecule identification has opened the field of body fluid identification to a new powerful tool capable of high accuracy, sensitivity and reproducibility. We hypothesize that a body fluid assay by mass spectrometry could be developed to meet the criteria of a common, single confirmatory test for nearly all body fluids. To test this hypothesis, we evaluated blood, saliva, semen, menstrual blood and vaginal fluid to identify body fluid specific markers and evaluate them by MALDI MS.

II. METHODS

Experimental Design, Methods and Materials

MATERIALS: Five body fluids, (blood, saliva, semen, menstrual blood and vaginal fluid) were collected from volunteers and stored at -80°C (see Experimental Design below for collection details) by procedures approved by the New York City Department of Health and Mental Hygiene IRB. Isoelectric focusing strips (pH 3-10, 11 cm) and precast 15% SDS PAGE gels were purchased from Bio-Rad (Hercules, CA). SDS, Tris, PBS, ammonium bicarbonate, urea, thiourea, sodium deoxycholate, RIPA buffer, HPLC grade organics (e.g. acetonitrile, methanol, chloroform, trifluoroacetic acid etc.), HPLC grade water and Bradford protein detection reagent were purchased form Fisher Scientific Inc. (Pittsburgh, PA). Trypsin (mass spectrometry grade) was purchased from Promega (Madison, WI). Dithiothreitol, iodoacetamide, CHAPS, NP-40, α-cyano-4-hydroxycinnamic acid (CHCA) and Enzymatic Protein Deglycosylation Kits were purchased from Sigma-Aldrich (St. Louis, MO). C-18 nano-columns (75 µm x 150 mm, 3 µm, 100 Å) were purchased from Dionex (Sunnyvale, CA). Lysing Matrix D (mechanical extraction media) was obtained from MP Biomedicals (Solon, OH). C18 ZipTips were purchased from Millipore (Billerica, MA). Isotopically labeled, HPLC purified peptides were purchased from the Rockefeller University Proteomic Research Center, New York, NY.

INSTRUMENTATION: An AB SCIEX 4800 MALDI TOF/TOF Analyzer was used for mass spectrometry. High performance nano-column liquid chromatography was performed on a Dionex UltiMate 3000 LC system; purified samples were spotted on MALDI plates using a Dionex Probot. Isoelectric focusing was done with a BioRad (Hercules, CA) Protean IEF Cell; SDS PAGE was performed with a BioRad Mini-Protean Tetra System. Samples were extracted with a MP Biomedicals FastPrep-24 mechanical cell disruptor (Solon, OH).

EXPERIMENTAL DESIGN & METHODS – The goal of this work was to use mass spectrometry to establish a rapid and sensitive method to identify five different body fluids in a single confirmatory test. The test was established by selecting appropriate peptide markers. to allow for quantitative analysis it was also an aim to develop and test a panel of isotopic peptide standards. Subsequently, experiments were designed to determine i) the best method for sample extraction and solubilization, ii) the limits of detection of pure samples, iii) the limits of detection of each body fluid in mixed samples, and iv) the assay's effectiveness with mock forensic samples. Additional work was done on menstrual blood and vaginal fluid. As requested, descriptions of the rationales and methods used for each of these experiments are detailed below.

1. Body Fluid Collection & Mock Forensic Sample Preparation –

Rationale: Sample collection from volunteers requires IRB approval; once obtained, samples collection and storage were done as described below.

Methods: An IRB application was approved by the New York City Department of Health and Mental Hygiene's Internal Review Board for the collection of five body fluids. Saliva and semen were collected at home by volunteers in 50 ml polypropylene tube, stored overnight at 4°C and returned to our laboratory where they were stored at -80°C. Blood was collected in our laboratory by finger-stick with 0.2 – 0.5 ml collected in sterile Microtainers containing EDTA and stored at -80°C. Menstrual blood and vaginal fluid were collected by volunteers at home and stored overnight at 4°C and returned the next day to our laboratory and stored at -80°C. Two collection methods were used. The first used tampons that were kept in place for a minimum of 4 hours, the second used a menstrual blood collection cup. A total of 12 volunteers provided 45 samples. Volunteer demographics are presented in Table 1.

Table 1. Demographics of Collected Body Fluid Samples

Blood	Sample No	Age	Gender	Race
1	#215	42	Female	Asian
2	#271	25	Female	White
3	#947	38	Female	White
4	#182	46	Female	Black
5	#454	50	Female	White
6	#256	56	Male	White
7	#560	37	Male	Asian
8	#885	55.5	Female	White
9	#982	48	Female	Black
10	#300	30	Female	White
11	#640	43	Female	white
12	#750	29	Female	White

Saliva	Sample No	Age	Gender	Race
1	#215	42	Female	Asian
2	#271	25	Female	White
3	#947	38	Female	White
4	#182	46	Female	Black
5	#454	50	Female	White
6	#256	56	Male	White
7	#560	37	Male	Asian
8	#885	55.5	Female	White
9	#982	48	Female	Black
10	#300	30	Female	White
11	#640	43	Female	white
12	#750	29	Female	White

Semen	Sample No	Age	Gender	Race
1	#256	56	Male	White
2	#560	37	Male	Asian
3	#T2388	35	Male	Asian

Menstrual Blood	Sample No	Age	Gender	Race	Time in Period	Contraception
1	#215	42	Female	Asian	Middle	
2	#271	25	Female	White	End	
3	#947	38	Female	White	Middle	
4	#454	50	Female	White	End	
5	#454	50	Female	White	Middle	
7	#982	48	Female	Black	Beginning	
8	#300	30	Female	White	Middle	Oral
9	#640	43	Female	White	End	
10	#750	29	Female	White	End	Oral

Vaginal Fluid	Sample No	Age	Gender	Race	Days Since Last Period Ended	Contraception
1	#1	NA	Female	Asian	NA	
2	#947	38	Female	White	14 days	
4	#885	55.5	Female	White	PM	
5	#293	43	Female	Asian	15 days	
6	#982	48	Female	Black	NA	

7	#750	29	Female	White	14 days	Oral
8	#215	42	Female	Asian	10 days	

Mock forensic samples were prepared by spotting known volumes of each body fluid on five different substrates: cotton, polyester, wood (tongue depressor), common plastic beverage bottles (polyethylene terephthalate, PETE), and condoms. Two methods for spotting were used. The first spotted samples directly on substrates; the second mixed the samples with their specific associated isotopically labeled peptides (5 nmoles isotope / 5 μl blood, 0.5 nmoles isotope /10 μl semen) which were then spotted onto substrates. The purpose of mixing a body fluid specific labeled peptide with its associated body fluid was to help determine extraction efficiency by later measuring the amount of labeled peptide recovered. Mock forensic samples were dried and stored at room temperature away from direct sunlight for periods up to 20 months.

2. Preparation and Testing of Isotopically Labeled Peptide Standards –

Rationale: The purpose of synthesizing isotopically labeled peptide standards was to be able to quantify the amounts of body fluid marker proteins in a sample. The criteria for selecting which peptide markers to label were: i) clear separation by nano-LC, ii) clear separation and strong signal by MALDI mass spectrometry, iii) masses between 1,000 and 2,500 Da, iv) avoidance of peptides with modified amino acids, v) avoidance of peptides with cryptic trypsin digestion sites and vi) avoidance of peptides with amino acids sequences difficult to synthesize. Most of these criteria must be determined empirically; consequently selections were made following the experiments described above. In total, 29 isotopically labeled peptides were selected and labeling performed by our collaborator Dr. Haiteng Deng at Rockefeller University. Methods for testing the quality and levels of detection of synthesized peptides follow.

Methods:

- **Quality:** Isotopically labeled peptides were resuspended in 50% acetonitrile /0.1% TFA, mixed with 5 mg/ml CHCA. The mixtures were spotted on a MALDI plate and subjected to MALDI TOF-TOF. Quality was checked by MALDI TOF/TOF by evaluating the mass, peptide sequence and position of labeled amino acids.
- **Levels of Detection:** To determine the linear range of detection for each labeled peptide standard, serial dilutions were made with the following amounts of peptide spotted on a MALDI plate in a 1 μl volume: 1 pmol, 0.5 pmol, 0.1 pmol, 50 fmol, 10 fmol, 5 fmol and 2 fmol. In addition to testing individual peptides, similar dilutions were made with mixtures of all peptides

for each body fluid. This was done to determine if ionization efficiencies of individual peptides were affected when present in a mixture (i.e. do levels of detection decline).

3. Extraction and Solubilization –

Rationale: Perhaps the most important aspect of sample detection is the initial extraction, as what is not extracted cannot be measured. Consequently, determining ideal extraction conditions is the first step toward establishing a sensitive assay. The two variables that most affect extraction efficiency are the substrate material on which a body fluid is present (as different substrates can bind samples with different affinities), and the components of the solution used for extraction. The substrates we examined were: cotton, polyester, wood, common beverage bottle (polyethylene terephthalate, PETE) and a condom. Solubilizing solutions were evaluated with respect to buffers, detergents and chaotropic salts.

Methods: Two methods were used to evaluate sample extraction efficiency. The first, SDS PAGE, was used to detect overall protein extraction from samples. The second involved spiking samples with isotopically labeled peptide markers prior to spotting them on different substrates and then comparing the amounts of recovered markers to the amounts originally spotted. Both freshly prepared and aged mock forensic samples were used to determine if extraction efficiency and protein stability were affected by aging.

- Sample Extraction Method: Extraction solutions included ionic and non-ionic detergents, as well as chaotropic salts. Five extraction solutions were tried (see below), extraction of fresh samples were used as control. Approximately 0.5 cm^2 of mock forensic samples were placed in a 2.0 ml screw cap lysing tubes containing 1.4 mm ceramic spheres (Matrix D, MP Biomedicals) with 300 µl of buffer. Sample extraction was performed on a FastPrep-24 (MP Biomedicals) reciprocating homogenizer set to vibrate at 6.0 m/s for 30 s at room temperature. After extraction, samples were spun at 20,000 g for 10 min and supernatants carefully removed. Relative extraction of individual bands and degradation of aged mock samples were evaluated by loading equal volumes of extracted samples on 15% PAGE gels (along with freshly prepared controls), staining with Coomassie blue, scanning, and quantifying bands by NIH ImageJ software. To evaluate extraction efficiency and levels of degradation of *target* proteins, body fluids were spiked with body fluid specific, isotopically labeled, markers peptides. Detection of labeled and endogenous peptides was performed by LC/MALDI (see below).

The five extraction solutions were:

1. Deionized water
2. PBS (10 mM Na2HPO4 - 2 mMKH2PO4 pH 7.4, 2.7 mM KCl, 137 mM NaCl) pH 7.4

3. 2% SDS in Tris buffer (25 mM Tris-HCl pH7.4, 2.7 mM KCl, 137 mM NaCl)

4. RIPA buffer (50 mM Tris-HCl pH7.4, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% Sodium deoxycholate)

5. 4% CHAPS , 7M urea and 2M thiourea

4. Detection of Body Fluid Samples, after Purification –

Rationale: In order to determine the lowest level of detection of each body fluid, as well as identify those specific peptide markers that give the most reproducible and strongest signals, known quantities of body fluids were serially diluted, and subjected to multiple methods of protein purification prior to detection by MALDI TOF-TOF. Protein purification techniques included: polyacrylamide gel electrophoresis (SDS PAGE), isoelectric focusing (IEF) and high performance nano-liquid chromatography (LC).

Methods: For SDS PAGE, samples were homogenized using a FastPrep24 instrument in RIPA buffer, debris removed by centrifugation at 20,000 g X 10 min. Supernatants were carefully removed, protein content measured by the Bradford method (BSA as standard) and 10 ug loaded on 15% PAGE gels run at 130 V constant voltage. Gels were stained with Coomassie brilliant blue. Lanes were cut into equal lengths (0.5 cm), destained, reduced, alkylated and subjected to in-gel trypsin digestion at 37°C overnight. Samples were dried in a SpeedVac, resuspended in 5% acetonitrile/0.1% TFA and separated by reverse phase high performance liquid nano-chromatography using a continuous gradient of 5% - 45% acetonitrile. This method is referred to as PAGE-LC-MALDI.

For isoelectric focusing experiments, samples were homogenized using a FastPrep24 instrument in rehydration buffer containing 7 M urea, 2 M thiourea and 2% CHAPS. Debris was removed by centrifugation at 20,000 g X 10 min. Supernatants were carefully removed, protein content measured by the Bradford method (BSA as standard) and 10 µg run on a 11 cm IEF strips with pH ranging from 3 - 10. The IEF run was started at 250V for 15 min followed by rapid voltage ramping for 3 hrs without exceeding 50 µA/strip current. The run was finished when the voltage reached 35,000 Vhr at 8,000 V. The voltage was held at 500 V to prevent diffusion. The IEF run was performed at 20°C. Samples typically ran 12-14 hrs. Following focusing, strips were cut into equal lengths (0.5 cm), washed, reduced, alkylated and subjected to in-gel trypsin digestion overnight at 37°C. Samples were dried in a SpeedVac, resuspended in 5% acetonitrile/0.1% TFA and separated by reverse phase high performance nano-chromatography using a continuous gradient of 5% - 20% acetonitrile. Detection after this sequence of steps will be referred to as IEF-LC-MALDI.

