Application of Raman Spectroscopy for an Easy-to-Use, on-Field, Rapid, Nondestructive, Confirmatory Identification of Body Fluids

2009 DNBXK186

Final Technical Report
to the
Department of Justice
February 29, 2012

Igor K. Lednev
Associate Professor
Department of Chemistry
University at Albany, SUNY, 1400 Washington Ave
Albany, NY 12222

Phone: (518) 591-8863
Fax: (518) 442-3462 (Chemistry)
E-mail: lednev@albany.edu
http://lednev.simpot.net/
10/29/2011
DUNS – 152652822
EIN Number – 14-1368361
The Research Foundation of SUNY, University at Albany
Office for Sponsored Programs, MSC 312
1400 Washington Avenue
Albany, New York 12222
RF # 1085330-1-53166
November 1, 2009 – December 31, 2011
Reporting Period End Date – December 31, 2011

This report describes project activity conducted at University at Albany, SUNY, which is aiming the development of a new method based on Raman spectroscopy for rapid, nondestructive, confirmatory identification of body fluid traces. The personnel involved in the project include Prof. Igor K. Lednev (PI) and Kelly Virkler PhD, Vitali Sikirzhytski PhD, graduate students Aliaksandra Sikirzhytskaya, Gregory McLaughlin and Aliea Afnan.

Abstract

According to the major goals of this project, a novel method was developed for the nondestructive, confirmatory identification of body fluid traces based on Raman spectroscopy combined with advanced statistics. Our prior hypothesis that the Raman spectroscopic signature of each body fluid is unique and can be used for identification purposes was confirmed. A library of Raman signatures for various body fluids, specifically blood, semen, vaginal fluid, saliva and sweat was built and software for the automatic identification of an unknown sample was developed. The method was expanded to include body fluid stains on various substrates and contaminated stains.

Human body fluids were collected from anonymous groups of donors. The donors' age, race and gender were disclosed to assure the required diversity and sample size. Samples of each body fluid were obtained from donors representing five races, including: Caucasian, African American, Asian, Hispanic and Native American. Dry traces of body fluids were tested using Raman microscope equipped with a computer-controlled stage for rapid mapping. It was found that dry traces of body fluids are intrinsically heterogeneous meaning that Raman spectra measured from different spots varied significantly. We proposed a new approach to identification by utilizing multidimensional spectroscopic signature to account for sample heterogeneity and variations with donor. Our developed software algorithms compare experimental Raman spectra with the library of Raman signatures, providing quantitative measures of similarity. It was demonstrated that an unknown stain of a pure body fluid can be identified with near 100% accuracy provided that good quality Raman spectra are acquired.

We demonstrated that mixed samples can be identified through automatic mapping if two body fluids are not thoroughly mixed. Thoroughly mixed samples can be identified using support vector machine (SVM)-based methods. Specifically, SVM-based method was able to identify a small contribution (2-5%) of thoroughly mixed semen in blood and blood in semen samples. As a specific objective of this project, the problem of substrate interference was also explored. We demonstrated that dry traces of blood and semen can be identified successfully on several substrates of practical importance.

Overall, this study demonstrates great potential of Raman spectroscopy for the nondestructive, confirmatory identification of body fluid traces, including stains on various

substrates and contaminated mixtures. Additionally, a program was developed for the automatic identification of body fluids in dry mixtures.

Table of Contents

Executive Summary

Synopsis of the problem and purpose

Although the identification of biological stains recovered from a crime scene is a key part of forensic investigation, currently there is no single identification method applicable to every body fluid. Several presumptive and confirmatory tests are currently in use.[1-5] In general, these tests are either destructive, require hazardous reagents and/or suffer from cross-reactivity with other biological fluids. The ***long-term goal of our research*** is to develop ***an easy-to-use, portable instrument for the rapid, nondestructive, confirmatory identification of body fluids in biological stains revealed at a crime scene.*** This instrument could also be used in a forensic laboratory for assessing the probative value of evidentiary samples. The ***main purpose of this project*** was the ***development and evaluation of the proposed novel methodology under controlled (laboratory) conditions***. Building and certifying a portable easy-to-use automatic instrument are the targets of future studies and beyond the scope of this project.

For this initial study, we chose five body fluids of high importance for forensic investigations (or science), specifically for potential DNA extraction, namely blood, semen, vaginal fluid, saliva and sweat. In addition to individual body fluids, we developed methods capable of identifying mixtures of two body fluids, samples contaminated with non-biological additives, and samples on various substrates of practical importance.

Research design

General Approach

We hypothesized that given every body fluid has a specific composition, the Raman spectroscopic signal of each body fluid is unique, and therefore can be used for identification purposes. ***Specific objectives*** of this study include: (1) obtaining statistically confident Raman spectroscopic characteristics for each body fluids, (2) developing a statistical program for the automatic identification of body fluids in dry mixtures, (3) extending the application of the developed method to body fluid stains on various substrates and body fluid mixtures contaminated with non-biological components, (4) and determining the optimum wavelength range and excitation for Raman spectroscopic measurements.

The development of the body fluid spectroscopic signatures should take into account the intrinsic heterogeneity of the stain and biochemical variations with donors. Donor group should be large and diverse to assure a statistically significant representation. Professor Howard Stratton, Department of Epidemiology and Biostatistics, School of Public Health, University at Albany advised us about the required diversity of donors for each research objective.

To take into account sample heterogeneity, we performed automatic mapping for each sample using a Renishaw inVia confocal Raman microscope with a PRIOR computer-controlled stage (WiRE 3.2 software) or, the lower plate of a Nanonics AFM MultiView 1000 system (using Quartz II and QuartzSpec software). Typically, Raman spectra were collected between 32 to 128 random spots on dry samples using 1-6 ten-second accumulations at each spot. Selected excitations within the 410 nm to 785 nm range were used for excitation. A 785-nm laser excitation was found to be the most suitable in terms of accuracy of body fluids identification and used for the majority of the reported results. The laser beam was focused on a small spot of the sample surface in the standard confocal mode with laser power varied between 11.5-115 mW. Raman spectra were recorded with a spectral resolution of about 1 cm^{-1}. Small portions (10-µL) of individual body fluids, body fluid mixtures, contaminated fluids and contaminated mixtures were dried overnight on a microscope slide covered with aluminum foil. Aluminum was used to eliminate the fluorescent interference from glass. The effect of cleaning materials was tested on semen and blood, by washing with various concentrations of all-purpose bleach (Clorox). Individual body fluids were also tested on various substrates.

Samples

Samples of blood, semen, saliva, sweat and vaginal fluid were purchased from certified companies. The certified companies disclosed the age, race and gender of the donor to assure requisite donor diversity (Table 1). Professor Howard Stratton (Department of Epidemiology and Biostatistics at the School of Public Health, University at Albany) advised us about the required number of samples from various donor groups to assure the diversity. As a practically important example of mixed body fluids and their contaminated mixtures, we chose semen and blood. Two sets of mixed samples were tested including those with and without thorough mixing before they were deposited and dried to simulate forensic evidence.

Table 1. Human body fluid sampling.

Biological fluid (Supplier name)	Partially specified samples*	Caucasian		Hispanic		African American		Native American		Asian	
		female	male	female	male	female	male	female	male	female	male
Blood (Bioreclamation, Inc)	18	2	2	2	2	2	2	1	1	1	2
Vaginal fluid (Lee Biosolutions, Inc)	2	5		-		-		-		1	
Semen (Lee Biosolutions, Inc; Biological Speciality Corporation)	50	2		2		2		-		1	
Saliva (Lee Biosolutions, Inc; Biological Speciality Corporation)	20	2	2	2	2	2	2	-	-	1	1
Sweat (Lee Biosolutions, Inc)	10	1	2	-	2	-	1	-	-	-	1
* samples for which only partial information about gender, race or age is available											

Dr. Duceman, Director of Biological Sciences at the NY State Police Forensic Investigation Center consulted us on the most relevant non-biological contaminants of body fluid stains as well as various pertinent substrates. Two types of sand, indoor dust, soil and all-purpose bleach were chosen as non-biological contaminants for blood and semen samples. Fresh portions of blood or semen on aluminum foil were completely covered with a specific contaminant and allowed to dry overnight. Raman spectra were recorded using the automatic mapping procedure previously described.

Samples containing all-purpose bleach and semen with various proportions from 1:1 to 10:1 (bleach:semen, volumes) were mixed thoroughly for 20 seconds using a Vortex Mixer, then deposited on aluminum foil and dried overnight. Similarly, mixed samples of bleach and blood were prepared with compositions of 10:1, 20:1, 50:1, and 100:1, respectively.

Diluted blood was prepared using deionized water in 2:1, 6:1, 10:1, 40:1, 60:1 and 80:1 concentrations. A typical glass microscope slide, denim from blue jeans and un-dyed cotton swatches and swabs were used as substrates. Portions of 100 µL of pure and diluted blood were deposited on un-dyed cotton swatches. A 100 µL portion of pure blood was deposited on denim samples. Aliquots of 10 and 100 µL of pure blood were deposited on a glass microscope slide. All stains were allowed to dry before data acquisition. Areas of neat substrate were sampled to obtain subtraction references.

Data analysis

Raman spectra were pre-treated using GRAMS/AI software to eliminate cosmic ray contributions from each spectrum before baseline correction. The spectra were imported into MATLAB 7.4.0 for statistical analysis and normalized to adjust for the varying amount of background interference in each spectrum.

Significant factor analysis (SFA) and PCA were performed to determine the number of principal components in the data set obtained for a single basis sample chosen randomly for each body fluid. The data were then cross-validated to confirm the number of principal components that best described the spectral data. The alternate least squares (ALS) algorithm was used to extract the individual spectral components for each body fluid. For human blood, semen, saliva and vaginal fluid, the components found for the basis sample were fitted to the average spectrum obtained from the remaining samples. The Curve Fitting Toolbox in MATLAB was used to perform the residual analysis of the difference between the fitted and experimental spectra. Characteristic spectroscopic features were assigned to the biochemical constituents of human body fluids (for example, molecules of hemoglobin in dried blood samples).

As discussed below, human sweat exhibited much higher donor variability than other body fluids tested in this study. To overcome this problem a modified approach was used: (1) multidimensional signatures were calculated using experimental Raman spectra obtained from *all* samples, (2) two data sets were formed for each of these body fluids. The first data set contained all Raman spectra after baseline correction using the adaptive and iteratively reweighted penalized least-squares baseline correction algorithm (AIRPLS). The second data set included only the fluorescent background. For each body fluid, both data sets were treated separately to extract virtually pure Raman and fluorescent components of the multidimensional signatures. The rest of the analysis was similar to that developed for other body fluid data.