For direct high performance nano-liquid chromatography experiments, samples were homogenized using a FastPrep24 instrument in RIPA buffer. Debris was removed by centrifugation at 20,000 g X 10 min. Supernatants were carefully removed, protein content measured by the Bradford method. Samples were, reduced, alkylated and digested overnight at 37°C with trypsin. Samples were dried in a SpeedVac, resuspended in 5% acetonitrile/0.1% TFA and separated by reverse phase high performance liquid nano-chromatography using a continuous gradient of 5% - 45% acetonitrile. Detection after this sequence of steps will be referred to as LC-MALDI.

5. Direct Detection of Body Fluid Samples Without Protein Purification –

Rationale: One of the goals of this work was to establish a body fluid assay amenable to high throughput testing, e.g. <500 samples/wk. While we demonstrated that the use of 2D protein/peptide separation techniques can identify multiple peptide markers from multiple specific proteins in each body fluid and at high sensitivity, these methods require a significant amount of time. Consequently, we decided to evaluate limits of detection for body fluid identification without prior protein purification. Because only body fluids with abundant markers would be detectable by this method, only blood (hemoglobin), saliva (amylase) and semen (semenogelin) were tested.

Methods: Ten microliters samples were homogenized in 100 μl RIPA buffer using a FastPrep24 reciprocating homogenizer and debris removed by centrifugation. Protein concentration was determined by the Bradford method and 20 μg of protein were reduced, alkylated and digested with trypsin as described above. Two micrograms of digest were loaded onto C18 ZipTips, washed and eluted with 10 μl of 50% acetonitrile/0.1% TFA. Samples were dried in a SpeedVac, resuspended in 10 μl of 50% acetonitrile/0.1% TFA. One microliter of sample was mixed with 1 μl of 7.5 mg/ml CHCA in 50% acetonitrile/0.1% TFA, spotted onto MALDI plates and subjected to MALDI TOF/TOF detection and analysis. Detection after this sequence of steps will be referred to as direct spotting MALDI or MALDI.

6. Limits of Detection of Body Fluids from Mixed Samples –

Rationale: The ability to distinguish two or more body fluids in a single sample depends on three factors: i) the abundance of each marker within its body fluid, ii) the ratio of body fluids in a mixture - based on total protein not volume, and iii) the ionization efficiency of peptide markers. Thus, a body fluid with a high protein concentration in which one or several protein markers

predominate will be more easily detected in a mixture than a body fluid with a low protein concentration and in which protein markers make up a smaller percentage of total protein.

Methods: Three body fluids were tested: blood, saliva and semen. (vaginal fluid and menstrual blood were not used at this time because of the difficulty of identifying reliably consistent markers. Each body fluid was mixed with a second body fluid at varying ratios. Mixtures were solubilized with extraction buffer (7 M urea, 2 M thiourea, 4% CHAPS, 60 mM DTT) using a FastPrep reciprocating homogenizer. Proteins were then separated by IEF, and those IEF fractions containing known body fluid marker proteins were digested with trypsin, separated by nano liquid chromatography and analyzed by MALDI-TOF/TOF. Mixtures were also solubilized with RIPA buffer using a FastPrep reciprocating homogenizer. Proteins were then digested with trypsin, separated by nano liquid chromatography and analyzed by MALDI-TOF/TOF.

7. Freshly Prepared and Aged Mock Forensic Samples –

Rationale: To determine the effects of aging on protein degradation and our ability to extract and solubilize them, mock forensic samples were prepared on five materials (cotton, polyester, wood, condom and plastic beverage bottles (PETE)) as described above. The five materials were cut into small pieces and 2 µl of each body fluid were applied on each. Ten stained pieces of each material were prepared for different aging periods. After drying, samples were placed in clear plastic sleeves in a biosafety hood with the glass door down, out of direct sunlight, but exposed to indirect daylight and artificial light on a 10 hr cycle and at room temperature. Overall samples were stored at room temperature for varying lengths of times up to 20 months. To determine the effect of aging, two methods were used. In the first, samples were extracted and run on SDS PAGE gels alongside equal amounts of freshly prepared samples in order to estimate recovery and levels of degradation. In the second, specific marker proteins were evaluated by sample extraction, digestion, LC and MALDI TOF/TOF. Quantitation of the level of degradation was also be evaluated from aged samples that have been spiked with isotopically labeled peptides.

Methods: A total of 5 µl of semen (about 100 µg of protein) were spotted separately on all five substrates. Similar amounts of protein from blood (1 µl) and saliva (10 µl) were also spotted. Samples were allowed to dry for 3 hrs and then extracted with 300 µl of RIPA buffer (see recipe above) using FastPrep at a setting of 6m/sec for 30 seconds. Samples were spun at 20,000 g for 10 minutes at room temperature and supernatants carefully removed. Total protein was measured by the Bradford methods with BSA as standard and run on 15% SDS PAGE gels

alongside equal amounts of freshly prepared samples. Gels were stained with Coomassie Blue and destained. Gels were scanned and images quantified by NIH ImageJ software.

For quantitation of isotopically labeled spiked body fluids, samples were prepared as follows: to 5 µl of blood 5 nmoles of corresponding isotopically labeled body fluid markers were added, and to 10 µl of semen 0.5 nmoles of corresponding isotopically labeled body fluid markers were added. Samples (2 µl) were then applied to cotton and allowed to age for 3 months; evaluation is currently underway. Here, samples will be extracted as described above, but analysis will not be by SDS PAGE, but MALDI TOF/TOF. Following solubilization, extraction, and centrifugation, supernatants will be reduced, alkylated, digested with trypsin and directly spotted on MALDI plates for TOF/TOF. Freshly made samples prepared at the same ratios of protein to isotopically labeled peptides will be used as controls.

8. Deglycosylation of Menstrual Blood Proteins –

Rationale**:** The original markers we proposed to use for menstrual blood identification, based on the presence of their mRNAs in menstrual blood, were matrix metalloproteinases (**MMPs**) 7, 10 and 11. However, even after employing two different 2D protein/peptide purification methods (SDS PAGE/LC and IEF/LC) followed by MALDI TOF/TOF, none of these proteins could be identified. Because the MMPs can be glycosylated to varying extents - thereby changing both their mass and isoelectric points, which would effectively dilute them on PAGE gels and IEF strips - deglycosylation experiments were carried out to try to detect them.

Methods: Menstrual blood samples were solubilized and extracted in RIPA buffer using FastPrep as describe above. After centrifugation supernatants were collected. Hb, Alb and IgGs were depleted form the supernatant using ProteoPrep Immunoaffinity Albumin and IgG Depletion Kits (Sigma-Aldrich, St. Louis, MO) and Ni-NTA agarose beads (see detail procedures in section 10-2). Protein content of depleted menstrual blood was measured by the Bradford method with BSA as standard. For deglycosylation, 30 µl of sample (300 µg protein) were reacted with 1 µl of a mixture of glycosidases (200 ng of each enzyme; Enzymatic Protein Deglycosylation Kit, Sigma-Aldrich, St. Louis, MO) for 3 hours at 37°C. Samples were then separated on IEF strip (pH 3-10, 11 cm). The strip was cut into 0.5 cm, proteins digested with trypsin, separated by nano liquid chromatography and analyzed by MALDI-TOF/TOF. While several promising marker proteins were identified, as well as several MMPs, MMPs 7, 10, and 11 were not identified.

9. Establishing a Menstrual Blood Proteome –

Rationale: As described above, we were unable to detect matrix metalloproteinases 7, 10 and 11 in menstrual blood. Consequently, we undertook a search for new menstrual blood protein markers. However, as no menstrual blood proteome has been published, it was necessary to perform a complete proteome analysis of menstrual blood. For thoroughness, several different starting materials and protein purification methods were used. These are discussed in methods below. It should be noted that the most characteristic proteins of blood (hemoglobin (**Hb**), albumin (**Alb**) and immunoglobulins (**Igs**)) cannot be used as specific markers for menstrual blood. In addition, as these three proteins are the most abundant proteins in blood, they have the potential to mask unique proteins of menstrual blood. Consequently, in addition to whole menstrual blood, Hb, Alb and IgG depleted menstrual was also subjected to a full proteomic analysis.

Methods:

Types of samples used:

i) *WHOLE MENSTRUAL BLOOD:* Whole menstrual blood was solubilized by FastPrep using RIPA buffer and spun at 2,000 x g for 10 minutes. Both the supernatant and pellets were used. The supernatant was digested with trypsin as described above. The digested supernatant was dried in a SpeedVac, resuspended in 5% acetonitrile/0.1%TFA for LC-MALDI analysis. The pellet was solubilized with SDS sample buffer which consists of 2% SDS, and separated on a 15% SDS PAGE gel. The gel lane was cut into small pieces (0.5cm) and the proteins from each gel piece were in-gel digested with trypsin, separated by nano liquid chromatography and analyzed by MALDI-TOF/TOF.

II) HB, ALB AND IGG DEPLETED MENSTRUAL BLOOD: Depleted menstrual blood was extracted using the same method described above. The supernatant was mixed with Ni-NTA agarose beads to deplete hemoglobins (Ringrose JH, Solinge WW, et al. 2008). After mixing for 1 hr at room temperature with the beads, the supernatant was removed. The beads were washed twice using same lysis buffer. The supernatant and washing solutions were combined and submitted to Alb-IgG depletion using ProteoPrep Immunoaffinity Albumin and IgG Depletion Kit (Sigma-Aldrich, St. Louis, MO) following the manufacturer's instructions.

Types of Analysis: Depleted menstrual blood samples were analyzed by three different methods.

i) Total proteins were digested with trypsin, separated by nano liquid chromatography and analyzed by MALDI-TOF/TOF

ii) Proteins were separated on a 15% SDS PAGE gel, lanes were cut into 0.5 cm pieces, followed by in-gel trypsin digestion, separation by nano liquid chromatography and analysis by MALDI-TOF/TOF

iii) Proteins were separated on IEF strips (pH 3-10, 11 cm), strips were cut into 0.5 cm pieces, followed by in-gel trypsin digestion, separation by nano liquid chromatography and analysis by MALDI-TOF/TOF

10. MALDI-TOF/TOF Analysis and Database Search

A 4800 MALDI TOF/TOF Analyzer (AB SCIEX, Foster City, CA) was used to collect MS and MS/MS data. MS data were acquired over a mass range of 800-4000 Da in positive-ion reflector mode with 4000 Series Explorer software at a laser repetition rate of 200 Hz with 1500 laser shots/spectrum (50 laser shots/sub-spectrum). For MS/MS spectra, the 30 most intense precursors per spot with a minimum S/N of 20 were selected for subsequent fragmentation by collision-induced dissociation at 200 Hz in 1 kV MS/MS mode with 4200 laser shots/spectrum (25 laser shots/sub-spectrum) with the following TOF/TOF Series Explorer software Stop Conditions: Maximum shots per spectrum 4200; Minimum shots per spectrum 1000; number of MS/MS fragments 8; S/N of each fragment 75.

The data were processed by ProteinPilot 3.0 software using the Paragon search algorithm. MS/MS data were searched against the IPI (Human) database. Trypsin was selected as the digestion enzyme with a fixed carbamidomethylation modification on cysteine residues, and one missed cleavage site was allowed. All peptide mass values were considered monoisotopic and mass tolerance was set at 50 ppm for MS analysis, and 0.35 Da for MS/MS analysis. Peptide score 1.0 (>90%) was selected as detected protein threshold.

- *CONFIDENCE OF IDENTIFYING PROTEINS FROM MS DATA* – Peptide identification by mass spectrometry is based on comparing masses of unknown sample peptides against known protein databases. One of the most accurate methods for identifying peptides uses tandem mass spectrometry in which a precursor peptide identified by the first mass spectrometer is then directed into a gas collision chamber where it is fragmented. The masses of the generated fragments are then measured in the second mass spectrometer and compared against an *in silico*-generated database of peptide spectra using search algorithms in an uninterpreted manner to determine the mass of the precursor peptide with a high level of confidence.

In our experiments we use MALDI TOF/TOF and ProteinPilot software for protein identification. ProteinPilot combines the Mascot search engine from Matrix Science with the Paragon search algorithm from Applied Biosystems to identify peptides. Consequently, protein

identification is flexible with an option to choose the Paragon or Mascot search engine for protein ID. We have compared the confidence calling of these two programs by analyzing the same set of MALDI data with each; both algorithms give similar confidence scores and identify the same proteins.

We routinely use the Paragon search engine. A peptide score of 1.0 using the Paragon algorithm represents a 90% confidence of identification. We use the score of 1.0 merely as a threshold for research data analysis - basically, we want to see which proteins are identified at the 90% level. In terms of protein identification calling, we require a score ≥ 1.3 (confidence ≥ 95%). For routine forensic testing, we believe that sample spectra should be compared against synthetic peptides of known sequence; in this case confidence approaches 100%.

A review of the criteria employed by different algorithms to determine calling confidence is beyond the scope of this report, however we refer the reader to Shilov et al. 2007 and Kapp et al. 2005 for good published reviews.

III. RESULTS

This section is divided into nine subsections representing major findings.

1. Evaluation of Proposed Body Fluid Markers: The main criteria for selecting body fluid marker proteins were: i) specificity (i.e. proteins that are unique or highly enriched in a particular body fluid), and ii) abundance. An ideal marker would be both specific and abundant, but specificity is required. Initially we had proposed protein markers identified in the literature, some of which were based on the presence of their mRNAs (Table 2). However, experiments using the Nano-LC–MS could not detect all these markers (e.g. PBGD, PRM1, STATH, HTN3, Muc4, HBD1, MMP-7, MMP-10, MMP-11). Consequently, several new marker proteins were identified for semen, blood and saliva that better fit our criteria of specificity and abundance than those originally described. These markers were easily identified in 2 ng of blood and semen, and 10 ng of saliva, which corresponds to 0.006 nl blood, 0.1 nl of semen and 4 nl of saliva, demonstrating their sensitivity. Consequently we have modified our original marker list for these three body fluids (Table 2).