Identification of unknown blood stains using multidimensional Raman signatures

The quality of body fluid identification based on the spectral similarity can be quantitatively evaluated using the sum of squares due to error (SSE), R-square (R^2) and root mean square error (RMSE), are indicating how well the spectroscopic signature fits the experimental spectra. An ideal fit would yield calculated profiles that completely match the experimental spectra. Using the Curve Fitting Toolbox in MATLAB, the intensity values for the

basis experimental spectrum and fitted spectrum were plotted on an axis as the x- and y-coordinates, respectively. All of the spectra were normalized to a maximum value of 1, making it the highest value for both the x- and y-axes. Two identical spectra would yield a scatter plot matching a line with the equation of y = x, so this line was used for comparison and was fit to the plotted data points of the basis sample evaluation to determine how close of a match the experimental spectrum and signature were. The SSE, R^2 and RMSE values represent the total deviation of the fitted data points from the experimental data. The SSE value measures the total deviation of the data points from the y = x line, and a value closer to 0 means there are fewer random errors. The R^2 value indicates how well the y = x best fit line explains variation in the data, and a value closer to 1 indicates that a higher proportion of the variance is accounted for by the line. A value closer to 1 also means that the fitted signature and experimental spectrum are a better match. Finally, the RMSE value estimates the standard deviation of the random data components. Again, a value closer to 0 indicates that the y = x line is a better fit, and that the signature better fits the experimental spectrum. The good-fit SSE and RMSE values will be closer to 0 and R^2 will be closer to 1 if the signature corresponds to a body fluid. In this case, the variations in the goodness-of-fit parameters will be caused mainly by noise contribution, and the residuals from the ALS fitting will resemble the normal distribution. A quantitative statistical analysis using SSE, R^2, and RMSE was performed to confirm a satisfactory fitting of all experimental spectra.

Findings and conclusions

Multidimensional Raman spectroscopic signatures were determined for dry traces of human blood, semen, vaginal fluid, saliva, and sweat. Raman spectroscopic signatures were tested on multiple samples from a variety of pure body fluids obtained from donors of different races, genders, and ages. These signatures were fitted to the experimental Raman spectra of various dried body fluid stains. High goodness-of-fit statistical results (SSE, R^2, and RMSE) were obtained for accurate body fluid-Raman signature pairs.

We developed software that automatically identifies a spectrum based on the multidimensional signatures using advance statistical analysis. The program analyzes a set of recorded spectra and compares them with the library of multidimensional spectroscopic signatures built for the five body fluids. Multidimensional signatures help to overcome the

intrinsic heterogeneity of dry samples and variations with donors. The obtained results were tested by extensive cross-validation.

A novel statistical approach based on combination of SVM classification and SVM regression methods was proposed to examine the composition of body fluid mixtures. The detection limit can be as low as several percents for minor contributors. The proposed methodology was successfully probed with semen-blood mixtures and is being expanded to characterize other complex mixtures.

The concept of multidimensional signatures was extended to contaminated body fluid mixtures. This approach is based on the comparison of recorded spectra with the library of multidimensional spectroscopic signatures built for the body fluids. The proposed methodology was successfully tested using semen and blood samples contaminated by sand, dust and soil.

It was decided that compiling a full library of contaminants was impractical due to the huge variability of the population. For instance, at least one hundred characteristic spectra are required to account the intrinsic heterogeneity of dust components. Instead of developing a spectroscopic signature of the impurity, we propose to analyze the data recovered from "hot spots", where the contribution of body fluid dominates.

We demonstrated that identification of unknown stains could be performed without any prior knowledge about the body fluid or the nature of the contaminant. The identification process involves determining which experimental data can be fitted by the predefined set of signatures. This concept is based on the assumption that an experimental Raman spectrum that is dominated by blood or semen contributions will be/can be fitted by a multi-dimensional spectroscopic signature; whereas a spectrum dominated by contaminant features will not.

Various data treatment and experimental approaches were evaluated to eliminate substrate interference. Substrate interference can be in different forms, depending on the specific substrate material. Specifically, the substrate subtraction methods developed relies on the ability of Raman mapping to characterize the neat substrate and the body fluid stain. Most dramatically, we demonstrated that this method has the ability to extract the signal from a trace amount of semen even when the spectra are dominated by a heterogeneous substrate. Surprisingly, we found that in common scenarios with blood stains, the Raman signal of the fluid can be extracted with modest data manipulation.

All excitation wavelengths within our capabilities (406-785 nm) were tested from the viewpoint of substrate interference. We determined that the optimum wavelength for the identification pure body fluids is 785 nm.

Implications for policy and practice

These results clearly demonstrate the great potential of Raman spectroscopy t for an easy-to-use, on-field, rapid, nondestructive, confirmatory identification of body fluid traces discovered at a crime scene. This instrument could also be used in a forensic laboratory for assessing the probative value of evidentiary samples. Our 2009 paper is the most downloaded article from *Forensic Science International*, the top journal in the field, in 2010-2011. Total of 15 articles were published or submitted for publication in peer-reviewed journals, more than 30 presentations were made at meetings conferences. Several scientific and general news agencies published press coverage about our research:

BioOptics World press coverage. SPECTROSCOPY/FORENSICS: CSI: Multidimensional Raman spectroscopy. This feature article made the magazine cover.

January 1, 2012

SPECTROSCOPY/FORENSICS: CSI: Multidimensional Raman spectroscopy

http://www.bioopticsworld.com/articles/print/volume-5/issue-1/features/multidimensional-raman-spectroscopy.html

Forensic science: CSI Raman

Raman Newsletter from SpectroscopyNOW.com, November 1, 2011

http://www.spectroscopynow.com/coi/cda/detail.cda?chId=6&id=26318&type=Feature&page=1

Albany school on cutting edge of forensics

The Daily Gazette. January 30, 2011 by Sara Foss

Seat at forensic table

A University at Albany professor has been appointed by the White House to help shape the field of forensic science for the next two decades

Times Union newspaper, January 20, 2011

http://www.timesunion.com/default/article/Opera-offers-vision-of-a-peaceful-Mideast-966144.php

UAlbany professor develops CSI Raman technique

Press coverage in *spectroscopyNOW.com*, August, 2010

http://www.spectroscopynow.com/coi/cda/detail.cda?id=23983&type=News&chId=2

University at Albany Researcher Develops CSI Technique

Press coverage in *Forensic* Magazine, July 16, 2010

http://www.forensicmag.com/news/university-albany-researcher-develops-csi-technique

UAlbany professor develops CSI technique

Times Union newspaper, July 14, 2010 by Carol DeMare

http://www.timesunion.com/AspStories/story.asp?storyID=950434&category=REGION

UAlbany Researcher Identifies Novel Method to Improve Crime Scene Investigations

Press release at the University at Albany website. July 12, 2010

http://www.albany.edu/news/release_9789.php?WT.svl=headline

Kelly Virkler, '09. Discovering New Frontiers of Forensic Science

Press release at the University at Albany website. 2010

http://www.albany.edu/news/profile_8195.php

Non-destructive spit test

Raman Newsletter from SpectroscopyNOW.com, 1 March, 2010

http://www.spectroscopynow.com/coi/cda/detail.cda?id=23147&type=Feature&chId=6&page=1

Forensic science calls woman

FLUID ANALYSIS EXPERT: Watertown High School grad making waves

Watertown Daily Times, Thursday, December 31, 2009, by Jamie Munks

http://www.watertowndailytimes.com/article/20091231/NEWS03/312319970

Blood will tell

By Hank Hogan, Biophotonics, November/December 2009, 16-17

Raman Spectroscopy: As Seen in CSI? *October 11, 2009*:

http://calvinus.wordpress.com/2009/10/11/raman-spectroscopy-as-seen-in-csi/

Species in a snap: Raman analysis of blood

A news story published by Analytic Chemistry, number one journal in the field

(by Erika Gebel, Anal. Chem., 2009, 81, 7862):

http://pubs.acs.org/doi/pdf/10.1021/ac901827u

http://pubs.acs.org/journal/ancham

Main Body of the Final Technical Report

I. Introduction

1. Statement of the problem

Although the identification of biological stains recovered from a crime scene is a key part of forensic investigation, today there is no single method for the identification of every body fluid. Several presumptive and confirmatory tests are currently in use.[1-5] Generally, these tests are destructive, require hazardous reagents and/or suffer from cross-reactivity with other biological fluids. The ***long-term goal of our research*** is to develop ***an easy-to-use, portable instrument for rapid, nondestructive, confirmatory method for identification of body fluids in biological stains revealed at a crime scene.*** This instrument could also be used in a forensic laboratory for assessing probative value of evidentiary samples. The ***main purpose of this project*** was the development and evaluation of the proposed novel methodology under controlled (laboratory) conditions. Building and certifying a portable easy-to-use automatic instrument are the goals of future studies and beyond the scope of this project.

2. Literature citations and review

Analysis of body fluids for forensic purposes

Our review article (K. Virkler and I.K. Lednev) entitled ***Analysis of Body Fluids for Forensic Purposes: From Laboratory Testing to Non-Destructive Rapid Confirmatory Identification at a Crime Scene*** was published in *Forensic Science International* journal.[6] This is a 17-page (123 references) comprehensive review of current methods used for body fluid identification followed by a discussion of new prospective methods for non-destructive, confirmatory identification of body fluids at a crime scene. The major conclusion of this review article is that ***although it is most desirable for a body fluid test to be confirmatory, non-destructive, and applicable to multiple fluids, there are no methods which fit this description at the moment***. There are many current techniques for the presumptive identification of certain body fluids, and some fluids have confirmatory tests. Blood and semen, which are the two most commonly encountered body fluids at crime scenes, can be confirmed with specific immunological tests, and some of these can determine the species (human vs. animal nature) of the sample. There are numerous articles[7-15] and book chapters[2,3,5,16,17] that pertain to presumptive tests for blood. These tests can be performed either in a laboratory or at a crime scene, but most

are destructive to the sample. Some confirmatory tests for blood[18-23] can also be used at a crime scene, but again the destructive nature of the tests is a problem. The presumptive[1,4,17,24-44] and confirmatory[45-64] tests for semen are similar in nature. They are mostly destructive and sometimes have to be performed in a laboratory. Saliva has no accepted confirmatory test at the moment, and the primary form of presumptive identification[1,4,17,65-69] relies on the presence of amylase in the sample which can lead to false positive results due to the presence of amylase in other fluids. Presumptive tests for body fluids such as urine[1,4,16,17,70-78] and vaginal fluid[1,17,79-81] are unreliable due to the inconsistent nature of these fluids that increases the possibility of a false positive or false negative result. Most presumptive tests can be performed in the field, but some sample preparation such as extraction is often necessary. Many confirmatory tests must be done in the laboratory,[62] so forensic experts at the scene of a crime will not know the confirmed identity of a fluid until much later on.

There are emerging techniques to identify the presence of blood[82-94], semen[82-86,89,90,93,95-98], saliva[16,82-86,90,91,95,98-103], and vaginal fluid[83-86,89,90,95,104]. Many of these techniques depend on the analysis of mRNA to detect certain genetic markers. Much of this work has been performed by Juusola and Ballantyne[84-86,90], and several other laboratories have also contributed to the progress in this area[83,87,89,91,95]. The new method of analyzing mRNA sequences is a major breakthrough in the confirmatory identification of body fluids, especially for saliva and vaginal fluid since they did not have a previous confirmatory technique. The problem is that these new methods are in fact still destructive, and they have to be performed in a laboratory. There is no one identification method, which is confirmatory, portable, and applicable to all body fluids.