Table 2. Body Fluid Marker Proteins

Body Fluid	Proposed Biomarker Protein	Identified Biomarker Protein
Blood	Beta-spectrin (SPTB) Hemoglobin alpha (HBA) Porphobilinogen deaminase (PBGB)	Hemoglobin alpha (HBA) Hemoglobin beta (HBB) Beta-spectrin (SPTB) Alpha-spectrin (SPTA1)

		Band 3 anion transport protein (SLC4A1)
Saliva	Statherin (STATH) Histatin 3 (HTN3) Mucin-5B (MUC 5B)	Alpha-amylase 1 (AMY1) Cystatin SA (CST2) Uncharacterized protein C6orf58 (C6oor58) Histatin-1 (HTN1)
Semen	Prostate specific antigen (PSA) Kallikrein 4 (KLK 4) Protamine 1 (PRM1)	Semenogelin-2 (SEMG2) Semenogelin-1 (SEMG1) Mucin 6 (MUC6) Prostate specific antigen (PSA) Progestagen-associated endometrial protein (PAEP) Cysteine-rich secretory protein 1 (CRISP1)
Vaginal Fluid	Mucin 4 isoform a (MUC 4) Human beta-defensins 1 (HBD 1)	(See Candidates in Table 3)
Menstrual Blood	Matrix metalloprotease 7 (MMP-7) Matrix metalloprotease 10 (MMP-10) Matrix metalloprotease 11 (MMP-11)	(See Candidates in Table 3)

To identify menstrual blood and vaginal fluid specific protein markers, we have performed complete proteome analyses of the two body fluids, and compared results with the proteomes of blood, saliva and semen. Promising candidate markers based on their uniqueness and abundance are listed in Table 3.

Table 3. Candidate Markers for Menstrual Blood and Vaginal Fluid

Body Fluid	Protein Name	General Annotation
Menstrual Blood	Hemoglobin subunit zeta (HBZ)	HBZ is expressed in human embryo and human endometrium (Dassen et al. 2008). HBZ was identified in our Menstrual blood sample.
	15-hydroxyprostaglandin dehydrogenase [NAD+] (HPGD)	HPGD was found in our menstrual blood sample, but wasn't found in our other body fluids and is not present in databases and literatures of other body fluids.
	Progestagen-associated endometrial protein (PAEP)	PAEP is synthesized and secreted in the endometrium from mid-luteal phase of the menstrual cycle and during the first semester of pregnancy. At least two differentially glycosylated forms of this protein are found. Glycodelin-A (GdA) form is in amniotic fluid) and glycodelin-S (Gds) form is in seminal plasma.
	Mucin-5B (MUC5B)	MUC5B is highly expressed in saliva and VF, but not in venous blood. This protein is highly abundant in our Menstrual blood sample, and can serve as a marker to distinguish menstrual blood and venous blood.

	Proteasome subunit beta type-5 (PSMB5)	PSMB 5 is a component of proteasome which is a multicatalytic proteinase complex and is characterized by its ability to cleave peptides with Arg, Phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or slightly basic pH. This protein was found in our menstrual blood sample, but not in other body fluids.
Vaginal Fluid	small proline-rich protein 3 (SPRR3)	SPRR3 is a marker of human esophageal neoplastic progression. this protein may play a role in the maintenance of normal esophageal epithelial homeostasis, and aberrant expression of SPRR3 may contribute to the tumorigenesis of esophageal squamous cell carcinoma. SPPR3 is highly abundant in vaginal fluid.
	Dimethylaniline monooxygenase 3 (FMO3)	A major function of FMO3 is oxidative metabolism of a variety of xenobiotics such as drugs and pesticides. This protein was found in our vaginal fluid sample, but not in other body fluids.
	Transmembrane protease, serine 11B (TMPRSS1B)	TMPRSS11B is a single pass type II membrane protein. Its cDNA is expressed in cervical (Bechtel et al. 2007; Kneller et al. 2007). This protein was identified in vaginal fluid (Shaw, Smith et al. 2007).
	Cartilage intermediate layer protein 1 (CILP1)	CILP1 and CILP2 are isoforms of cartilage intermediate layer protein, which play a role in cartilage scaffolding. These two proteins were found in our vaginal fluid sample, but not in other body fluids.
	Cartilage intermediate layer protein 2 (CILP2)	
	Beta-crystallin S (CRYGS)	CRYGS is a major protein of vertebrate eye lens and maintains the transparency and refractive index of the lens. This protein was identified in our vaginal fluid, but not in other body fluids.
	Isoform 1 of A-kinase anchor protein 13 (AKAP13)	AKAP13 is a member of the family of the A-kinase anchor proteins (AKAPs), which have the common function of binding to protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell. This protein is expressed in endometrium (Hearns-Stokes et al. 2006), and was found in vaginal fluid. (Shaw, Smith et al. 2007).

2. Isotopically Labeled Marker Peptides: Selection, Synthesis, and Evaluation - An important aspect of body fluid identification is the ability to quantify the amounts of endogenous marker peptides present in a sample; this is particularly true when trying to identify mixed samples that may be present in extreme ratios. To quantify peptides by mass spectrometry samples are spiked with a known amount of isotopically labeled (heavy) peptides and their signal intensities compared with those of endogenous peptides.

2.1 SELECTION OF PEPTIDE MARKERS: The criteria for selecting which peptides to label are: i) clear separation and strong signal by MALDI mass spectrometry; ii) masses between 1,000 and 2,500 Da; iii) avoidance of peptides with cryptic trypsin digestion sites; iv) avoidance of peptides

with modified amino acids; v) avoidance of peptides with amino acids sequences difficult to synthesize; and vi) clear separation by nano-LC. Peptides meeting these criteria were synthesized and isotopically labeling by our collaborator Dr. Haiteng Deng, Director of the Rockefeller University Proteomics Resource Center. Peptides were synthesized in two groups as additional information became available – e.g. candidate markers for menstrual blood and vaginal fluid. They are listed in Tables 4 and 5.

Table 4. Isotopically labeled marker peptides (synthesized in July 2009)

Body Fluid	Peptide Name	Peptide Sequence	Detection Level Peptide Alone	Detection Level in peptide Mix	Detection Level in 50 ng Body Fluid
semen	SEMG1-1	LPSEFSQFPHGQK	2 fmol	2 fmol	2 fmol
	SEMG1-2	DIFSTQDELLVYNK	5 fmol	5 fmol	5 fmol
	SEMG1-3	QEPWHGENAK	10 fmol	10 fmol	10 fmol
	SEMG2-1	DVSQSSISFQIEK	50 fmol	ND	ND
	SEMG2-2	GQLPSGSSQFPHGQK	5 fmol	5 fmol	5 fmol
	SEMG2-3	STQKDVSQSSISFQIEK	2 fmol	2 fmol	2 fmol
	KLK3-1	HSQPWQVLVASR	2 fmol	2 fmol	2 fmol
	KLK3-2	LSEPAELTDAVK	10 fmol	10 fmol	10 fmol
saliva	AMY1-1	NVVDGQPFTNWYDNGS NQVAFGR	2 fmol	2 fmol	
	AMY1-2	TSIVHLFEWR	2 fmol	2 fmol	
	AMY1-3	ALVFVDNHDNQR	2 fmol	2 fmol	
	CST2-1	ALHFVISEYNK	2 fmol	2 fmol	
blood	HBB-1	VLGAFSDGLAHLDNLK	2 fmol	2 fmol	2 fmol
	HBB-2	LLVVYPWTQR	2 fmol	2 fmol	2 fmol
	HBB-3	VNVDEVGGEALGR	2 fmol	2 fmol	2 fmol
	HBA-1	TYFPHFDLSHGSAQVK	2 fmol	2 fmol	2 fmol
	HBA-2	VGAHAGEYGAEALER	2 fmol	2 fmol	2 fmol

Isotopically labeled amino acids are in red fonts in peptide sequences, which are confirmed by MS and MS/MS data. SEMG1, Isoform 1 of Semenogelin-1; SEMG2, Semenogelin-2; KLK3, kallikrein-related peptidase 3; AMY1, Alpha-amylase 1; CST2, Cystatin-SA; HBB, Hemoglobin subunit beta; HBA, Hemoglobin subunit alpha. ND, none detectable.

Table 5 Isotopically labeled marker peptides (synthesized in June 2010)

Body fluid	Protein	Peptide Sequence	Detection Level Peptide Alone	Detection Level in peptide Mix	Detection Level in 50 ng Body Fluid
Semen	SEMG1	DIFSTQDELLVYNK	2 fmol	2 fmol	2 fmol
	SEMG1	QEPWHGENAK	2 fmol	2 fmol	2 fmol

	SEMG2	GQLPSGSSQFPHGQK	2 fmol	2 fmol	2 fmol
	SEMG2	GSISIQTEEQIHGK	2 fmol	2 fmol	2 fmol
Saliva	AMY1	AHFSISNSAEDPFIAIHAESK	10fmol	10fmol	
	AMY1	GHGAGGASILTFWDAR	2 fmol	2 fmol	
	CST2	IIEGGIYDADLNDER	2 fmol	2 fmol	
	C6orf58	ESPGQLSDYR	2 fmol	2 fmol	
Blood	HBB	VNVDEVGGEALGR	2 fmol	2 fmol	2 fmol
	HBB	EFTPPVQAAYQK	2 fmol	2 fmol	2 fmol
	HBA	TYFPHFDLSHGSAQVK	2 fmol	2 fmol	2 fmol
	HBA	VGAHAGEYGAEALER	2 fmol	2 fmol	2 fmol
	SLC4A1	ADFLEQPVLGFVR			
	SLC4A1	LQEAAELEAVELPVPIR			
Vaginal Fluid	FMO3	GVFPPLLEK			
	TMPRSS11B	FKFPPAEGVSMR			
	TMPRSS11B	FKFPPAEGVSMR			
	CRYGS	KPIDWGAASPAVQSFR			
Menstrual Blood	HBZ	TYFPHFDLHPGSAQLR			
	PAEP	HLWYLLDLK			
	PAEP	VHITSLLPTPEDNLEIVLHR			

Isotopically labeled amino acids are in red fonts in peptide sequences, which are confirmed by MS and MS/MS data . SEMG, semenogelin; AMY1, alpha-amylase 1; C6orf58, uncharacterized protein C6orf58; CRYGS, beta-crystallin S; CST2, cystatin-SA; FMO3, dimethylaniline monooxygenase 3; HBB, hemoglobin subunit β; HBA, hemoglobin subunit α; HBZ, hemoglobin subunit zeta; PAEP, progestagen-associated endometrial protein; SLC4A1, band 3 anion transport protein; TMPRSS11B, transmembrane protease, serine 11B.

2.2 QUALITY OF ISOTOPICALLY LABELED PEPTIDES – The quality of synthesized peptides was evaluated by MALDI TOF-TOF. MS and MS/MS data demonstrated that the peptides synthesized in July 2009 had correct sequences with isotopically labeled amino acids present at designed positions (Table 4). Sequences and isotopic labeling of semen peptides, HBA and HBB peptides and saliva peptides synthesized in June 2010 were also confirmed by MALDI (Table 5). Evaluation of the quality of other peptides is underway.