Applications of Raman spectroscopy and microspectroscopy for body fluid identification for forensic purposes

As quoted by Mann and Vickers, Raman spectroscopy "is unusually, if not uniquely, suited to be the process control star of the next century."[105] This is because "the intrinsic selectivity of Raman spectroscopy allows for accurate identification of organic, inorganic and biological species, which provides an advantage that is lacking in many other analytical techniques, such as ultraviolet absorbance and fluorescence spectroscopies".[106] The list of ***Forensic applications of Raman spectroscopy*** includes drug identification,[107] fiber[108] and ink comparison,[109] age estimation,[110] and fingerprint analysis[111]. Raman spectroscopy is based on the inelastic scattering of low-intensity, nondestructive laser light by a solid, liquid or gas

sample. ***No sample preparation is needed, and the required amount of tested material could be as low as several picograms or femtoliters*** (10^{-12} grams or 10^{-15} liters, respectively). A typical Raman spectrum consists of several narrow bands and is considered a unique signature of the material. Raman spectroscopy shows minimal interference from water which simplifies the analyses of biological fluids and their traces. Typically, strong interference from fluorescence limits the application of Raman spectroscopy.

Portable Raman instruments have recently become available and provide a possibility for analyzing forensic evidence at the scene of a crime.[112] The current price of a portable Raman instrument is around $20K. We anticipate a substantial price reduction due to laser and detector technology development in the near future.

The problem of contamination and substrates

Additional consideration as to whether trace material found on a set of items did in fact come from the source proposed for them, it is always important to investigate if any of the material in question could have arisen as a result of accidental contamination between one item and another.[113] The problems with contaminations are related to their diverse composition. Thus, indoor dust, consists mostly of human and animal hairs, human skin cells, fibers, minerals of soil and sand, etc.[114] The chemical composition of sand depends on the origin of the rock, and formation conditions of the particular type of sand. The main constituents of the sand are silica, quartz and other mineral particles. Organic fragments from decaying plants and animals, small particles of broken rocks, water and air are the main constituents of soil. [115] The components of contaminants contribute into the spectroscopic characteristics of the studied sample, making further analysis more complicated.

Raman microspectroscopy has been successfully used for many forensic purposes. However, confocal microscopy suffers from relatively poor z-direction resolution. Consequently, a Raman confocal instrument will likely detect the body fluid trace and what lies directly underneath it. In instances where the body fluid is incorporated into the substrate, as is the case with fabrics, increased spatial resolution will not eliminate substrate interference. Generally speaking, body fluids are weak to moderate scatterers, such that luminescent substrates will often contribute a signal that is an order of magnitude more intense than the fluid in question. Boyd et al. was the first group to investigate the substrate problem pertaining to the forensic identification of body fluids[8]; they discovered luminescence interference to be crippling on even

the most commonplace substrates, such as cotton, denim and glass. Their recommendation is that the identification of blood on luminescence substrates be avoided in favor of a reconstitution method. This reconstitution method involves immersing the bloodstain in water and drying the extract on a non-interfering aluminum surface for analysis. Their research demonstrated that blood is detectable at a low concentration on a non-interfering substrate, making this reconstitution method a possible solution. We are pursuing an alternative approach, in which the identification of the fluid is established automatically with spectra obtained from evidence *in situ*. A non-destructive method that can eliminate substrate interference and that is applicable to all body fluids is most desirable.

Effect of cleaning materials

Dr. Henry C. Lee wrote "Chemical evidence found at a crime scene may be composed of many different types of substances… Liquid substances may be pure liquids, mixtures or solutions, either in water or in other solvents. The most common liquids submitted to the laboratory for analysis include acids, gasoline and other types of accelerants, beverages, body fluids, poisons, cleaning fluids, and medications".[117]

In detecting latent blood at a crime scene, the reagent most often used is luminol in presumptive tests. If bloodstains are cleaned with bleach, then presumptive tests are void. [118] Bleach has the most damaging effect on DNA profiles when compared with soap or a non-chlorine disinfectant. Using the Kastle-Meyer (KM) presumptive blood test on non-porous profiles was found unreliable in comparing duplicate samples. Bleach treated materials and their respective profiles continued to decrease in quality over time, showing a raise in heterozygote imbalance (Hbx), signifying an ongoing degradation of the DNA. [119]

3. Statement of hypothesis or rationale for the research

We hypothesized that since every body fluid has its specific biochemical composition the Raman spectroscopic signature of each body fluid is unique and can be used for identification purposes. The spectroscopic signature should take into account the intrinsic heterogeneity of dry body fluids and possible variations with donors. The latter means that the donor group should be large and diverse enough to assure a statistically significant representation.

The ***major goal*** of this project was to ***develop Raman spectroscopic methodology for the identification of body fluids in the form of mixed dry traces under controlled (laboratory) conditions.*** To achieve this goal we targeted the following ***objectives***:

1. Obtain statistically confident Raman spectroscopic characteristics for various body fluids including ***blood, semen, vaginal fluid, saliva,*** and ***sweat***.
2. Develop a ***statistical program for automatic identification of body fluids in the dry mixtures***. Evaluate major performance characteristics and limitations of the developed method as presumptive and confirmatory tests.
3. Extend the application of the developed method to body fluid stains on ***various substrates*** including human skin, fabrics, carpets, hard surfaces, etc.
4. Extend the application of the developed method to the dry body fluid ***mixtures contaminated with non-biological components***.
5. Determine the ***optimum wavelength range for Raman spectroscopic measurements*** from the viewpoint of the method selectivity, sensitivity, and reduction of fluorescence interference.

II. Methods

1. Instrumentation

A Renishaw inVia confocal Raman microscope with a 20-50x long-range objective was used for all measurements. Automatic mapping (Figure 1) was performed using a PRIOR computer-controlled stage (using WiRE 3.2 software) and the lower plate of a Nanonics AFM MultiView 1000 system (using Quartz II and QuartzSpec software). Typically, Raman spectra were collected from 32 to 128 random spots on the dry sample using 1-6 10-s accumulations at each spot. Selected laser lines from 410 nm to 785 nm range were used for excitation. An excitation of 785 nm was found to be the most suitable and used for the majority of the reported results. The laser power was varied between 11.5-115 mW on the sample. The laser beam was focused on a small spot of the sample surface in the standard confocal mode (Figure 1). Raman spectra were recorded with a spectral resolution of about 1 cm^{-1} and calibrated using a silicon standard. Small droplets (10-µL) of individual body fluids, body fluid mixtures, contaminated pure fluids and contaminated mixtures were dried overnight on a microscope slide covered with

aluminum foil to eliminate the fluorescent interference from glass. Individual body fluids were also tested on various substrates.

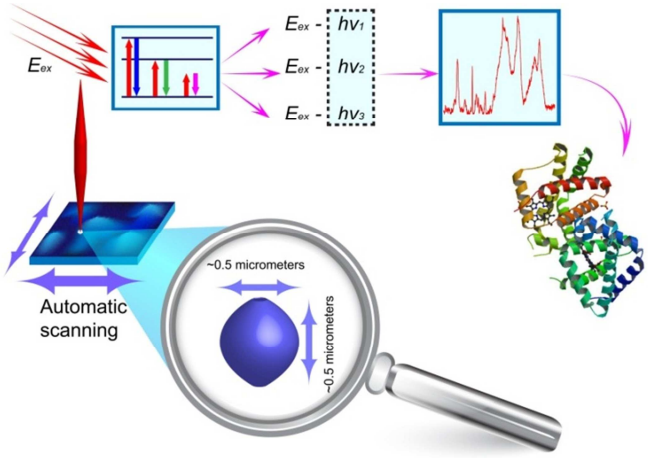

Figure 1. Theory of Raman spectroscopy and experimental design. Raman spectroscopy is based on the inelastic scattering of laser light through its interaction with vibrating molecules. Scattered photons with shifted frequencies provide information regarding molecular structure and physical state. Raman imaging is accomplished by moving the sample in a stepwise manner and acquiring spectral information from each spot.

Raman mapping of body fluid stains on different substrates was performed with a 20x objective. Multiple accumulations were performed for substrates and stains. Spectral range was set to 300-3200 cm^{-1} on the earlier stages of our studies. Spectral range 355-1730 cm^{-1} was chosen as the most informative region on the later stages of method development. Automatic cosmic ray removal was accomplished with Wire 2.0 and Wire 3.2 (software version was updated).

2. Body fluid samples

Samples of human body fluids, including blood, semen, saliva, sweat and vaginal fluid were purchased from certified companies. Samples were collected from anonymous donors whose age, race and gender were disclosed to us to assure donor diversity. Professor Howard Stratton (Department of Epidemiology and Biostatistics at the School of Public Health, University at Albany) advised us about the required number of samples from various donor groups to assure the diversity. (Table 1).

Table 1. Human body fluid sampling.

Biological fluid (Supplier name)	Partially specified samples*	Caucasian		Hispanic		African American		Native American		Asian	
		female	male	female	male	female	male	female	male	female	male
Blood (Bioreclamation, Inc)	18	2	2	2	2	2	2	1	1	1	2
Vaginal fluid (Lee Biosolutions, Inc)	2	5		-		-		-		1	
Semen (Lee Biosolutions, Inc; Biological Speciality Corporation)	50	2		2		2		-		1	
Saliva (Lee Biosolutions, Inc; Biological Speciality Corporation)	20	2	2	2	2	2	2	-	-	1	1
Sweat (Lee Biosolutions, Inc)	10	1	2	-	2	-	1	-	-	-	1
* samples for which only partial information about gender, race or age is available											

For the analysis of body fluid mixtures, samples were prepared with varying blood/semen ratio (5:95, 10:90, 20:80, 30:70, 40:60, 50:50, 70:30, 75:25, 85:15, 85.5:12.5, 92.75:6.25, 96.88:3.12 and 98.44:1.56) by thoroughly shaking for 20 seconds. Experimental data were collected and analyzed for semen/vaginal fluid mixture according to the following ratios: 50:50, 75:25, 85.5:12.5, 92.75:6.25, 96.88:3.12 and 98.44:1.56. All samples were allowed to dry overnight and Raman spectra were measured from 108 points using automatic mapping from a sample area of 3.5x2.5 mm with one 10-s accumulation at each point on a Renishaw PRIOR automatic stage with 785-nm excitation and a 50x long-range objective. The spectral resolution was about 0.8 cm^{-1}, and the CCD camera was calibrated using a silicon standard.

Mixed blood/semen samples were prepared by thoroughly mixing for 20 seconds in plastic tubes. Then, ~10 µL of each mixture was placed on an aluminum-covered microscope slide. In order to model the most challenging circumstances, heterogeneous impurities (sand, domestic dust and soil) were added in such way, that body fluid samples were covered completely when they were wet, and left to dry overnight. Samples were tested in the same manner as body fluid mixtures. Additionally, Raman spectra from particles of all types of contaminants were collected with one 10-s accumulation from each particle with 785-nm excitation, a 50x long-range objective and 100% laser power for sand and 50-10% for soil and dust.

Dilute blood was prepared using deionized water in 2:1, 6:1, 10:1, 40:1, 60:1 and 80:1 concentrations. A typical glass microscope slide, denim from blue jeans and un-dyed cotton swatches and swabs were used as substrates. Portions of 100 µL of pure and dilute blood were deposited on un-dyed cotton swatches. A 100 µL portion of pure blood was deposited on denim samples. Aliquots of 10 and 100 µL of pure blood were deposited on a glass microscope slide. All stains were allowed to dry before data acquisition. Areas of neat substrate were sampled to obtain subtraction references.

Dilutions of 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, and 8:1 of bleach:semen were mixed, as well as dilutions of 10:1, 20:1, 50:1, and 100:1 of bleach:blood, and pure samples of bleach, semen, and blood . All samples/mixtures were mixed well on Vortex Mixer for approximately 20 seconds, and aliquots of 10 µL each were dried overnight. All spectra were collected using automatic mapping technique, at 785nm with three 10-second exposure accumulations with a spectral range of 300-3200cm^{-1}, 100% laser power. Pure blood was recorded at about 11.5 mW laser power. After data collection was complete, all data was treated with GRAMS/AI software, undergoing baseline correction, and cosmic ray removal when necessary.

3. Data treatment

Raman spectra were pre-treated using GRAMS/AI software to eliminated cosmic ray contributions from each spectrum before baseline correction. The spectra were imported into MATLAB 7.4.0 for statistical analysis and normalized to adjust for the varying amount of background interference in each spectrum.

Significant factor analysis (SFA) and PCA were performed to determine the number of principal components in the data set obtained for a single basis sample chosen randomly for each body fluid. The data were then cross-validated to confirm the number of principal components that best described the spectral data. The alternate least squares (ALS) algorithm was used to extract the individual spectral components for each body fluid. For human blood, semen, saliva and vaginal fluid, the components found for the basis sample were fitted to the average spectrum obtained from the remaining samples. The Curve Fitting Toolbox in MATLAB was used to perform the residual analysis of the difference between the fitted and experimental spectra. Characteristic spectroscopic features were assigned to the biochemical constituents of human body fluids (for example, molecules of hemoglobin in dried blood samples).

A modified approach was tested for sweat and vaginal fluid Raman spectroscopic data. Two data sets were formed for each of these body fluids. The first data set contained all Raman spectra after baseline correction using the adaptive and iteratively reweighted penalized least-squares baseline correction algorithm (AIRPLS). The second data set included only the fluorescent background. For each body fluid, both data sets were treated separately to extract virtually pure Raman and fluorescent components of multidimensional signatures.

For human sweat, experimental Raman spectra obtained from all samples were used to determine the multidimensional signature to overcome the effect of significant spectral variations between donors. The residual analysis was similar to that used for other body fluid data.

Substrates

A 50 μL portion of semen was deposited on the chosen substrates and allowed to dry overnight before spectral accumulation. For every semen stain on substrate experiment, two datasets were accumulated: the neat substrate was accumulated with at least 30 points and then the stain spectra were collected. The datasets were then imported into a MATLAB environment for analysis. In this environment, standard deviation and averaging were derived from the datasets. The algorithm for automatic subtraction was also executed in the MATLAB environment.

The datasets acquired from the prepared semen stains on substrates follow a progressive analysis. First, the spectra of substrate and stain were visually compared to gauge the degree of interference. If there is no interference, then identification progresses without subtraction. If there is significant substrate interference, the substrate heterogeneity is evaluated using the standard deviation. If the substrate is homogeneous or moderately heterogeneous, automatic subtraction efforts are undertaken. If there is severe heterogeneity, then PLS modeling will be used to find hotspots to target for manual subtraction.

Body fluid mixtures

Spectral pretreatment was performed in the same manner as for pure fluids i.e. using GRAMS/AI 7.01 and MATLAB 7.4.0 software for statistical analysis.[120] The normalized and corrected spectra were subjected to dimension reduction by PCA. The number of principal components was chosen based on significant factor analysis (SFA) and root-mean-square error of

cross-validation (RMSECV) parameters of leave-one-out cross-validated PCA.[121] PCA scores were used for SVM regression by compiling the experimental data into three groups: two groups were formed using the Raman spectra of the pure body fluids, and one group was formed using only the Raman spectra of the selected body fluid mixtures. Selection rules were established to minimize any interference of the third class with the first two. The resulting three groups of data were used to build the SVM classification model. The full data set, including the Raman spectra of body fluids mixtures omitted in the previous step, revealed a high discrimination power of the developed SVM classification model.

Contaminated body fluids and their mixtures

Cosmic ray interference (for preliminary analysis) was removed with GRAMS/AI 7.01 software from all spectra of contaminated samples and pure contaminants. Extremely noisy spectra were excluded from the data set. Visual inspection of the dataset revealed the presence of complex and variable fluorescent backgrounds. Therefore, an adaptive iteratively reweighted penalized least squares (airPLS) algorithm was used for baseline correction. The spectra were then normalized by total area to account for the varying amount of Raman signal. The prepared spectra were subjected to alternating least squares (ALS) fitting according to the blood signature to estimate the contribution of blood in the experimental spectra. The SSE, R2, and RMSE statistical parameters were calculated for the fitting of pure blood, dust, sand and soil Raman spectra. Those parameters were used to test the quality of fitting and to create a PLS-DA classification approach for the identification of nearly pure blood spots within the contaminated stains. All statistical calculations were performed using the Matlab 7.12.0 environment (Mathworks, Nattick) and PLS toolbox (Eigenvector Research, Inc., Wenatchee).

4. Identification of unknown blood stains using multidimensional Raman signatures

The quality of body fluid identification is based on spectral similarity and can be quantitatively evaluated using sum of squares due to error (SSE), R^2 and root mean square error (RMSE) which indicate how well the spectroscopic signature fit the experimental spectra.

$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual variation, which reflects the fact that real observations do not all lie on the fit profile $\hat{y}$. The coefficient of determination, R^2, was calculated as the square of a Pearson correlation coefficient, often called a *correlation coefficient*, between the original and fitted data values:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2}}\right)^2$$

where $\bar{y} = \frac{1}{n}\sum(y_i)$ and $\overline{\hat{y}} = \frac{1}{n}\sum(\hat{y}_i)$ represent the means of the experimental y_i and the calculated $\hat{y}_i$ Raman intensities, respectively.

An ideal fit would yield calculated profiles that completely match the experimental spectra. The SSE, R^2 and RMSE values represent the total deviation of the fitted data points from the experimental data. The good-fit SSE and RMSE values will be closer to 0 and R^2 will be closer to 1 if the signature corresponds to a body fluid. In this case, the variations in the goodness-of-fit parameters will be caused mainly by noise contribution, and the residuals from the ALS fitting will resemble the normal distribution. The results of fitting for different body fluids are illustrated in Table 2. A quantitative statistical analysis using SSE, R^2, and RMSE was performed to confirm a satisfactory fitting of all experimental spectra.

Table 2. The sum of squares due to error (SSE), R-square (R^2), and root mean squared error (RMSE) calculated for the sweat and vaginal fluid signature fittings. Shaded rows show the results of fitting when the signature is consistent with the type of body fluid stain.

Sample\Signature	SSE	R^2	RMSE
Vaginal fluid\Vaginal fluid	0.298	0.995	0.010
Sweat\Sweat	0.365	0.984	0.018
Saliva\Saliva	0.550	0.979	0.018
Semen\Semen	0.055	0.998	0.0063
Blood\Blood	0.049	0.998	0.0067
Sweat\Blood	6.250	0.925	0.059
Saliva\Semen	5.100	0.905	0.053
Saliva\Vaginal fluid	2.406	0.952	0.044
Semen\Vaginal fluid	5.410	0.924	0.053
Blood\Vaginal fluid	7.396	0.910	0.070

Figure 2 demonstrates several examples of Raman signature fitting. The two upper pairs of spectra show fitting using the correct signature, whereas the three lower pairs are illustrations of “improper” fitting.

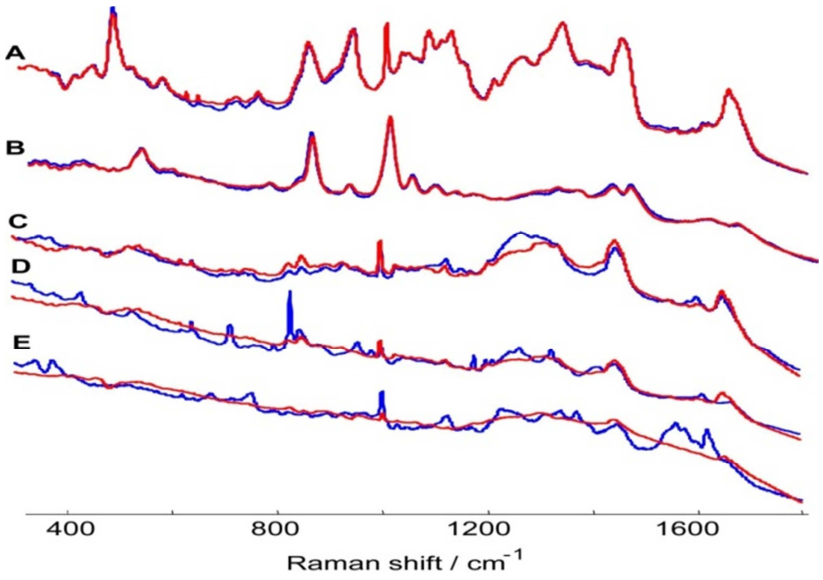

Figure 2. The average Raman spectra (blue lines) of various body fluid samples are fitted by different Raman signatures. The red lines represent the calculated spectra. (A) Raman spectrum of vaginal fluid fitted using the signature of vaginal fluid. (B) Raman spectrum of sweat fitted using the signature of sweat. (C-E) Raman spectra of saliva, semen and blood fitted with the spectroscopic signature of vaginal fluid.

III. Results

Statement of Results

Raman spectroscopic characteristics (spectroscopic signatures) were obtained for blood, semen, vaginal fluid, saliva, and sweat. A statistical program for the automatic identification of body fluids in the dry mixtures was developed. The application of the developed method was extended for various substrates including cotton, denim, glass and tile; as well as various contaminants such as sand, dust, soil and all-purpose bleach. The optimum wavelength range for Raman spectroscopic measurements from the viewpoint of method selectivity, sensitivity, and reduction of fluorescence interference will also be determined in the last half year of the project.

1. Obtain statistically confident Raman spectroscopic characteristics for various body fluids including blood, semen, vaginal fluid, saliva, and sweat (Objective 1)

1.1 Raman spectroscopic signatures of blood, semen, saliva, sweat and vaginal fluid

Blood

The Raman spectrum of blood (Figure 3A) has several contributing components including hemoglobin, albumin, and glucose. Hemoglobin is the major contributor since it contributes to 95% of the dried weight of red blood cells. Reference spectra of hemoglobin[122,123] were used to make the assignments in the blood Raman spectrum. Again, the 1001 cm^{-1} peak is due to the presence of phenylalanine in albumin[124] similar to the phenylalanine-containing proteins, lipase and amylase, in saliva. The amide III vibrational mode, which consists of C-N bond stretching, N-H bending, and C-C stretching,[125] is evident at 1247 cm^{-1} .[126]

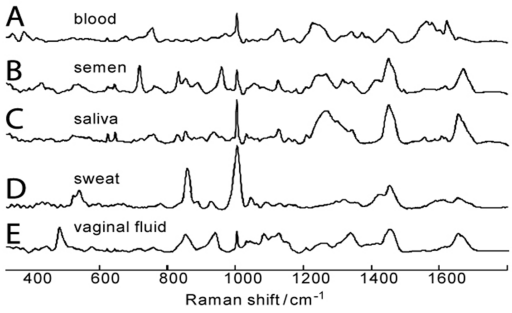

Figure 3. Raman spectra of blood, semen, saliva, sweat and vaginal fluid after fluorescent background subtraction.