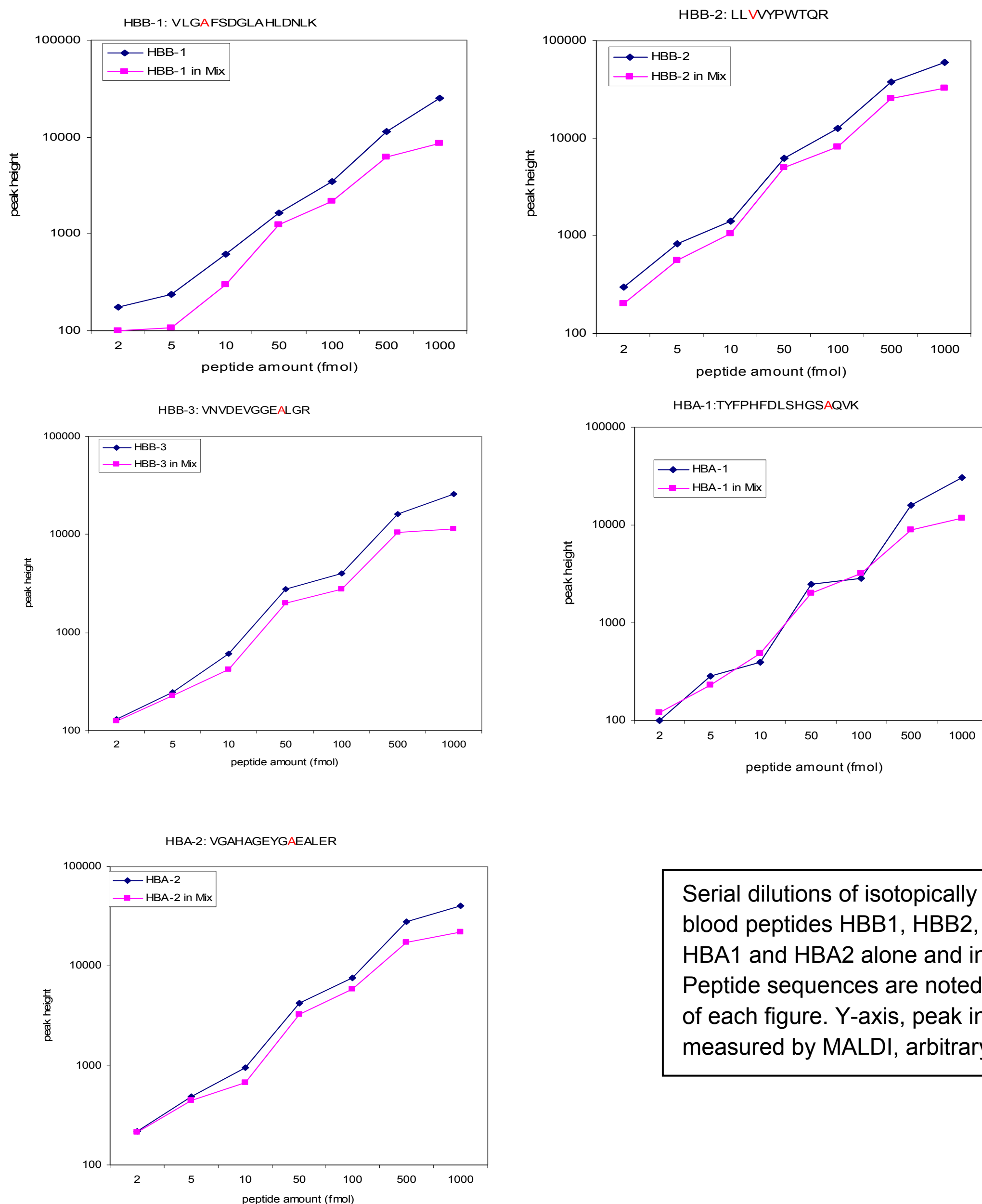
7.3 LEVELS OF DETECTION OF PURE ISOTOPICALLY LABELED PEPTIDES - To determine the linear range of detection for each peptide, serial dilutions were made with the following amounts of peptide spotted on a MALDI plate in a 1 µl volume: 1 pmol, 0.5 pmol, 0.1 pmol, 50 fmol, 10 fmol, 5 fmol and 2 fmol. In addition, to testing individual peptides, similar dilutions were made with mixtures of all peptides for each body fluid (Table 6). This was done to determine if ionization efficiencies of individual peptides were affected when present in a mixture (i.e. do levels of

detection decline, Figure 1). The concentration of each peptide in mixed samples was the same as those stated above for individual peptides, i.e. a mixed 2 fmol sample had 2 fmol of each peptide.

Table 6. Composition of Isotopic Peptide Mixtures

Body Fluid	Peptide in Mixture
Semen	SEMG1-1, SEMG1-2, SEMG1-3, SEMG2-1, SEMG2-2, SEMG2-3, KLK3-1, KLK3-2
Saliva	AMY1-1, AMY1-2, AMY1-3, CST2-1
Blood	HBB-1, HBB-2, HBB-3, HBA-1, HBA-2

As may be seen in Table 4, most of the isotopically labeled peptides were detected by MALDI at 2 fmol level. Some peptides could only be detected at the 5 fmol and 10 fmol level, and SEMG2-1 was only detected at the 50 fmol level when assayed alone and could not be detected when present in semen samples. Those peptides which showed low levels of detection were re-synthesized for higher quality peptides or replaced by newly selected peptides. As shown in Table 5, the peptides which were synthesized in June 2010 and had been tested by MALDI could be detected at 2 fmol level.



Serial dilutions of isotopically labeled blood peptides HBB1, HBB2, HBB3, HBA1 and HBA2 alone and in mixtures. Peptide sequences are noted at the top of each figure. Y-axis, peak intensity measured by MALDI, arbitrary units

Figure 1 Serial Dilution of Blood Isotopic Labeled Peptides. Serial dilutions of isotopically labeled blood peptides HBB1, HBB2, HBB3, HBA1 and HBA2 alone and in mixtures were made with the following amounts of peptide spotted on a MALDI plate in a 1 µl volume: 1 pmol, 0.5 pmol, 0.1 pmol, 50 fmol, 10 fmol, 5 fmol and 2 fmol. MS peak intensities of each peptide at different amounts were measured by MALDI.

7.4 LEVELS OF DETECTION OF ISOTOPICALLY LABELED PEPTIDES IN BODY FLUIDS - Because peptide signal intensity can be affected by the presence of other proteins, we further tested lower limits of detection and effective ranges for isotopically labeled peptides when present in actual body fluid samples. Here we report on minimal levels of detection and effective ranges of marker peptides when mixed with blood and semen. The results in Table 4 and Table 5

demonstrate that all isotopically labeled markers of blood and semen can be detected at the same levels when mixed with their respective body fluids as when assayed alone.

To determine the linear range over which quantitation is effective two experiments were performed. In the first, isotopically labeled peptides concentrations were increased in the presence of a constant amount of body fluid (50 ng protein). In the second, labeled peptide was held constant (100 fmol) while body fluid amounts were increased. Figure 2 show results for blood (using peptide markers HBB-2 & HBA-2), and demonstrates a linear range of quantita-tion of between approximately 2 and 50 ng of endogenous blood proteins (corresponding to 0.006 nl & 0.15 nl of blood). Semen was also quantified in a linear range of 2 and 50 ng of endogenous semen proteins by using the isotopically labeled peptide markers (corresponding to 0.1 nl & 2.5 nl of semen). Similar results were obtained for semen. Quantitation of saliva is underway.

Isotopically labeled peptides are useful for identifying the position of their endogenous counterparts in body fluid samples that are present in vanishingly small amounts. In order to use this technique, it is first necessary to determine the linear range of their detection in the presence of their respective body fluids. The above experiments, identified the amounts of isotopically labeled peptides and the linear range over which they can be used for detecting limiting amounts of body fluids.

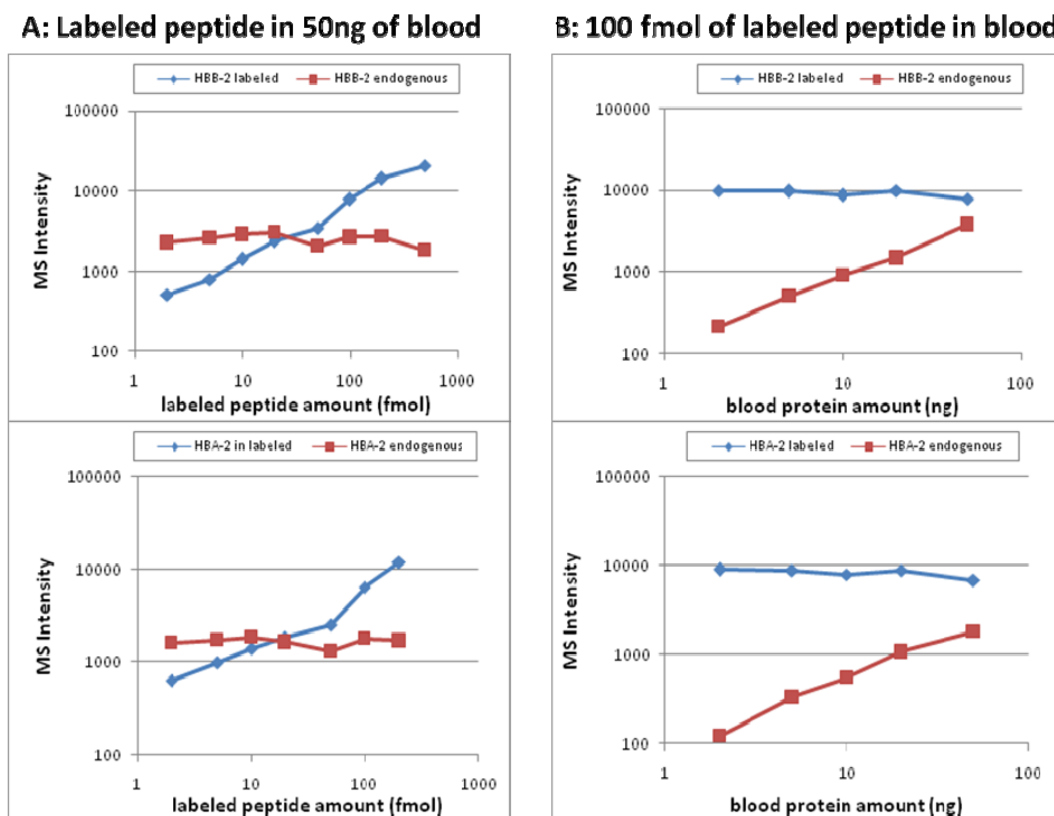


Figure 2 Determining the Linear Range of Quantitation For Blood. A. Isotopically labeled peptides concentrations were increased in the presence of 50 ng of total blood protein. B. Labeled peptide was held constant at 100 fmol while blood amounts were increased from 2 ng to 50 ng. MS peak intensities of each peptide (labeled and endogenous) were measured by MALDI.

3. Body Fluid Extraction and Solubilization - We evaluated extraction conditions for optimum recovery of body fluids from five stained materials: cotton, polyester, wood, polyethylene terephthalate (plastic beverage bottle) and condom. Five different extraction solutions were used as described in section 1 of Methods, above.

Following extraction, samples were centrifuged to remove debris. Equal volumes of extracted proteins were separated on 15% SDS PAGE gels, after which they were stained with Coomassie blue, and scanned. Protein extraction efficiency was evaluated by quantifying protein amounts from scanned gel images using NIH ImageJ software. The extractions were repeated four times. Data (n=4) are plotted as shown in the Figure 3 with standard deviations noted.

Results show that ionic detergents and chaotropic salts are most effective for protein extraction. Although extraction with RIPA buffer and CHAPS/Urea solution were somewhat less effective than 2% SDS, they have advantages of downstream sample processing. Consequently, for IEF-LC-MALDI the CHAPS/Urea solution was used. For LC-MALDI or direct spotting-MALDI, RIPA buffer is used.

Figure 3 Extraction of Blood, Saliva and Semen from Different Materials. Blood, semen and saliva were extracted from cotton, polyester, wood, plastic beverage bottle and condom using 5 different extraction solutions. Equal volumes of extracted proteins were loaded on 15% SDS PAGE gels, stained with Coomassie blue, and scanned. Relative abundance of total protein recovered was calculated using NIH ImageJ software, n=4.

● *PROTEIN EXTRACTION BUFFER COMPATIBILITY WITH DNA TESTING:* The ability to extract sufficient quantities of high quality DNA from vanishingly small samples is paramount to forensic testing laboratories. Consequently, serology testing on such samples must use methods that are compatible to DNA testing. In our research we tried four different extraction buffers and found that detergents and/or chaotropic salts gave all around the best results (see buffer components below). For DNA extraction the NYC OCME currently uses Qiagen M48 Digestion Buffer G2 which also contains detergents and chaotropic salts (see buffer components below) along with proteinase K. The Qiagen G2 buffer, however, would not be suitable for serology

testing as it contains both proteinase K (which would completely digest proteins) and 0.8 M Guanidine-HCl, which would denature trypsin, a necessary reagent in the preparation of samples for MS protein analysis. We have not attempted to isolate DNA from extracts prepared with our protein extraction buffers. However, considering that both DNA and protein extraction buffers function using similar principles, it should be possible to optimize an extraction buffer capable for isolating both types of biomolecules.

Currently, however, there are several products on the market that can simultaneously isolate DNA, RNA and proteins. These use both organic (e.g. Trizol) and non-organic (e.g. Qiagen) methods. Of course, validation experiments would need to be performed to determine if combined extraction methods give similar results to individual extractions methods in terms of quality and quantity of recovered DNA and proteins.

● *Components of Protein Extraction Buffers Used in Our Experiments*

1. PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4 pH 7.4, 2.7 mM KCl, 137 mM NaCl) pH 7.4
2. 2% SDS in Tris buffer (25 mM Tris-HCl pH 7.4, 2.7 mM KCl, 137 mM NaCl)
3. RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% Sodium deoxycholate)
4. 4% CHAPS, 7 M urea and 2M thiourea

● *QIAGEN M48 DIGESTION BUFFER G2 (USED IN CONJUNCTION WITH PROTEINASE K)*

0.8 M Guanidine-HCl, 30 mM EDTA, 30 mM Tris-HCl, 5% Tween-20, 0.5% Triton X-100, pH 8.0

Finally, we have observed that the urea/thiourea buffer gives the best results when extracting saliva from mock forensic samples. We do not believe that including urea in a "universal" extraction buffer for serology testing would cause a problem for urine testing, as a MS urine test would be based on identifying urine specific or enriched proteins (Adachi et al. 2006) rather than urea.

4. Limits of Detection for Pure Body Fluids – Our results (Figure 4) demonstrate that the LC-MALDI method using a purification step has equal or greater sensitivity for body fluid detection than do chemical, immunochemical, enzymatic and mRNA methods. Our marker proteins (Table 2) could be identified from 2 ng of blood and semen, and 10 ng of saliva, which corresponded to 0.006 nl blood, 0.1 nl of semen and 4 nl of saliva.