Human semen

The human semen spectrum (Figure 3B) is composed of several different components including vibrational signatures of albumin, fructose, lysozyme, lactate, and urea[127]. After comparison to reference spectra of albumin[128] and lysozyme[129], it was determined that these compounds are the major contributors to the spectrum. The peak at 536 cm^{-1} can be tentatively assigned to urea, lactate, and lysozyme. The complex nature of the 536 and 1004 cm^{-1} peaks is indicated by their different relative intensities in a pure urea spectrum (not shown) and the semen spectrum measured under the same conditions. Any slight discrepancies, which were found for the characteristic peak frequencies, between the body fluid spectra and the individual component

spectra are most likely due to the environment created by the dry semen. Assignments for the smaller peaks were not made at this time.

Saliva

Saliva (Figure 3C) contains many different components including amylase, lipase, and lysozyme. Lysozyme[128,129] contribution to the saliva Raman spectrum is evident through the peaks at 1343, 1449, and 1660 cm^{-1}, which match the corresponding peaks in semen and vaginal fluid also assigned to lysozyme. Both amylase[130] and lipase[131] contain phenylalanine in their amino acid sequences, and this amino acid residue results in a strong Raman peak at about 1000 cm^{-1}. The amide I vibrational mode of the polypeptide backbone (mainly C double bond; length as m-dashO stretching vibration) is evident around 1660 cm^{-1}. A similar peak is present in the semen and vaginal fluid spectra. The peak at 2065 cm^{-1} is very unique to saliva and is not present in any of the other fluids. The peak is likely caused by the presence of thiocyanate, which is commonly found in the saliva of a smoker. Thiocyanate is also found in smaller amounts in a non-smoker's saliva as an antibacterial agent.

Sweat

Dried human sweat appeared to be the most heterogeneous fluid in our study. We found that the Raman and fluorescent spectral features of sweat are highly variable. The fluorescent background of human sweat cannot be simplified; because of the variation of a single profile, the multidimensional signature requires two fluorescent and three Raman components (Figure 3D). The fluorescent and Raman components were determined independently by multivariate curve resolution MCR (alternating least squares, ALS) analysis of experimental data split by the AIRPLS algorithm. A quantitative statistical analysis using SSE, R2, and RMSE confirmed a satisfactory fitting of all experimental spectra using the resulting sweat signature.[132] The major constituents of human sweat are water (99%), lactate (lactic acid), urea (uric acid) [133], diethylene glycol, alanine, valine, and leucine. Spectral component 1 was characterized by an 856/1003 cm^{-1} doublet and a strong Raman band at 1003 cm^{-1}, which was assigned to the CN stretching of urea molecules.[134] The 856 cm^{-1} band was tentatively assigned to lactic acid.[135,136] Raman peaks at 552, 800, 1150, 1310, 1445, 1608, and 1651 cm^{-1} that were found in component 2 are indicative of urea, proteins, and lactic acid. Amino acids and protein significantly contribute to component 3 (662, 886, 1062, and 1297 cm-1).[137]

Vaginal fluid

The pure spectrum of vaginal fluid (Figure 3E) has some similarities with the semen spectrum, but there are still substantial differences. The major components of vaginal fluid are urea, lactate, lysozyme, acetic acid, and pyridine[138,139]. Several reference spectra were used in determining the peak assignments[140,141] including those from literature and ones measured in our laboratory using the same instrument. Lysozyme and urea peaks were common for both semen and vaginal fluid, which is in agreement with the composition of these body fluids. The peak around 1000 cm^{-1} likely originates from urea and/or phenylalanine which is present in vaginal fluid proteins. Further quantitative analysis of the correlation between the fluid composition and the Raman spectrum is required to establish the species with dominating contribution to the Raman signal. Again, slight shifts in peak frequencies are observed when compared to the pure component spectra due to the dry vaginal fluid environment.

Raman peaks assignment

Table 3. Peak assignment based on Raman spectra collected from dried stains of pure body fluids. [142-152]

Blood		Semen		Saliva		Sweat		Vaginal fluid	
R	A	R	A	R	A	R	A	R	A
744	trp	641	tyr	323	saccharide	538	urea	481	urea
967	fibrin	715	holin	521	saccharide	553	urea	540	urea
1000	phe	759	alb	544	arg	573	urea	577	urea
1122	heme	798	tyr	632	acetate	588	urea	850-	ama, lac
1248	fibrin	829	tyr	852	ν (C-N)	592	ama, proteins,	855	
1342	trp	848	tyr	919	arg		urea	930-	acetic
1368	heme	888	sph	927	(C-H)	662	-//-	940	acid
1448	trp	958	sph	991	arg	779	-//-	1002	phe,
1542	heme	983	tyr	1002	proteins	830	-//-		urea
1575	fibrin	1003	alb	1047		855-857	-//-	1045	lac,
1620	heme	1011	sph	1097		886	proteins		protein,
		1055	sph	1120		927-937	lac, ama		urea
		1065	sph	1125		1003	ama, proteins,	1082	lac
		1125	sph	1265			urea	1125	lac, urea
		1179	tyr	1295	acetate	1009	urea		
		1200	tyr	1434	acetate	1044	proteins, lac	1250-	protein
		1213	tyr	1444	proteins	1062	proteins	1260	
		1240	proteins	1653	proteins	1089	lac	1310	protein
		1265	tyr	1744	acetate	1105	urea	1330-	protein
		1317	sph			1128	urea, lac	1337	
		1327	tyr			1137	urea	1380	urea
		1336	alb			1280	proteins, urea	1445-	lac, urea
		1448	alb, holin			1297-1360	proteins	1455	
		1461	sph			1410-1420	urea	1610	lac
		1494	sph			1445-1451	lac, urea	1655	urea,
		1616	tyr			1554	urea		lac,
		1668	proteins			1606	lac		proteins
						1651-	lac, urea,		
						1690	proteins		

R – Raman peak, A – assignment, trp – tryptophan, phe – phenylalanine, alb – albumin, tyr – tyrosine, sph - spermine phosphate hexahydrate, arg- arginine, lac – lactate, ama – amino acids

1.2 Multidimensional Raman spectroscopic signatures of blood, semen, saliva, sweat and vaginal fluid

The heterogeneous nature of body fluid stains does not allow the use of a single spectrum of pure stains as a reference. A set of characteristic spectra is required to represent all of the possible variations among both a single donor's stains and different donor's. To achieve this goal, advanced statistical analysis was performed. The necessary and sufficient numbers of spectral components for each body fluid were determined using cross-validated PCA and significant factor analysis (SFA). Spectral components (Figure 4) were found using multivariate curve resolution (MCR) by an alternating least squares algorithm (ALS).

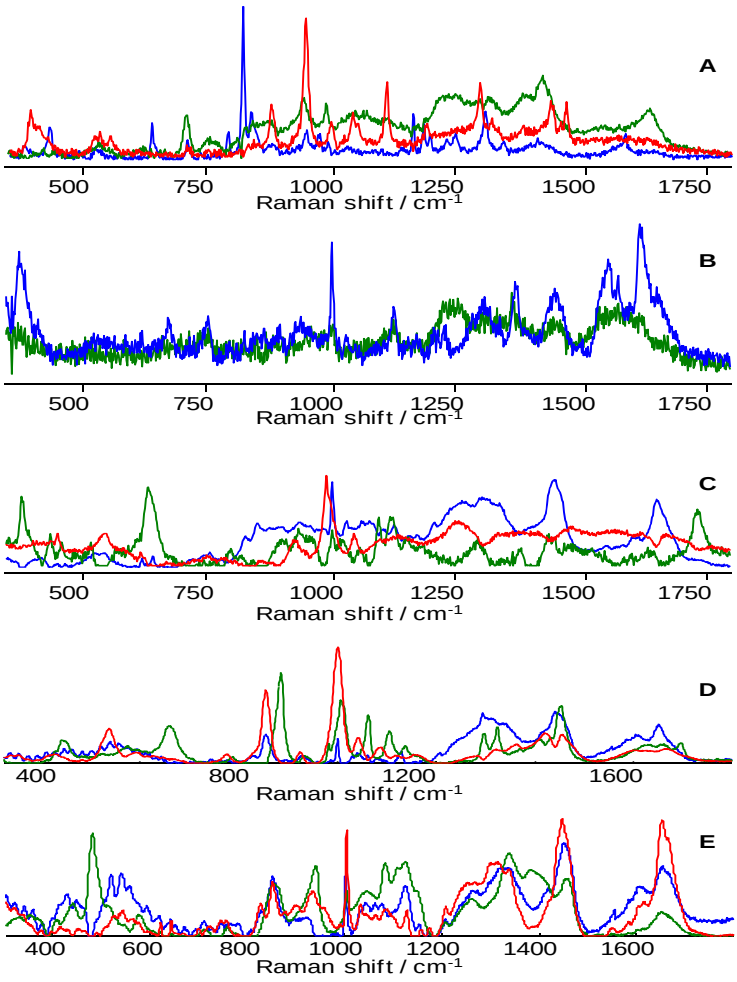

Figure 4. Multidimensional spectroscopic signatures of semen (A), blood (B), saliva (C), sweat (D) and vaginal fluid (E).

The Raman spectrum of blood is dominated by hemoglobin (Component 1) (1000, 1368, 1542 and 1620 cm^{-1}) and fibrin (Component 2) (967, 1248 and 1342 cm^{-1}) contributions, which are the main constituents of dried red blood cells and plasma respectively (Figure 4B). Noticeably, a Raman spectrum of liquid blood significantly differs from Raman spectra of dried stains. These differences can be explained by the coagulation process and chemical derivatization of hemoglobin that occurs when blood dries. The multidimensional signature of semen contains three Raman spectral components that were obtained by statistical analysis and one fluorescent background extracted from the average spectrum (Figure 4A). The Raman components of semen are complex and cannot be assigned to pure biochemical species. Spectral features of component 1 are consistent with the characteristic peaks of tyrosine, one of the most abundant free amino acids in semen. The pure spectrum of semen is composed of characteristic Raman bands at 1668 cm^{-1} (Amide I) and 1240 cm^{-1} (Amide III) associated with proteins (Component 2). Serum albumin (759, 1003, 1336 and 1448 cm^{-1}), choline (715 cm^{-1}) and spermine phosphate hexahydrate (888, 958, 1011, 1055, 1125, 1317, 1461 and 1494 cm^{-1}) also have strong contribution in the Raman spectra of semen stains. The presence of Raman bands at 641, 798, 829, 848, 983, 1179, 1200, 1213, 1265, 1327 and 1616 cm^{-1} originated from other biochemical species reflects the complex composition of semen. Three spectral components and one fluorescent background extracted from an average spectrum characterize the spectroscopic signature of saliva (Figure 4C). Proteins (1002, 1444 and 1653 cm^{-1}) - component 1, acetates (632, 1295, 1434 and 1744 cm^{-1}), and carbohydrates (323 and 521 cm^{-1}) - component 2, are major contributors of the Raman spectra of saliva. Raman spectra of both sweat and vaginal fluid are dominated by lactate/lactic acid and urea/uric acid. We found that the Raman and fluorescent spectral features of sweat are highly variable. The multidimensional signature requires two fluorescent and three Raman components. The fluorescent and Raman components were determined independently by multi-curve resolution MCR (ALS) analysis of experimental data split by the AIRPLS algorithm. Raman spectra of sweat can be characterized by a higher relative contribution of urea/uric acid - component 1, a strong Raman band at 1003 cm^{-1} was assigned to the CN stretching of urea molecules and moderated the contribution of single amino acids and proteins (Figure 4D). The 856 cm^{-1} band was tentatively assigned to lactic acid. Raman peaks at 552, 800, 1150, 1310, 1445, 1608, and 1651 cm^{-1} that were found in component 2 are indicative of urea, proteins and lactic acid. Amino acids and protein significantly contribute to component 3

(662, 886, 1062 and 1297 cm^{-1}). Vaginal fluid has higher protein content. Raman components 1 and 2 have similar spectral profiles and consist of the same Raman bands with different relative intensities (Figure 4E). These spectral components contain the characteristic features of lactic acid, urea and proteins. Raman component 3 contains unique bands at 481 and 1380 cm^{-1}, which correspond to the N-C=O bending and CN stretching vibrations of urea respectively. It was found that the fluorescent background in the case of vaginal fluid is more variable compared to sweat. A linear combination of three spectra is required to compensate for this variation. However, the variations in the Raman portion of the vaginal fluid spectra between donors are much smaller than those in sweat.

2. Develop a statistical program for automatic identification of body fluids in dry mixtures. Evaluate major performance characteristics and limitations of the developed method as presumptive and confirmatory tests (Objective 2)

2.1 Automatic identification of pure body fluids

Software for the discrimination of traces of individual body fluids with high level of confidence was developed. The program analyzes a set of recorded spectra and compares them with the library of multidimensional spectroscopic signatures built for various body fluids (all five body fluids). The multidimensional nature of the spectroscopic signatures accounts for intrinsic heterogeneity of dry samples and possible variations with donors.

Identification of an unknown species

Discriminant analysis (PLS-DA, SVM-DA) allowed for differentiating semen, blood, saliva, vaginal fluid and sweat traces with over 99% confidence under laboratory conditions. The ability to make this identification on-site at a crime scene would be a major advancement in forensic analysis (Figure 5).

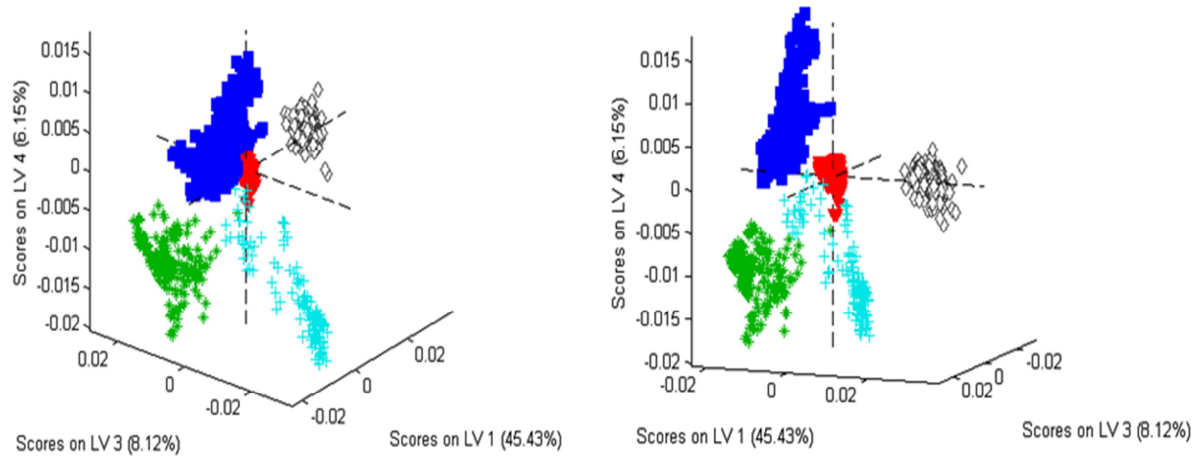

Figure 5. A three-dimensional latent variable plot for blood, semen, saliva samples based on the first, third and fourth latent variables (LV). Defined classes: blood (red triangles) - class 1, saliva (green stars) – class 2, semen (blue squares) – class 3, sweat (light blue crosses) – class 4, vaginal fluid (white diamonds) – class 5.

Automatic identification of body fluids

Data preparation, preprocessing and automatic confirmatory analysis were performed as described in the first semi-annual report (Figure 6):

✓ Data preparation includes defining the spectral range of raw data followed by interpolation on the 500-1800 cm^{-1} spectral region with predefined spectral resolution, "cosmic" lines depressing;

✓ Preprocessing includes automatic baseline subtraction and predefined preprocessing procedure;

✓ Body fluid stain identification is based on the results of automatic discriminant analysis.

Three-dimensional score plots built for the PLS-DA model demonstrated clustering among single body fluid samples indicating that more specific information about donors is accessible.

The major conclusion of this part of the project is that if (i) a minute spot (as small as a femtoliter or picogram) of the dry stain is dominated by a pure individual body fluid and (ii) the

Raman spectrum is acquired from this spot via automatic scanning, then identification of that part of the stain will be accomplished via the automatic data processing developed here.

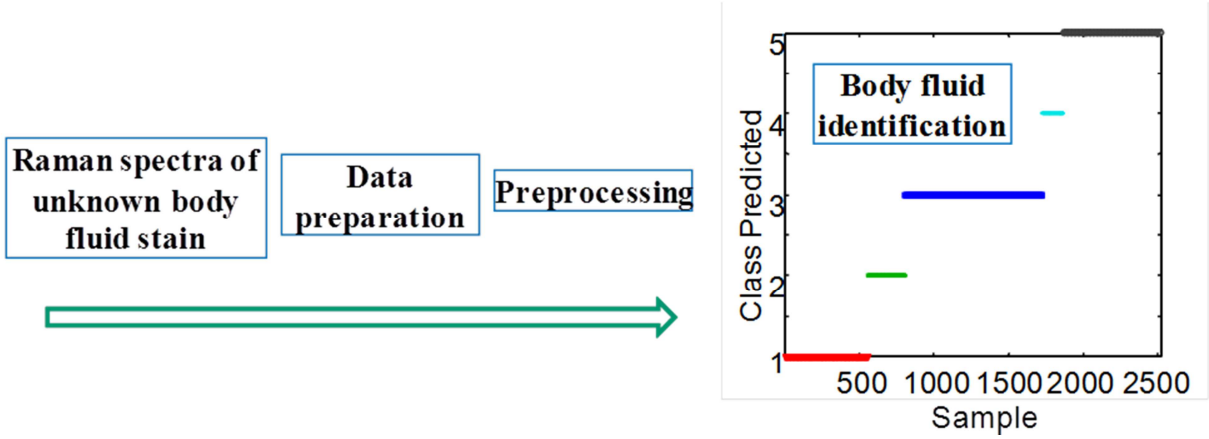

Figure 6. Scheme of the of automatic body fluids identification process. Defined classes: blood (red triangles) - class 1, saliva (green stars) – class 2, semen (blue squares) – class 3, sweat (light blue crosses) – class 4, vaginal fluid (white diamonds) – class 5.

2.2 Analysis of body fluid mixtures

Stains of body fluid mixtures are intrinsically heterogeneous. A single mixture stain can have spots with variable composition ranging from pure semen to pure blood. Heterogeneity of the samples along with sample-to-sample variability is a major challenge of spectroscopic characterization of any biomaterials. Despite recent progress in the field of chemometrical analysis, the problem of nonlinear effects present in the relationships between the concentration of the mixture component and the mixture spectra still does not have a universal solution. We developed a multistep chemometrical procedure to facilitate the detection of pure body fluids and their mixtures. Our approach is based on effective combination of SVM regression and classification methods, which overcomes the calibration problems and produces robust models in the case of spectral variations due to a nonlinear interference. Regression analysis assists with the selection of Raman spectra of binary mixtures that can be easily distinguished from spectra of pure body fluids and have distinct characteristics of both ingredients. Discriminant analysis based on the selected classes significantly outperforms the direct application of the regression method.

Figure 7 shows selected characteristic preprocessed Raman spectra acquired from pure blood and semen along with the spectra of blood/semen mixtures. Distinctive Raman bands of blood (754, 1003, 1226, and 1619 cm^{-1}) and semen (716, 830, 959, 1268, 1329, and 1671 cm^{-1}) can be used to trace their contribution in mixtures. While spots dominated by a single body fluid can be easily detected and identified using multidimensional Raman signatures1-3, analysis of thoroughly mixed fluids requires a more complex approach. Raman spectra of blood-semen mixtures are dominated by blood contribution. For example, according to visual inspection, all spectra with more than 50% of blood do not have noticeable evidences of semen contribution.

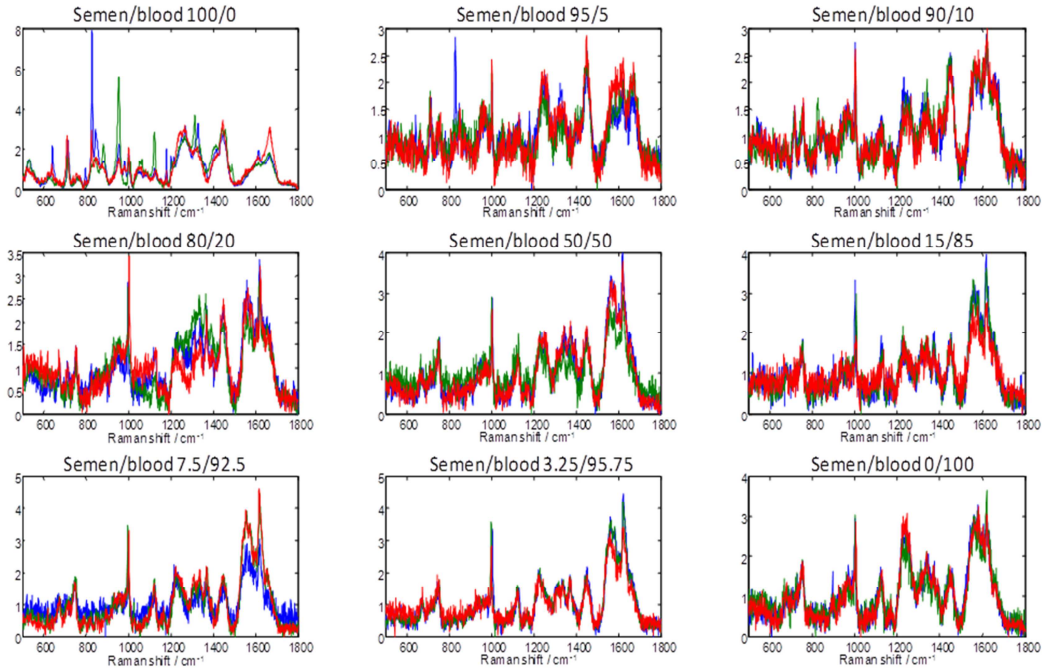

Figure 7. Selected characteristic of preprocessed Raman spectra acquired from pure blood and semen along with the spectra of blood-semen mixtures. Colored lines represent the variability of the Raman spectra due to the samples heterogeneity.