The marker proteins of semen, saliva and blood can also be identified by the less time consuming direct spotting MALDI method without prior protein purification, here digested body fluid proteins are directly spotted on MALDI target plates and analyzed by mass spectrometry. To determine limits of detection of pure body fluids using the direct-spotting MALDI method, serial dilutions of unpurified digested blood, semen and saliva samples were directly assayed by MALDI MS at following total protein levels: 100 ng, 50 ng, 20 ng, 10 ng, 5 ng, and 2 ng. Semen markers, SEMG 1 and SEMG 2, could be detected at the 2 ng level (0.1nl semen). Blood markers, HBA and HBB, were also detected when 2 ng level (0.006 nl blood), and AMY1, a saliva marker, was detected at the 5 ng level (2 nl saliva). Of course, the number of marker peptides detected by this method was less that observed when purified samples were used.

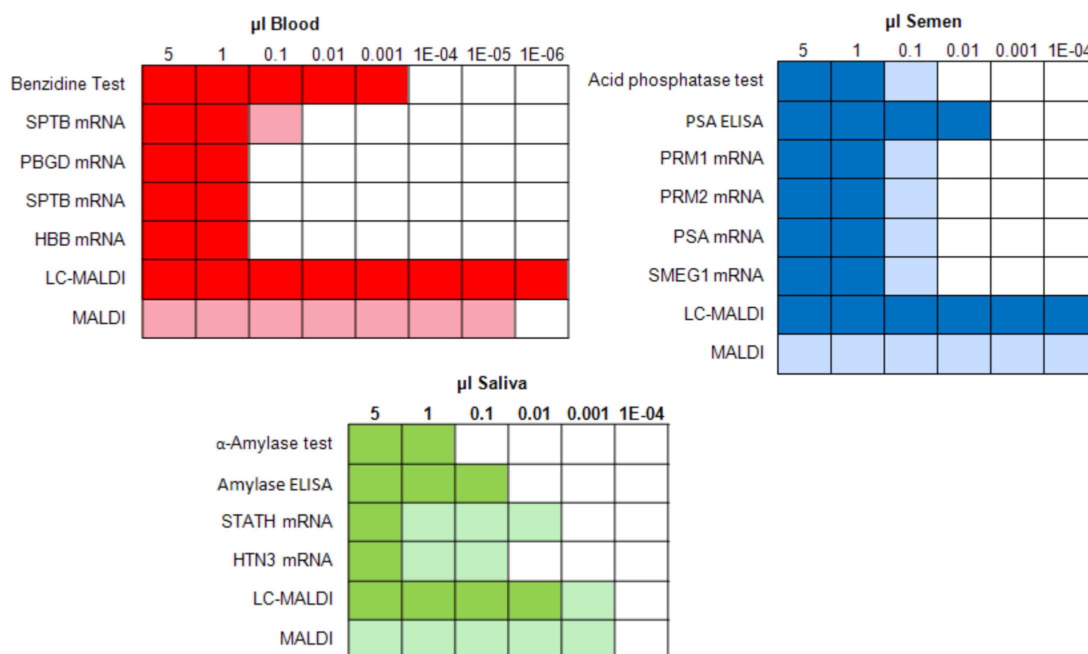


Figure 4 Sensitivity of MALDI Compared to Conventional Chemical, Immunochemical, Enzymatic and mRNA Profiling Methods for Body Fluid Identification. Light colored squares represent results obtained with only 1 to 2 of 3 experiments performed by mRNA assays, or not all marker proteins were found by MALDI (Table 2). Refs: Haas, C., and A. Kratzer, et al. 2009; Haas. C., W. Ba¨r et al. 2008

5. Limits of Detection in Mixed Body Fluid Samples - Body fluids are composed of hundreds of proteins present in an extremely wide range of concentrations. Mixtures of body fluids dramatically increase the complexity of the proteins present in a sample and simultaneously dilute target proteins.

We have established methods to extract, solubilize and detect marker proteins from individual body fluids. Here we determined the limits of detection for each body fluid when present in mixtures by identifying marker proteins for each body fluid.

The ability to distinguish two or more body fluids in a sample depends on three factors: i) the abundance of each marker within its body fluid, ii) the ratio of body fluids in a mixture - based on relative protein amounts, and iii) the ionization efficiency of marker peptides. Thus, a body fluid with a high protein concentration in which markers predominate will be detected more easily in a mixture than a body fluid with a low protein concentration in which markers make up a smaller percentage of total protein. As may be seen in Table 7, protein concentration and protein abundance of blood, semen and saliva are extremely different from each other.

Table 7 Protein Concentrations and Relative Abundance of Markers

Body Fluid	Protein Concentration	Predominate Marker	Marker % Total Protein
Blood	350 mg/ml	α & β hemoglobin	~30%
Saliva	2.5 mg/ml	α-amylase 1	~8%
Semen	20 mg/ml	Semenogelin 1 & 2	~12%

Each body fluid was mixed with a second body fluid at varying ratios (Table 8). Mixtures were solubilized with the urea/thiourea extraction buffer using a FastPrep reciprocating homogenizer. Extracted proteins were separated by isoelectric focusing (IEF). The IEF fractions containing body fluid marker proteins were digested with trypsin, separated by nano-LC and analyzed by MALDI-TOF/TOF.

As may be seen in Table 8, the ability to detect two body fluids in a mixture primarily depends on the ratio of protein amount in the two body fluids. On a volume/volume basis, results show that blood is detectable when it was diluted 1000-fold in saliva. However, in saliva/blood mixture, saliva could only be detected in blood at a ratio 1:1. In a semen/saliva mixture with 1/500 mixing ratio, semen still could be detected, but saliva could only be detected when its ratio to semen is greater than 1:10.

Importantly, the actual volume of body fluids needed to make body fluid identification is very small. Our data demonstrated that blood could be detected in mixture when 3 nl of blood was added to 50 nl of semen, or to 500 nl of saliva. Both blood and saliva could be detected in 5 nl of a 1:1 blood/saliva mixture, and in 400 nl of a 1000:1 saliva/blood mixture.

Table 8 Ratio of Body Fluids Detected by IEF-LC-MALDI

Body Fluid Mixture	Mix Ratio (Volume)	Mix Ratio (Protein Amount)	Marker Protein Found
Blood/Saliva	1 /1	140/1	AMY1, HBB, HBA, SLC4A1, SPTA1

Saliva/Blood	1000/1	7 /1	AMY1, HBB, HBA, CST2, C6orf58, SLC4A1
Blood/Semen	5 /1	87/1	HBB, HBA, SLC4A1, SPTA1, SPTB, KLK3, SEMG2, SEMG1
Semen/Blood	200 / 1	11 /1	HBB, HBA, SEMG1, SEMG2, KLK3, ACPP, MUC6, CRISP1, PAEP, HBG, SLC4A1
Saliva/Semen	500 / 1	62.5/1	AMY1, HBB, SEMG1, SEMG2, KLK3, CST2, C6orf58
Semen/Saliva	10 /1	80/1	AMY1, HBB, SEMG1, SEMG2, KLK3, ACPP, MUC6, CRISP1, PAEP

AMY1, alpha-amylase1; HBB, Hemoglobin subunit beta; HBA, Hemoglobin subunit alpha; SLC4A1, Band 3 anion transport protein; SPTA1, Spectrin alpha; CST2, Cystatin-SA; C6orf58, Uncharacterized protein C6orf58; SEMG1, Semenogelin isoform 1; SEMG2, Semenogelin isoform 2; KLK3, Prostate-specific antigen; MUC6, Mucin 6; CRISP1, Cysteine-rich secretory protein 1; ACPP, Prostatic acid phosphatase; PAEP, Progestagen-associated endometrial protein.

Results in Table 8 demonstrated that using the IEF-LC-MALDI method, we could identify nearly all body fluid markers (i.e. multiple markers for each body fluid) over a wide range of mixture ratios. However, the use of IEF prior to LC significantly increases the time required for processing sample. Consequently, we re-evaluated the limit of detection of body fluid/body fluid mixtures by the LC-MALDI method in which only LC was used for sample separation.

Results in Table 9 demonstrates that the body fluids can be detected in mixtures by the LC-MALDI method at nearly identical ratios to what can be determined by the IEF-LC-MALDI method, although, fewer numbers of body fluid markers were identified. For example, while we were able to detect six marker proteins in a saliva/semen mixture at a ratio of 50/1 (3 for saliva [AMY1, CST2, C6orf58] and 3 for semen (SEMG1, SEMG2, KLK3) by the IEF-LC-MALDI method, we could only detect four markers, two for saliva (AMY1, CST2) and two for semen (SEMG1, SEMG2) using LC alone. Without IEF separation, saliva could not be detected in a semen/saliva mixture at a ratio of 10/1.

Table 9 Ratio of Body Fluids Detected by LC-MALDI

Body Fluid Mixture	Mix Ratio (Volume)	Mix Ratio (Protein amount)	Marker Protein Found
Blood / Saliva	1/1	140/1	AMY1, HBB, HBA, SLC4A1
Saliva / Blood	1000 / 1	7/1	AMY1, HBB, HBA, CST2 C6orf58
Blood / Semen	5/1	87/1	HBB, HBA, SLC4A1, SPTB, KLK3, SEMG2, SEMG1
Semen / Blood	200 / 1	11/1	HBB, HBA, SEMG1, SEMG2, KLK3, ACPP, MUC6, PAEP
Saliva / Semen	500 / 1	62.5/1	AMY1, SEMG1, SEMG2, CST2

Semen / Saliva	10/1	80/1	HBB, SEMG1, SEMG2, KLK3, MUC6, PAEP *No saliva markers detected.*

AMY1, alpha-amylase1; HBB, hemoglobin subunit beta; HBA, hemoglobin subunit alpha; SLC4A1, band 3 anion transport protein; CST2, cystatin-SA; C6orf58, uncharacterized protein C6orf58; SEMG1, semenogelin isoform 1; SEMG2, semenogelin isoform 2; KLK3, prostate-specific antigen; MUC6, mucin 6; PAEP, progestagen-associated endometrial protein.

6. Evaluation of Aged Mock Forensic Samples

We processed samples of blood, saliva and semen which were stained on cotton and aged for 14 months (recent data for some samples aged for 20 months are presented below). Samples were extracted with RIPA buffer and run on 15% SDS PAGE gels to estimate the levels of protein degradation. Equal amounts of freshly prepared body fluid samples were run alongside as positive controls. Aged samples were also digested, separated by nano-LC and analyzed by MALDI TOF/TOF. As shown in Figure 5, after aging for 14 months, some protein degradation occurred, however, all aged body fluids could be identified by MALDI at similar levels of sensitivity as freshly prepared samples.

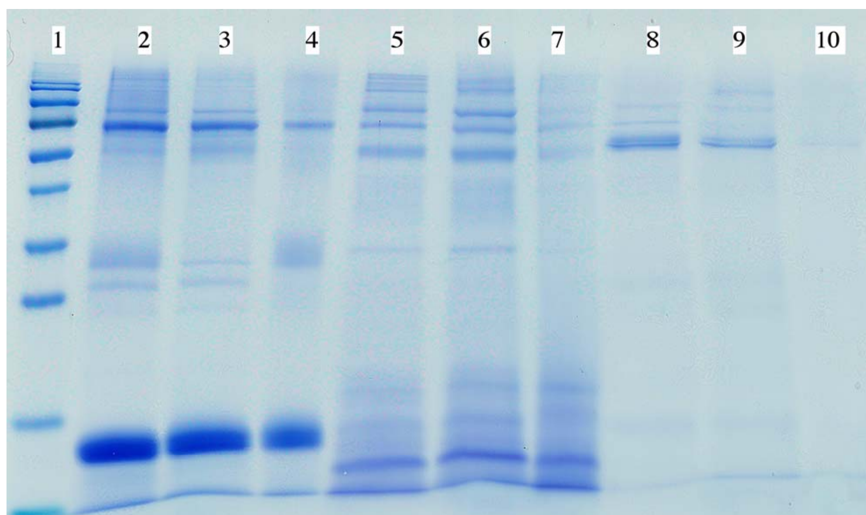


Figure 5 SDS PAGE Analysis of Proteins Extracted from Aged Body Fluid Samples Stained on Cotton. Blood, semen and saliva were extracted with RIPA buffer and run on SDS PAGE gels alongside equal amounts of freshly prepared samples. Lanes: 1 protein markers, 2 fresh blood, 3, blood stained on cotton 3 hrs, 4 blood stained on cotton 14 months, 5 fresh semen 6, semen stained on cotton 3 hrs, 7 semen stained on cotton 14 months, 8 fresh saliva, 9, saliva stained on cotton 3 hrs, 10 saliva stained on cotton 14 months.

The samples of blood stained on 5 materials and aged for 20 months were also analyzed by SDS-PAGE. Figure 6, shows protein degradation as diffuse background staining in lanes 2-5, as well as a diminution of major band intensities. However, after aging for 20 months, blood, semen and saliva can still be detected at nl level by MALDI.

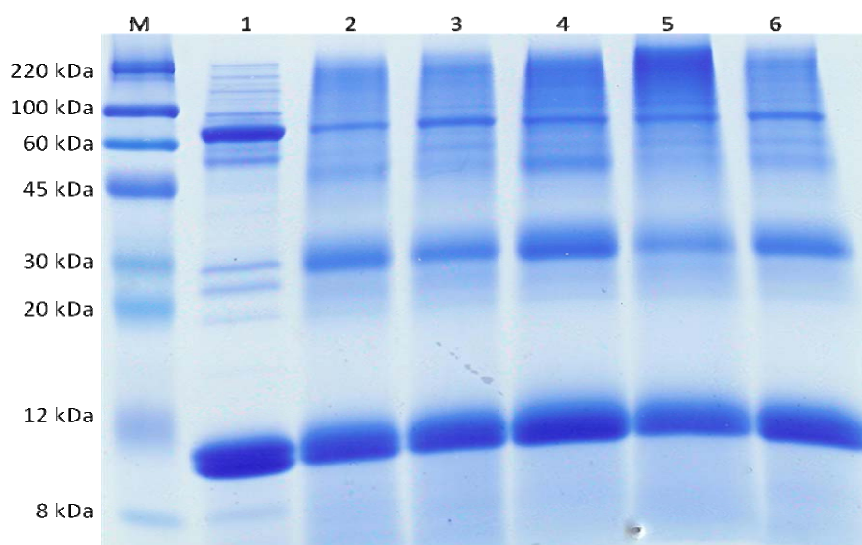


Figure 6 SDS PAGE Analysis of Proteins Extracted from Blood Aged 20 Months on Different materials. Blood was extracted from five materials, proteins quantified and equal amounts separated on a 15% PAGE gel. Lanes: M, protein markers; 1, fresh blood ; 2, cotton; 3, wood; 4, condom; 5, polyester; 6, PETE (beverage bottle plastic).

We also prepared aged body fluid samples that had been spiked with isotopically labeled peptides. To 5 µl of blood 5 nmoles of corresponding isotopically labeled marker peptides were added, and to 10 µl of semen 0.5 nmoles of corresponding isotopically labeled marker peptides were added. Two microliters of spiked samples was applied to cotton and allowed to age for 3 months. Evaluation of spiked mock samples is currently underway.

7. Menstrual Blood - We proposed to use matrix metalloproteases 7, 10 and 11 (MMP-7, 10 & 11) as menstrual blood markers as mRNAs for these proteins have been shown to be present in menstrual blood. However, we have not been able to detect these proteins. Consequently, we have taken two independent approaches to resolve this problem, 1) we have tried alternative methods of preparing menstrual blood samples that might enrich for MMPs, and 2) we have established a complete menstrual blood proteome for finding other specific protein markers.

7.1 ALTERNATIVE METHODS FOR PREPARING MENSTRUAL BLOOD SAMPLES FOR ANALYSIS:

7.1.1 HEMOGLOBIN, ALBUMIN AND IMMUNOGLOBULIN DEPLETION FROM MENSTRUAL BLOOD: The predominant proteins in menstrual blood are hemoglobin (**Hb**), albumin (**Alb**) and immunoglobulins (**Igs**). As these common and abundant proteins can conceal the presence of less abundant menstrual blood specific proteins, samples of menstrual blood were depleted of Hb, Alb and Igs in the hope of unmasking the MMP marker signals.

Figure 7 shows the depletion of Hb, Alb and Igs from blood and menstrual blood. Samples were sequentially depleted of Alb and Igs using ProteoPrep Immunoaffinity Alb and IgG Depletion Kit (Sigma-Aldrich, St. Louis, MO), followed by Hb depletion using Ni-NTA Agarose (Qiagen). (Ringrose JH, Solinge WW, et al. 2008). Starting materials may be seen in lanes 1. Alb and IgG heavy chain co-migrate on PAGE gels and are indicated by Al, hemoglobin is designated by Hb. Lanes 2 shows samples after Alb-IgG depletion. Lanes 3 shows proteins released from Alb/IgG affinity columns. Samples were then depleted of Hb by Ni affinity chromatography. Lanes 4 shows Hb depleted samples, lanes 5 shows proteins eluted from the Ni-NTA Agarose. Depleted samples were used for MALDI TOF/TOF analysis in order to evaluate their entire proteomes. However, none of the menstrual blood specific MMPs described in the literature was detected in the depleted menstrual blood sample,

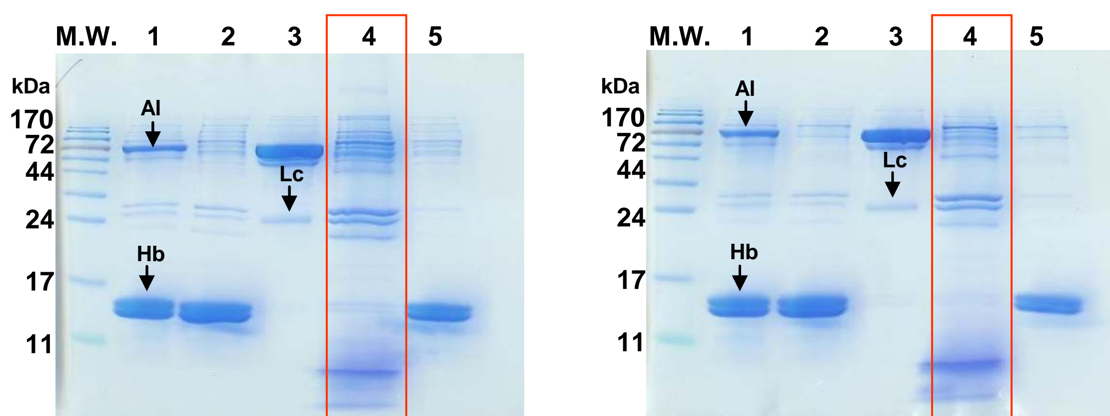


Figure 7 15% SDS-PAGE gels showing the depletion of Hb-Al-IgGs from blood (A) and menstrual blood (B). Lanes: 1. Starting materials (blood and menstrual blood), 2. Al-IgG depleted samples, 3. Proteins eluted from the Al-IgG immunoaffinity columns, 4. Hb depleted samples (Final), 5. Proteins eluted from the Ni affinity column. M.W. = molecular weight standards in kDa, Al = albumin & IgG heavy chain, Hb = hemoglobin, light chain = IgG light chain. 15 µg of protein were loaded in each lane.

7.1.2 PROTEIN IDENTIFICATION IN MENSTRUAL BLOOD PELLETS: After solubilizing and centrifuging whole menstrual blood samples, a small pellet is always observed. As MMPs are glycosylated membrane bound proteins, it is possible that they are poorly solubilized and disproportionately located within the pellet. Therefore, we extracted the menstrual blood pellet with 2% SDS sample buffer, and separated pellet proteins on a 15% SDS PAGE gel. Gel lanes were cut into small pieces (0.5 cm) and the proteins from each gel piece were in-gel digested with trypsin, subjected to LC separation and analyzed by MALDI TOF/TOF. MMP 7, 10 and 11 were not detected in the menstrual blood pellet.

7.1.3 DEGLYCOSYLATION OF MENSTRUAL BLOOD PROTEINS: As mentioned the matrix metalloproteinases are glycoproteins. Glycosylation adds complexity to proteins and, consequently, protein analysis. There are several reasons for this. First, the carbohydrate trees added to each protein are typically not identical, and since glycans typically carry an electric charge, variable glycosylation can cause proteins to lose focus on IEF – effectively diluting them to a point where they cannot be detected. Second, glycosylation can limit protease digestion efficiency (complete trypsin digestion an important step in MS analysis), and consequently, highly glycosylated proteins may not be fully represented in a proteome analysis. Third, even when digested, glycopeptides typically do not ionize completely during MS ionization, and consequently tend to have lower detection sensitivity. Finally, glycopeptides masses are not typically available in protein databases, so glycopeptide information is lost.

Thus the complexity of glycosylation could explain why original markers MMP 7, 10 & 11 were not identified by mass spectrometry. To test this hypothesis, we depleted menstrual blood of Hb, Alb and Igs, and treated it with a mixture of glycosidases (Enzymatic Protein Deglycosylation Kit, Sigma-Aldrich, St. Louis, MO). The deglycosylated samples were then fractionated by IEF, digested by trypsin, subjected to LC separation and analyzed by MALDI TOF/TOF.

A comparison of deglycosylated with non-deglycosylated control samples demonstrated the efficacy of this approach. As shown in Figure 8, deglycosylation both tightened protein focusing on IEF and increased protein recovery. For example, progestagen-associated endometrial protein (PAEP) is a menstrual blood enriched glycoprotein with a calculated pI of 5.21. In non-deglycosylated menstrual blood, PAEP was found in seven IEF fractions in two broad peaks covering pH ranges 4.0-5.0 and 7.0-8.3. Following deglycosylation, PAEP focused in three IEF fractions in a single pH range (5.0-5.9) which is consistent to its theoretical pI value of 5.21. These data indicate PAEP is differentially glycosylated which causes it to lose focusing on MMP, and consequently reduces one's ability to detect it in small sample sizes. Other proteins showed similar improvements in focusing and detection following deglycosylation. However, we were still unable to detect MMP 7, 10 & 11 in deglycosylated menstrual blood.

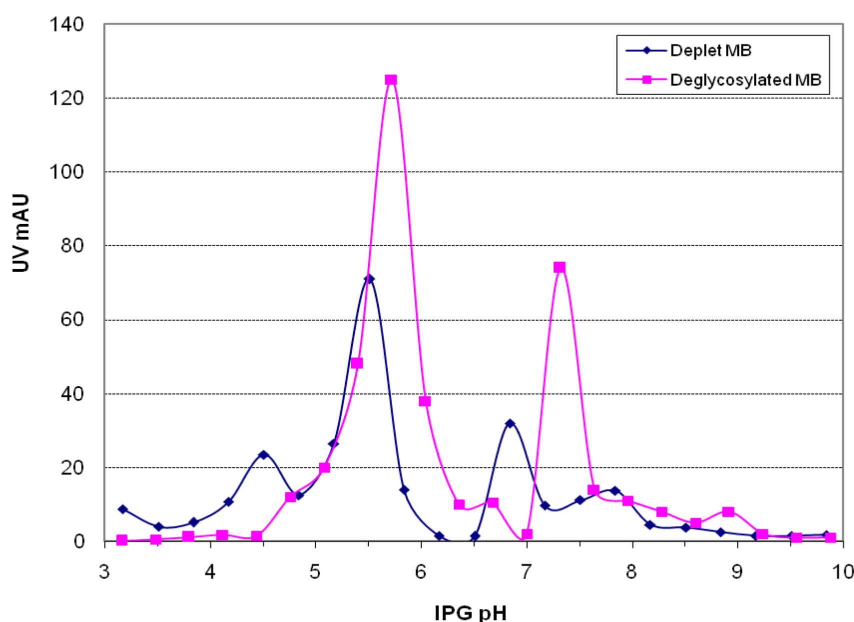


Figure 8. Comparison of Deglycosylated and Non-Deglycosylated Control Menstrual Blood by IEF

7.2 Establishing a COMPLETE MENSTRUAL BLOOD PROTEOME - Because MMP proteins 7, 10 and 11 were not detected in menstrual blood (only their mRNAs have been reported), we chose to establish a complete menstrual blood proteome using a variety of sample purification and MS methods.

As may be seen in Table 10, a total of 1,067 individual proteins were identified from six methods. Depletion of Hb, Alb and IgG from menstrual blood samples significantly increased the numbers of proteins identified form 67 to 129. Fractionation of depleted samples by both SDS PAGE and IEF doubled again the numbers of protein identified. There were 631 proteins identified in menstrual blood pellet.

Table 10. Methods and Sample Preparations for Menstrual Blood Proteome Discovery

Menstrual Blood Sample	Method	Proteins Detected (>95% Confidence)
Whole Menstrual Blood	LC- MALDI	67
Alb, IgG and Hb Depleted	LC-MALDI	129
Alb, IgG and Hb Depleted	PAGE - LC- MALDI	243
	PAGE -LC- ESI Orbitrap	270

Alb, IgG and Hb Depleted	IEF -LC- MALDI	276
Menstrual Blood Pellet	PAGE -LC- MALDI	631

To our knowledge this is the first proteome database for menstrual blood. In order to identify menstrual blood specific markers, i.e. to eliminate proteins common to venous blood as well as the four other body fluids, we compared the 1,067 menstrual blood proteins against the proteins we identified in blood, vaginal fluid, saliva and semen, as well as protein databases for these four body fluids found in the literature and online. Five candidate markers were identified (Table 3). However, to be useful, these markers must be present during the length of a woman's period, and consistently identified in all women. In our current grant we are testing the value of these markers in fifty women over the course of their entire periods.

8. Vaginal Fluid – Mucin 4 isoform a (MUC 4) and human β-defensin 1 (HBD 1) were the proteins we proposed to use as specific markers for vaginal fluid. As was the case with menstrual blood, these markers were originally identified in vaginal fluid by mRNA, and, as with menstrual blood, we are unable to detect either protein by mass spectrometry. We are not alone. The vaginal fluid proteome has been studied by other labs in hope of identify protein markers that might prove useful for diagnosis of reproductive health problems and cancers. Of the several publications and online protein databases, only one found MUC 4 in vaginal fluid (Shaw, J. L. V., Smith C. R., et al. 2007). In addition, this protein has also been identified in human plasma (http://www.biosino.org/bodyfluid/) making it less useful as a vaginal fluid specific marker. No group has identified HBD1 in vaginal fluid. There are several reasons why MUC 4 and HBD 1 proteins could not be detected in vaginal fluid, i) their mRNAs may not be translated, ii) their mRNAs may be translated but at extremely low levels and therefore expressed proteins are rare, or iii) these proteins have post-translational modifications that interfere with their detection.