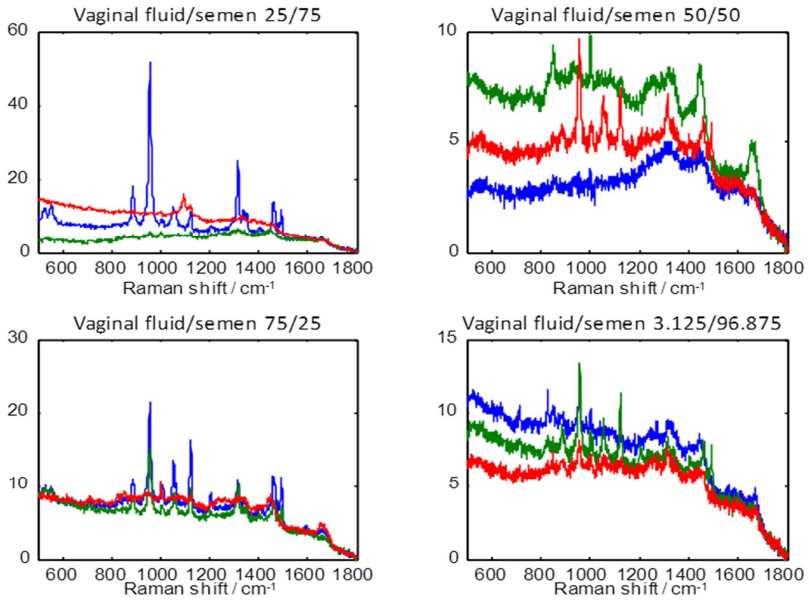

Figure 8. Selected characteristic of raw Raman spectra acquired from vaginal fluid-semen mixtures. Colored lines represent the variability of the Raman spectra due to the samples heterogeneity.

Dry stains of vaginal fluid are more heterogeneous when compared to semen or blood. Figure 8 shows selected spectral characteristics in raw Raman spectra acquired from vaginal fluid/semen mixtures. Characteristic Raman bands of vaginal fluid (887, 957, and 1316 cm^{-1}) can be easily used to detect low concentrations of vaginal fluid. Detection of semen traces among mixed samples at low concentration can be achieved using the developed statistical method.

A multistep chemometrical procedure for the forensic investigations of dried stains of body fluid mixtures was developed using experimental data selected by SVM regression, which is optimized for the detection of pure body fluids and their mixtures. Discriminant analysis based on selected classes clearly demonstrates a great potential of Raman spectroscopy coupled with advanced statistical analysis for the nondestructive characterization of body fluid mixtures.

Figure 9 shows the results of calibration using leave-one-out cross-validation. The rough plot of the predicted concentration versus spectra reflects the level of spectral data variability after the performed filtration. The high variability of the Raman spectra reflects inherent heterogeneity of samples. SVM regression demonstrates that mixtures with blood content higher than 80% cannot be distinguished from pure blood in the first step of data treatment. Conversely,

even 5% of blood resulted in Raman spectra is distinguishable from Raman spectra of pure semen. So, the lowest detectable by SVM regression model levels of blood and semen are around 5 and 25 %, respectively. These results were used to group all experimental data in to three groups. The first two group were formed using Raman spectra of pure body fluids, while the third group was built using all spectra that had fallen within the 15-75% range. Boundaries were chosen to cut off 100% of pure fluids spectra. Raman spectra of mixtures outside of 15-75% were not included in any group.

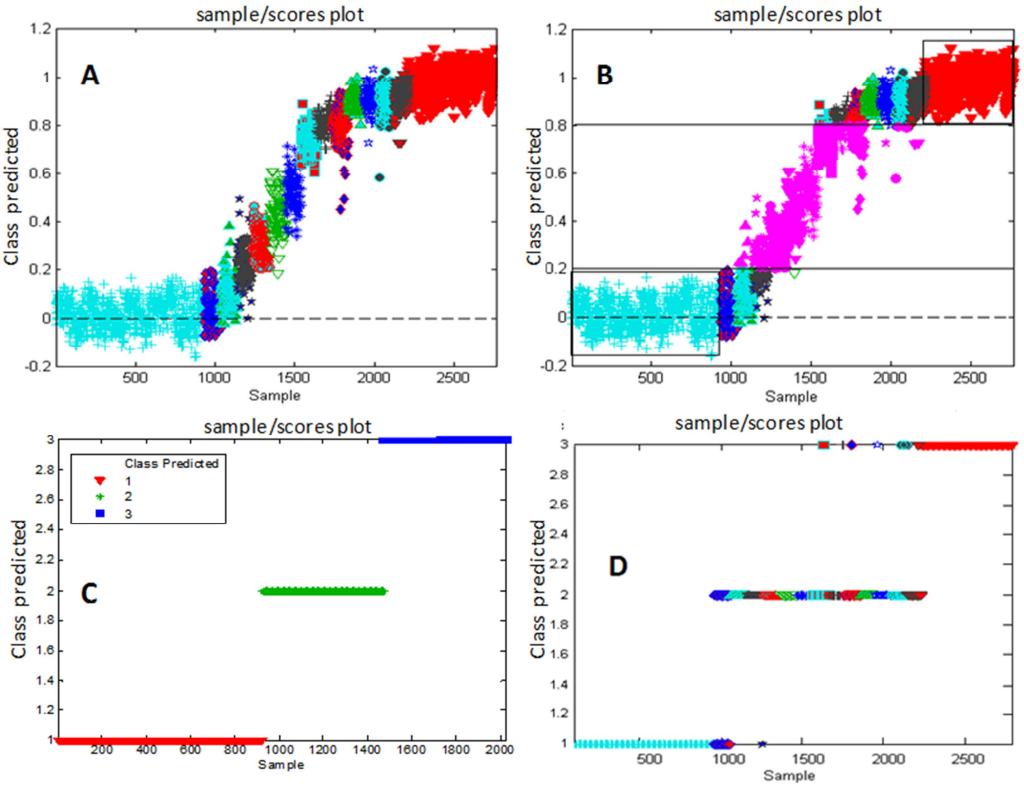

Figure 9. A. SVM regression, B. SVM selection. C. SVM classification model based on the selected spectra D. SVM discriminant analysis of all experimental data.

The formed groups were assigned to classes 1 (semen), 2 (blood-semen mixtures with various compositions) and 3 (blood) and subjected to SVM discriminant analysis (Figure 9.C). The PCA method was used to reduce dimensionality and data block compression. Cross-validation was performed by 1000 splits of entire data set into training and test data sets. Each splitting was done by random selection of 75 and 25% of data for training and test data sets, respectively. Figure 9.C demonstrates that 100% of classified Raman spectra were identified

correctly. For comparison, a complete data set was subjected to analysis using the calculated SVM-DA model (Figure 9.D). It is evident that most of the mixtures' data were assigned to class 2 (mixtures), and only for mixture with 5% of blood significant part of spectra were assigned to pure semen. These results considerably outperform both direct classifications, when recorded concentrations are used as classifiers, and composition determination using SVM regression method itself. Real samples are usually not thoroughly mixed and scanning of the entire stain could reveal very small droplets of the intruded fluid pointing out spots for further DNA analysis. Taking in to account that those droplets are detectable even if the concentration of the minor contributor is small (5%), the overall effectiveness of Raman spectroscopy is very good.

3. Extend the application of the developed method to body fluid stains on various substrates including human skin, fabrics, carpets, hard surfaces, etc. (Objective 3)

Spectroscopic contribution of substrates in Raman spectra acquired from body fluid stains can significantly impair their identification. In this study we addressed (1) the problem of strong fluorescence originated from substrates, which may significantly depress the Raman signal of body fluid, (2) subtraction of spectroscopic contribution of heterogeneous substrates, and (3) identification of body fluids deposited onto substrates with various surface morphologies. The following substrates were selected: glass (fluorescent, flat and homogeneous surface), cotton (slightly fluorescent, rough, porous and relatively homogeneous surface from spectroscopic point of view), denim (fluorescence, rough, porous and relatively homogeneous from spectroscopic point of view surface), and glazed bathroom tile (intensely fluorescent, flat and highly homogeneous surface).

Blood was used first to investigate the substrate problem because it is a strong scatterer and pigmented, which prevents light from penetrating to the substrate. Glass, cotton and denim substrates are relatively homogeneous. These considerations were chosen to prove that, in principal, spectral subtraction works to eliminate substrate interference.

3.1 Blood reference spectra

A reference spectrum of blood on aluminum was obtained for each excitation wavelength surveyed (Figure 10). Spectra collected at 406 nm and 457 nm (Fig. 1 A, B) excitations have the most similar profiles; they are almost indistinguishable in terms of the number of peaks, their locations and relative intensities, and both have a similar flat baseline profile.

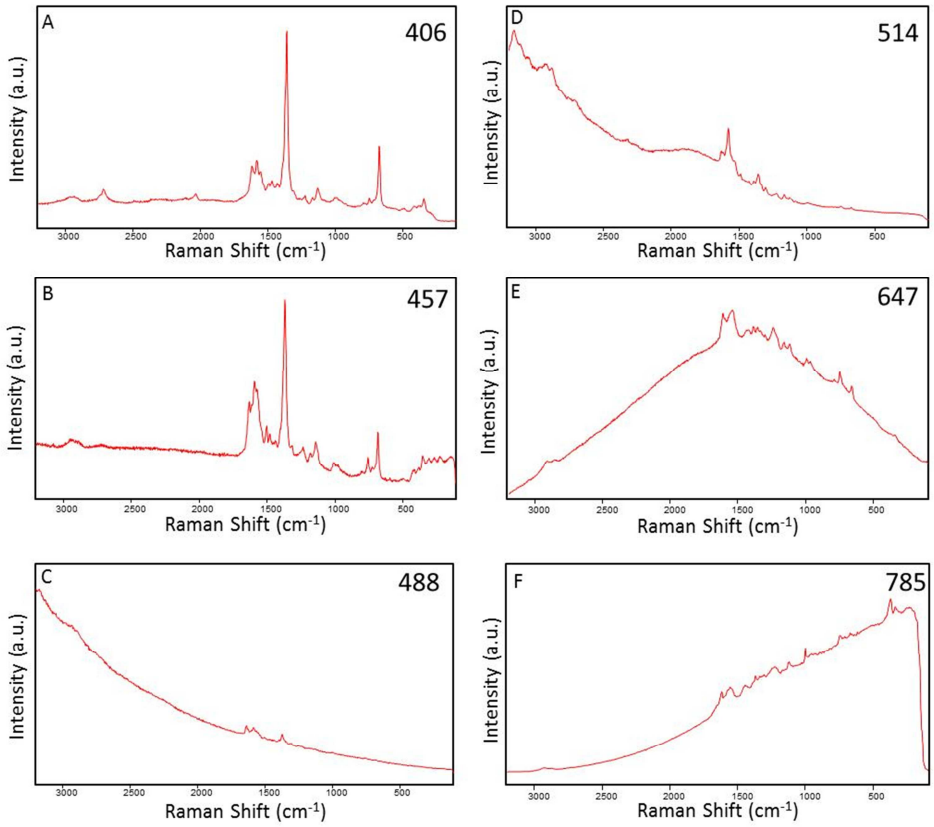

Figure 10. Raman spectra of pure blood acquired using 406 (A), 457 (B), 488 (C), 514 (D), 647 (E) and 785 (F) nm excitations.