As with menstrual blood, we have chosen to look for new vaginal fluid specific markers by establishing a vaginal fluid proteome. To date we have used three methods to achieve this goal: i) LC-MALDI, and ii) IEF + LC-MALDI, and iii) Alb-IgG depletion +IEF + LC -MALDI. These methods have identified 60, 504 and 227 proteins respectively, and a total of 687 individual proteins, seven of which appear to be promising candidates (see in Table 3). Again, as with menstrual blood, we are currently examining their expression in fifty women to determine if they are consistently expressed, and present during different time between a woman's period.

9. Assay Costs: Body fluid identification by mass spectrometry is a sensitive, confirmatory test that can be used to identify unknown samples in a single assay. Here we developed three MS methods to assay body fluids: i) Direct Spotting, ii) LC-MALDI and iii) IEF-LC-MALDI. Brief descriptions of these assays, costs, time (from extraction to analysis) and advantages of each are described below. Sample extraction is the same for all methods. Labor costs are not included; however, because the MS assay identifies all body fluid in a single test, labor costs are expected to be significantly less than current body fluid assays which require separate tests for each body fluid. Costs outlined below are based on workflows from sample extraction through data acquisition.

The NYC OCME processes over 25,000 serology tests annually (blood, semen and saliva), and so our objective is to automate the process by using a Biomek liquid handling device and BioTek plate reader/spectrophotometer. However, this is not necessary for smaller case numbers, in fact we have been processing research samples “by hand” from the inception of this project and it is simpler and more efficient than employing the current agarose diffusion (saliva) and ELISA (semen) methodologies, and would require less analyst time than these methods. Cost breakdowns are based on reagents/consumables used, and while there would be an “economy of scale” savings with regard to labor by automating this process for large sample numbers, smaller scale non-robotic testing would still be ~$1/sample for reagent and consumables. With regard to chain of custody and all other standard operating procedures, we did not take these into account as they are part of the current processing of serology samples and would not change.

i) DIRECT SPOTTING: In the direct spotting method a 96-well plate format is used throughout in conjunction with a Biomek liquid handling robot. Samples are extracted in 96-well plates (MP Biomedical FastPrep homogenizer), spun, and aliquots removed for protein quantitation by the Bradford method using a BioTek plate reader/spectrophotometer. Sample digestion (trypsin) is performed with the Biomek robot, as is mixing with matrix, and spotting on MALDI assay plates. Four 96-well plates are run simultaneously and 384 samples are spotted on the MALDI plate for MS analysis.

Direct spotting is high-throughput, time and cost-saving method that we estimate could process between 600-700 samples/week at a reagent/consumable cost of <$1.00/sample, making this confirmatory method competitive with current (and often non-confirmatory) methods. Detailed estimations of consumable/reagent costs as well as assay time are listed in the tables below and are based on running 384 samples.

Direct Spotting Method - Cost Estimate For 384 Samples		
Item	Quantity	Cost Estimate
1. FastPrep Matrix D	40ml	$100.00
2. RIPA buffer	38.4 ml	$25.00
3. Deep well 96-plate	4 plates	$2.00
4. 96-microplate	16 plates	$8.00
5. 96-tip box	16 boxes	$80.00
6. Bradford reagent	38.4ml	$40.00
7. Ammonium bicarbonate	320mg	$4.00
8. Urea	100g	$12.00
9. DTT	80mg	$4.00
10. IAA	90mg	$7.00
11. Trypsin	20ng/ul 2ml	$20.00
12. Acentonitrile	50% solution 100ml	$5.00
13. TFA	100ul	$1.00
14. MALDI Matrix	CHCA 10mg	$15.00
	Total Estimate	$323.00

Direct Spotting Method - Time Estimate for 384 Samples	
1st day	extraction, concentration measurement, in-solution digestion
2nd day	peptide purification, spotting, MALDI data collection
3rd day	data analysis, reporting

ii) LC-MALDI: The LC-MALDI method is most useful for identifying mixed samples or those samples in which specific body fluid biomarkers are not the most abundant proteins, e.g. vaginal fluid and menstrual blood. This method uses nano-HPLC to separate marker peptides into distinct fractions so that they can be detected. Each run takes approximately one hour, followed by a 30-minute column wash. By staggering sample processing we estimate that eight samples can be run/week. Newer ultra-HPLC instruments which can complete a run in < 10 minutes, as well as dramatically faster mass spectrometers (e.g. ABI 5800 MALDI TOT/TOF) would significantly improve throughput of this method.

Workflow is as follows: Samples are extracted, quantified and digested similarly to what is done in the direct spotting method (above) except only four samples are processed at a time. After digestion is complete, samples are loaded sequentially onto a nano-HPLC (with a washing step in between) and eluted by acetonitrile gradient onto a MALDI plate in 96 fractions. After all

four samples have been spotted onto the plate (384 fractions in total) it is loaded into the MALDI MS for analysis. At this time, a second group of four samples could be started. We estimate that it would take five days to run eight samples and cost approximately $11/sample (see Tables below). The major difference in costs between this method and direct spotting are the reagents and consumable for the nano-HPLC.

LC-MALDI Method - Cost Estimate for 4 Samples		
Reagent/Consumable	Quantity	Cost Estimate
1. FastPrep Matrix D	1ml	$2.50
2. 96-tip box	1boxes	$5.00
3. RIPA buffer	0.4ml	$0.33
4. Bradford reagent	2ml	$2.00
5. Ammonium bicarbonate	3mg	$0.04
6. Urea	1g	$0.12
7. DTT	1mg	$0.05
8. IAA	1mg	$0.08
9. Trypsin	20ng/ul 40ul	$0.40
10. Acentonitrile	100ml	$10.00
11. water HPLC grade	100ml	$2.50
12. TFA	200ul	$2.00
13. HPLC column	4 run	$10.00
14. Sample vials	4 vials	$2.00
15. MALDI Matrix	CHCA 5mg	$7.50
	Total Estimate	$44.52

LC-MALDI Method Time Estimate for 4 Samples	
1st day	extraction, concentration measurement, in-solution digestion
2nd day	LC run, plate spotting
3rd day	MALDI data collection
5th day	data analysis, reporting

iii) IEF-LC-MALDI: Because IEF-LC-MALDI separates proteins prior to separating their constituent peptides, it is the most sensitive method for identifying rare, low abundant markers that might otherwise be obscured in samples. As this method requires multiple techniques, it takes more time and fewer samples can be processed. This method is predominately used for

research purposes to identify new biomarkers in complex body fluids that have abundant non-specific proteins, e.g. hemoglobin, albumin and immunoglobulins in menstrual blood which make identification of specific markers difficult. Once specific body fluid markers are identified, simpler bulk, upstream separation techniques followed by LC-MALDI can typically be used for routine forensic casework.

Workflow: Samples are extracted and quantified as described above and then loaded onto 10 cm IPG strips for overnight protein separation. Strips are cut into 0.5 cm sections which are then individually digested and peptides subsequently separated by a nano-HPLC gradient into 96 fractions. Four pieces can be spotted onto each MALDI plate (total 384 fractions) for analysis.

We estimate that it would take ten days to run two sample, nearly half this time used for MS data collection of the 1,920 fractions generated by in IEF step (see Tables below). As with the LC-MALDI method, processing of a second sample (extraction, IEF separation and in-gel digestion) could begin during the LC-MALDI analysis stage of the first sample. Cost per sample is about $90. A detailed estimation is listed in the table below.

IEF-LC-MALDI Method - Cost Estimate for 1 Sample		
Item	Quantity	Cost Estimate
1. FastPrep Matrix D	0.2ml	$0.50
2. RIPA buffer	0.2ml	$0.20
3. 96-tip box	2 boxes	$10.00
4. Bradford reagent	2ml	$2.00
5. 10cm IEF strip	1	$6.00
6. Ammonium bicarbonate	15mg	$0.20
7. Urea	5g	$0.60
8. DTT	5mg	$0.25
9. IAA	5mg	$0.40
10. Trypsin	20ng/ul 200ul	$2.00
11. Acentonitrile	200ml	$20.00
12. Water HPLC grade	200ml	$5.00
13. TFA	0.4ml	$4.00
14. HPLC column	8 run	$20.00
15. Sample Vials	8 vials	$4.00
16. MALDI Matrix	CHCA 10mg	$15.00
	Total Estimate	$90.15

IEF-LC-MALDI Method - Time Estimate for 1 Sample	
1st day	extraction, concentration measurement, IEF run
2nd day	cutting IEF strip, in-gel digestion of IEF fractions
3rd day	LC, plate spotting
4th day	MALDI data collection
8th day	data analysis, reporting

SUMMARY: The costs of serology testing must be taken in consideration when selecting a method for routine forensic casework. It was for this reason, as well as assay time, that we chose to evaluate direct spotting as an alternative to the LC-MALDI method. We believe that in terms of costs, time and reliable confirmatory results, direct spotting is superior to current testing methods for blood, semen and saliva. The Kastle-Meyer test while fast and cheap is not confirmatory, while direct spotting costs < $1/sample, is confirmatory, and capable of processing ~600-700 samples/week if necessary. The amylase diffusion test is also a non-confirmatory assay, depends on samples retaining their biological activity (and is therefore limited by sample age and environmental conditions) and relies on the subjective interpretation of analysts. Finally, the P30 antibody test suffers from the selection of a single epitope for identification, which must retain its confirmation for detection. As P30 is also found in other body fluids and can been be found in women's blood, it is not as accurate as a test as the MS direct spotting which identifies multiple peptides of two semen specific proteins. Finally, MS direct spotting can in one test identify blood, semen and saliva, whereas current methods would require three separate tests to be performed. Consequently, we believe that MS body fluid assay are not only superior to current methods in terms of quality, but are more than competitive in terms of cost, especially as the MS test are confirmatory.

IV. DISCUSSION & CONCLUSIONS

A. Discussion of Findings: Body fluid identification plays an important role in forensic investigation. Yet unlike DNA STR typing, there is currently no single method or uniform set of standards for body fluid identification. Multiple techniques are used, even for the same body fluid, with varying levels of sensitivity and reproducibility; many techniques are not confirmatory. By necessity, these methods require multiple types of instrumentation as well as training in assay procedures and interpretation. The over arching goal of this grant was to determine if mass spectrometry offered a single viable alternative for the many diverse techniques currently

in use with respect to quality, sensitivity, time and cost. We believe our data demonstrates that it does.

We have developed a highly sensitive, confirmatory protein-based body fluid assay that can detect blood, semen and saliva simultaneously using a single, common method. Levels of detection, based on volume, are in the nanoliter range for all three body fluids. While samples aged for periods up to twenty months show some evidence of protein degradation, all can be detected at levels similar to those of freshly prepared samples because even partially degraded marker peptides can be used for identification. Samples of mixed body fluids are automatically detected by this method, as the MS technique is unbiased and identifies all markers present. However, the level at which mixtures can be distinguished depends on the ratio and nature of the body fluids in question. Blood, for example, can be detected in saliva when present at a ratio of 1:1000 (v/v) respectively, while saliva can be detected in blood only at a ratio of 1:1. Significantly improved sensitivities could be achieved, however, by introducing protein purification methods upstream of detection. We also developed a method to quantify individual body fluid contributions to mixtures using isotopically label peptide markers.

For routine daily testing of individual body fluids, experiments were initiated to develop a high throughput assay that can detect one or more body fluid markers without any prior purification techniques. Initial results are promising, and we estimate this method could process between 600-700 samples/week at an estimated reagent/consumable cost of < $1.00/sample, making this confirmatory method competitive with current, and often non-confirmatory, methods. Finally, progress was made toward establishing mass spectrometry assays for menstrual blood and vaginal fluid. Five candidate protein markers were identified for menstrual blood and seven for vaginal fluid. Work on the high throughput assay, as well as menstrual blood and vaginal fluid assays are being pursued with our new grant.

In summary, our results demonstrate that mass spectrometry is a robust, sensitive and above all confirmatory assay that can identify in a single test whatever body fluid is present in an unknown sample, thus eliminating the multiple methodologies currently needed, and reducing assay time, cost and the amount of sample consumed.

B. Implications for Policy and Practice – Current methods used for body fluid identification encompass a wide range of technologies including chemical, biochemical, immunochemical enzymatic, spectrophotometric and others. Some tests take advantage of modern scientific advances such as monoclonal antibodies, while others use chemical technologies that are a many decades old (e.g. Kastle-Meyer, Takayama), some are confirmatory, others, presumptive,

some sensitive, other not, some rapid, other require overnight incubations. This assortment of techniques presents practical laboratory as well as policy problems, as courts and juries come to expect test results that meet the high standards set by DNA testing.

The practical laboratory problems are many; some tests require biological activity, meaning that sample degradation in the field or in storage limit their usefulness. The need to perform multiple tests on an unknown sample to find out which body fluid is present is both time and sample consuming and requires multiple, and often expensive, instruments, as well as trained personnel to operate them. Finally, with progress that has been made in the field of proteomics, it is not necessary to settle for presumptive body fluid testing The demonstration that a single methodology, mass spectrometry, can rapidly, and cost effectively, identify all body fluids in a sample simultaneously and confirmatively will have implications for both the policy and practice of forensic science. The 2009 National Academy's report on forensic science has identified the need to bring all fields of forensics to the level of quality routinely provided by DNA testing, and policies to establish common standards and practices for body fluid testing are inevitable. Because mass spectrometry identifies the amino acid sequences of body fluid specific peptides, it offers a high level of certainty for the results. Also, as the use of MS becomes more prevalent, its methods more streamlined and user friendly, sample extraction and preparations "kits" may become available and in our opinion mass spectrometry may become, in practice, the gold standard of body fluid testing.

C. Implications for Further Research: Our work has not only demonstrated the usefulness of MALDI MS for the detection of blood, saliva and semen, but proof of principle that this method could be expanded to perhaps all body fluids and tissues. Preliminary results also demonstrate that the assay is amenable to high throughput automation when marker proteins are sufficiently abundant, as they are in blood, saliva and semen. In our new grant we proposed to follow up on the progress made here with the following projects:

- Establish a High Throughput Assay for Blood, Saliva and Semen – Because the protein markers for these body fluids represent a significant proportion of the total protein in each body fluid, it is possible to detect them without upstream purification techniques. This means that IEF, PAGE and LC can be eliminated, simplifying procedures and saving significant time. Here samples could be extracted in a 96 well plate format, transferred to a liquid handling robot for digestion, cleanup and MALDI plate spotting. We estimate that 600-700 samples could be processed in one week, at a cost of <$1.00/sample. Our goal is also to keep the level of sensitivity near that of purified samples, although fewer marker peptides will be detected.

• Establish Menstrual Blood and Vaginal Fluid Assays – The candidate markers we identified for menstrual blood and vaginal fluid need to be validated. Because protein expression in both body fluids can vary during the menstrual cycle, we have proposed to examine menstrual blood markers during every day of a woman's period and vaginal fluid protein twice her menstrual cycle. If confirmed, we hope to establish MS assays for both body fluids. While some upstream purification techniques will be required for both body fluids, since marker proteins for both represent a small fraction of total proteins, new technologies, such as "cartridge-based" 2D LC might allow relatively rapid identification of both.

• Species Identification – Inherent in a protein's sequence is its evolutionary descent as speciation is a result of changes in DNA and consequently protein sequence. Consequently, by careful selection, it should be possible to identify peptides that can distinguish species. Our goal here is to look for abundant proteins that can be identified with little or no upstream purification and give species identification simultaneously with body fluid identification.

• Gender Identification – Several proteins have been identified in blood that are unique products of the Y chromosome. If they are transcribed in sufficient quantities it should be possible to use them for gender identification during routine blood testing.

• Effect of Luminol on Protein Degradation – As luminol is a common reagent used at crime scene investigations, and is known to degrade DNA, we plan to evaluate its effects on proteins.

• Additional Research – Other research projects of value would include: i) establishing a protocol to simultaneously extract DNA and proteins from stained material for STR and body fluid testing, thus reducing sample size necessary for conducting both tests, ii) evaluation of other common household cleaners to determine their effect on protein degradation, iii) extend the search for body fluid markers beyond the five evaluated here, and iv) extend mixture discrimination sensitivity through new technologies that can reduce protein separation times.

V. REFERENCES

Adachi, J., et al. (2006) "The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins" Genome Biology **7**(9):R80.

Adkins, R.M. et al. (2011epub) "Racial differences in gene-specific DNA methylation levels are present at birth", Birth Defects Res A Clin Mol Teratol: Epub

Ambion Technical Bulletin 178. "RNase and DEPC Treatment: Fact or Laboratory Myth"

Andraos. C. et al. 2011 "Vitamin D receptor gene methylation is associated with ethnicity, tuberculosis, and TaqI polymorphism" Human Immunology **72**: 262–268

Anderson, N. L., M. Polanski, et al. (2004). "The Human Plasma Proteome: A Nonredundant List Developed by Combination of Four Separate Sources." Mol Cell Proteomics **3**(4): 311-326.

Bauer, M. and D. Patzelt (2008). "Identification of menstrual blood by real time RT-PCR: technical improvements and the practical value of negative test results." Forensic Sci Int **174**(1): 55-9.

Bechtel S, I. Schupp et al. (2007) “The full-ORF clone resource of the German cDNA Consortium”. BMC Genomics **31**(8): 399-399.

Blank, A., C. Dekker, et al. (1981). "Human body fluid ribonucleases: detection, interrelationships and significance." Nucleic Acids Symp Ser(10): 203-9.

Brown, T., K. Mackey, et al. (2004). "Analysis of RNA by northern and slot blot hybridization." Curr Protoc Mol Biol **Chapter 4**: Unit4 9.

Bustos, C.D. et al. (2009) “Tissue-specific variation in DNA methylation levels along human chromosome 1” Epigenetics & Chromatin **2**: 1-21

Choi, S-W. et al. (2010) "Epigenetics: A New Bridge between Nutritionand Health" Adv. Nutr. **1**: 8–16

Dasari, S., L. Pereira, et al. (2007). "Comprehensive proteomic analysis of human cervical-vaginal fluid." J Proteome Res **6**(4): 1258-68.

Dassen H, P. Groothuis et al. (2008) “Haemoglobin expression in human endometrium.” Hum Reprod. **23**(3):635-41.

Eckhardt, F. et al. (2006) “DNA methylation profiling of human chromosomes 6, 20 and 22” Nature Genetics **38**:1378-1385

Egger, G., et al. (2004) "Epigenetics in human disease and prospects for epigenetic therapy" Nature **429**, 457–463

Ehrich M, et al. (2007). "A new method for accurate assessment of DNA quality after bisulfite treatment". Nucleic Acids Res **35 (5)**: e29.

El-Maarri, O. et al. (2007) “Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males” Hum Genet **122**: 505–514

Ende, N. (1961). "Amylase activity in body fluid." Cancer **14**: 1109-1114.

Fleming,R.I., S. Harbison, (2009) “The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids” Forensic Sci Int: Genetics: in press

Frumkin, D., et al. (2010epub) “DNA methylation-based forensic tissue identification” Forensic ScienceInternational:Genetics: Epub

Ghosh, S. et al. (2010) “Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues” Epigenetics **5**: 527-538

Gilman, M. (2002). "Preparation of cytoplasmic RNA from tissue culture cells." Curr Protoc Mol Biol **Chapter 4**: Unit4 1.

Gronniger, E. et al. (2010) “Aging and Chronic Sun Exposure Cause Distinct Epigenetic Changes in Human Skin”, PLoS Genetics **6**: 1-10

Grunau C, et al. (2001). "Bisulfite genomic sequencing: systematic investigation of critical experimental parameters". Nucleic Acids Res. **29 (13)**: E65.

Guo, T., P. A. Rudnick, et al. (2006). "Characterization of the Human Salivary Proteome by Capillary Isoelectric Focusing/Nanoreversed-Phase Liquid Chromatography Coupled with ESI-Tandem MS." J. Proteome Res. **5**(6): 1469-1478.

Haas. C., W. Ba¨r et al. (2008) “mRNA profiling for body fluid identification” Forensic Sci Int: Genetics Supplement Series **1**: 37–38.

Haas, C., and A. Kratzer, et al. (2009) “mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR” Forensic Sci Int: Genetics **3**: 80–88.

Hanson, E.K., and J. Ballantyne, et al. (2009) “Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs” Anal Biochem **387**: 303–314.

Hearns-Stokes R, L. Nieman et al.(2006) “Expression of the proto-oncoprotein breast cancer nuclear receptor auxiliary factor (Brx) is altered in eutopic endometrium of women with endometriosis.” Fertil Steril. **85**(1):63-70.

Hong, S-J. et al. (2009) “DNA Methylation and Expression Patterns of Key Tissue-specific Genes in Adult Stem Cells and Stomach Tissues”, J Korean Med Sci **24:** 918-29

Hu, S., J. A. Loo, et al. (2006). "Human body fluid proteome analysis." Proteomics **6**(23): 6326-53.

Hu, S., Y. Xie, et al. (2005). "Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry." Proteomics **5**(6): 1714-28.

Inoue, H., A. Kimura, et al. (2002). "Degradation profile of mRNA in a dead rat body: basic semi-quantification study." Forensic Sci Int **130**(2-3): 127-32.

Juusola, J. and J. Ballantyne (2005). "Multiplex mRNA profiling for the identification of body fluids." Forensic Sci Int **152**(1): 1-12.

Juusola, J. and J. Ballantyne (2007). "mRNA profiling for body fluid identification by multiplex quantitative RT-PCR." J Forensic Sci **52**(6): 1252-62.

Kapp, E.A. et al. (2005) "An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis" Proteomics **5**: 3475–3490

Klein LL, KR. Jonscher KR, et al. (2008) "Shotgun proteomic analysis of vaginal fluid from women in late pregnancy." Reprod Sci **15**:263-273.

Kneller JM, SJ. Jones et al. (2007) “Using LongSAGE to Detect Biomarkers of Cervical Cancer Potentially Amenable to Optical Contrast Agent Labelling”. Biomark Insights **11** (2): 447-461.

Kohn, J. (1986). "Measurement of amylase activity--a useful indication of tissue damage after major radiation accidents?" Br Med J (Clin Res Ed) **292**(6534): 1523-4.

Liu, X., and D. E. Clemmer, et al. (2007) “Mapping the Human Plasma Proteome by SCX-LC-IMS-MS” J Am Soc Mass Spectrom **18**, 1249–1264.

Nussbaumer, C., E. Gharehbaghi-Schnell, et al. (2006). "Messenger RNA profiling: a novel method for body fluid identification by real-time PCR." Forensic Sci Int **157**(2-3): 181-6.

Ofran,Y., and J. Ritz, et al. (2010) “Diverse Patterns of T-Cell Response against Multiple Newly Identified Human Y Chromosome–Encoded Minor Histocompatibility Epitopes” Clin Cancer Res 16(5): 1642–51.

Ringrose JH, Solinge WW, et al. (2008) "Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically." J Proteome Res **7**: 3060-3063.

Pasini, E.M., et al. (2006) " In-depth analysis of the membrane and cytosolic proteome of red blood cells." Blood **108**(3): 791-801.

Pereira, L., A.P. Reddy, et al. (2007) “Identification of Novel Protein Biomarkers of Preterm Birth in Human Cervical-Vaginal Fluid” J Proteome Research **6**: 1269-1276.

Pilch, B. and M. Mann (2006). "Large-scale and high-confidence proteomic analysis of human seminal plasma." Genome Biol **7**(5): R40.

Quinzio, D. MK, K.Oliva, et al.(2007) " Proteomic analysis and characterisation of human cervico-vaginal fluid proteins." Aust N ZJ Obstet Gynaecol **47**:9-15.

Saferstein, R. (1982-1993). Forensic Science Handbook. Eaglewood Cliffs, Prentice Hall Regents.

Shaw, J. L. V. and E. P. Diamandis (2007). "Distribution of 15 Human Kallikreins in Tissues and Biological Fluids." Clin Chem **53**(8): 1423-1432.

Shaw, J. L. V., C. R. Smith, et al. (2007). "Proteomic Analysis of Human Cervico-Vaginal Fluid." J. Proteome Res. **6**(7): 2859-2865.

Shilov, I.V. et al. (2007) "The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra" Molecular & Cellular Proteomics **6**:1638–1655

Tang, L. J., F. De Seta, et al. (2007). "Proteomic analysis of human cervical-vaginal fluids." J Proteome Res 6(7): 2874-83.

Tanner, M. A., L. S. Berk, et al. (2002). "Substantial changes in gene expression level due to the storage temperature and storage duration of human whole blood." Clin Lab Haematol **24**(6): 337-41.

Venkataraman N, AL. Cole, et al. (2005) "Cationic polypeptides are required for anti-HIV-1 activity of human vaginal fluid. " J Immunol **175**:7560-7567.

Vercelli, D. **(**2004) "Genetics, epigenetics, and the environment: Switching, buffering, releasing" J Allergy Clin Immunol **113**:381-6.)

Vuillard, L., D. Madern, et al. (1995). "Halophilic protein stabilization by the mild solubilizing agents nondetergent sulfobetaines." Anal Biochem 230(2): 290-4.

Vuillard, L., N. Marret, et al. (1995). "Enhancing protein solubilization with nondetergent sulfobetaines." Electrophoresis 16(3): 295-7.

Xie, H., N. L. Rhodus, et al. (2005). "A Catalogue of Human Saliva Proteins Identified by Free Flow Electrophoresis-based Peptide Separation and Tandem Mass Spectrometry." Mol Cell Proteomics **4**(11): 1826-1830.

Yang, S. and C. M. Huang (2007). "Recent advances in protein profiling of tissues and tissue fluids." Expert Rev Proteomics **4**(4): 515-29.

Zegels, G., and X. WM Van Ostade, et al.(2009) “Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples” Proteome Sci **7**(17): 1-16

Zolotarjova, N., J. Martosella, et al. (2005). "Differences among techniques for high-abundant protein depletion." Proteomics 5(13): 3304-13.

Zubakov, D., and M.Kayser, et al. (2010) “MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation” Int J Legal Med: Epub.

VI. DISSEMINATION OF RESEARCH FINDINGS

In addition to the 2009 and 2010 NIJ Conferences held in Washington D.C. our data have been presented at three other scientific meetings: 1) The 57th Annual Conference of the American Society for Mass Spectrometry, May, 2009, Philadelphia, PA. 2) The 62nd Annual Meeting of the American Academy of Forensic Sciences, February, 2010, Seattle, WA. 3) The 58th Annual Conference of the American Society of Mass Spectrometry, May, 2010, Salt Lake City, UT.

In addition, two journal articles are in preparation: the first, *Characterization of the Human Menstrual Proteome*, the second, *Establishment of a Rapid, Confirmatory Body Fluid Assay Using MALDI Mass Spectrometry.*