The 647 nm blood spectrum displayed the greatest similarity to the 785 nm spectrum (Fig. 1 E, F), although there are some easily distinguishable features, most notably the background profile. The baseline blood spectrum at 647 nm excitation is similar to an inverted parabola, while the 785 nm spectrum has a strong upward trend toward lower frequencies. Both spectra have similar band shapes within the amide II (1520-1580 cm^{-1}), phenylalanine (1180-1220 and 1600-1620 cm^{-1}), amide III fibrin (1220-1300 cm^{-1}), and heme (1105-1135 cm^{-1}) spectral regions. The same strong upward trend at the greater frequencies is observed in the 488 and 514 nm blood spectra. All six reference spectra have several peaks in the 1652-1511 cm^{-1} (amide) spectral region, although they vary in position and intensity. In addition, all reference spectra display a peak at approximately 1364 cm^{-1}, which has been assigned to a CH_3 stretch of heme[2]; it is the most intense feature in the 406 and 457 nm reference spectrum.

3.2 Blood on glass

The reference Raman spectrum of glass for each instance displays strong luminescence regardless of the excitation (Figure 11, black lines); the profile varies for each excitation wavelength, but most display a broad peak centered at 1095 cm^{-1}.

Despite strong luminescence, the raw spectra of the blood stain on glass matched the blood reference in all but one excitation wavelength (Figure 11). At 785 nm excitation, the features of the spectrum were dominated by glass (Figure 11, F). The only feature of blood that was visible was the sharp Phe peak (1003 cm^{-1}) and the peak at 375 cm^{-1}. To extract the contribution from blood, spectral subtraction was applied using a reference spectrum for glass. The signal from glass was calculated to be approximately 23x greater than the signal from blood. Glass was observed to be very homogeneous, making subtraction in nearly all instances artifact free. After subtraction, all features were visible and matched both a reference spectrum for blood (Figure 11, F green trace) and the developed spectroscopic signature.

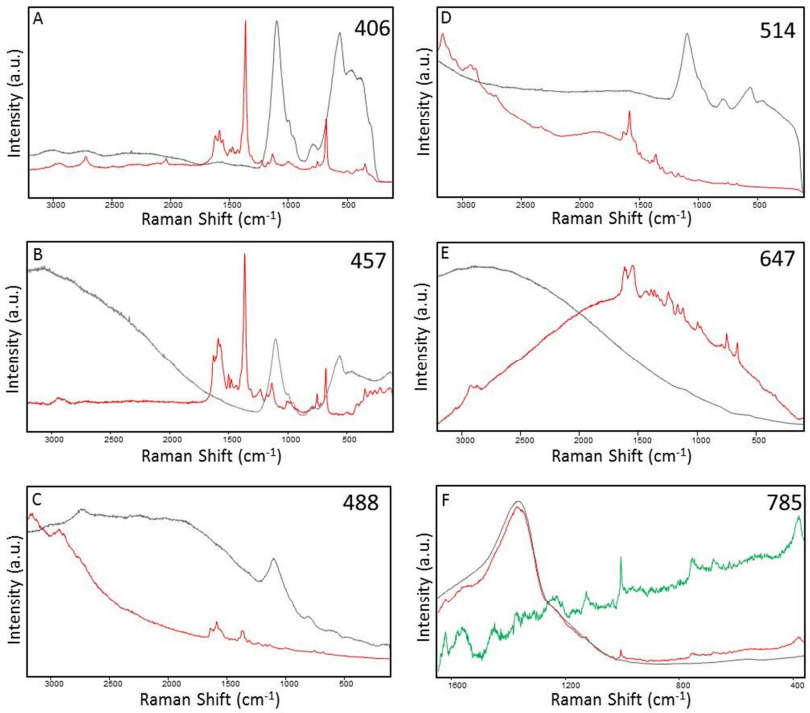

Figure 11. Raman spectrum of a blood-stain on glass acquired using 406 (A), 457 (B), 488 (C), 514 (D), 647 (E) and 785 (F) nm excitations. For comparison, the red traces are the spectra for the blood reference. In F, the green trace is the result of manual substrate subtraction.

It is worth noting that a high magnification objective helps to a limited degree to reduce substrate interference with glass. With a 50x objective, the substrate contribution still dominated the spectrum of a blood-stain on glass and tile, although this spectrum was more favorable than that obtained with the 20x objective (data not shown). Using high magnification could be helpful to produce a better quality spectrum when the fluid is located on top of the substrate. However, high magnification is not preferable for mapping experiments because of the narrow focal point.

3.3 Blood on tile

Similar to glass, tile is a highly luminescent substrate (Figure 12). The reference spectra for this substrate contained mostly broad features, with a few sharp peaks with for some excitation wavelengths. The Raman spectra of blood-stains were completely dominated by blood features, except at 785 nm excitation (Figure 12.F). At this excitation, the spectrum of the stain is indistinguishable from the substrate reference. However, after spectral subtraction, the contribution of blood is evident (Figure 12.F, red line); as in the case of glass, all features of blood are observable after subtraction. However, a sharp negative feature is observed at 1364 cm^{-1} (Figure 12.F, shaded area). This subtraction artifact corresponds to the most prominent feature of the tile's spectrum and is a result of substrate heterogeneity.

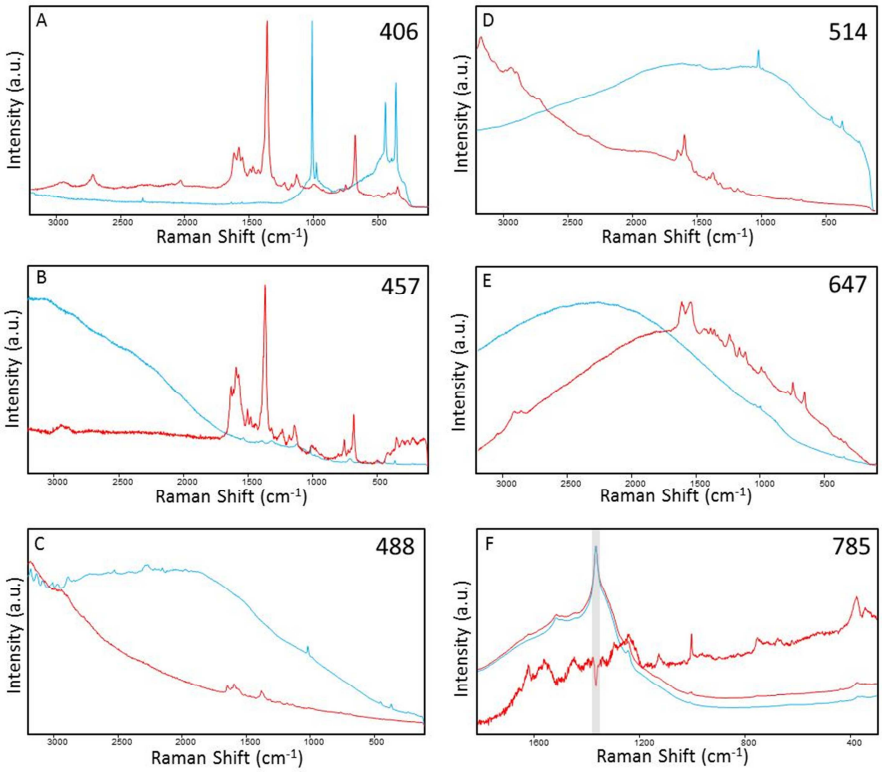

Figure 12. Raman spectrum a blood-stain on tile acquired using 406 (A), 457 (B), 488 (C), 514 (D), 647 (E) and 785 (F) nm excitations. For comparison, the red traces are the spectra for the blood reference. In F, the red trace is the result of manual substrate subtraction.

3.4 Blood on cotton

The fluorescence of cotton varied greatly depending on the excitation wavelength (Figure 13). At 406 nm excitation (near UV), the fluorescence of cotton is extreme. The laser power needed to be reduced below 1% to obtain a spectrum. The fluorescence was also strong with 457, 488 and 514 nm excitations. This result is consistent with that of Boyd et al., who found cotton to be highly luminescent at 532 nm excitation[8]. Because of strong the fluorescence, the spectrum for cotton contained few Raman peaks. At 647 and 785 nm, the fluorescence contribution is comparatively weak and the Raman peaks of cotton are more pronounced. Cotton displays a characteristic doublet Raman band centered at approximately 1108 cm^{-1}.

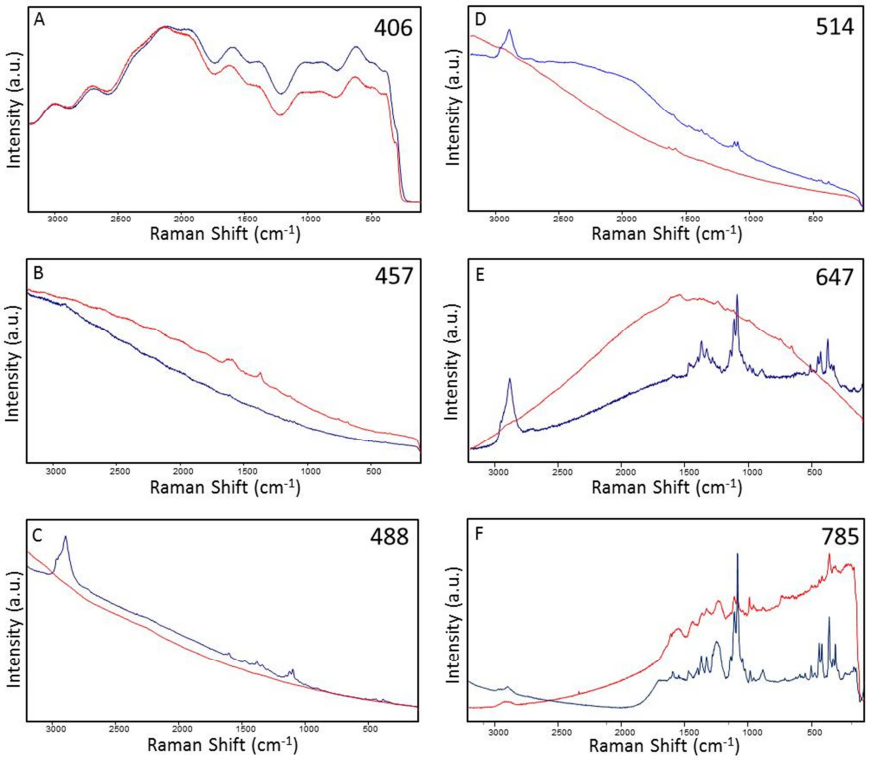

Figure 13. Raman spectrum of a blood-stain on cotton acquired using 406 (A), 457 (B), 488 (C), 514 (D), 647 (E) and 785 (F) nm excitations. For comparison, the red traces are the spectra for the blood reference.

With a laser excitation for which the fluorescence of cotton is high, the blood stain spectrum is essentially featureless as the substrate contributions overwhelm the spectrum. However, at 785 nm, blood features dominate the spectrum. The small cotton contribution, which can be identified by the doublet at 1124 and 1095 cm^{-1}, can be easily subtracted using a reference spectrum. Cotton yields a relatively consistent signal, which makes subtraction nearly artifact free, even though blood and cotton have several overlapping peaks throughout the spectrum.

These results indicate that there is an important distinction between detecting a fluid that is located on top of a substrate and a fluid that is incorporated into a substrate. The absorption properties of blood allow light at 785 nm to penetrate through the dried trace and cause substrate interference. However, blood has a relatively high absorption for light shorter than 650 nm which would block light from penetrating to the surface below, thereby reducing interference. Another

benefit of using a shorter wavelength for excitation is that scattering efficiency is enhanced by an approximate factor of $\sigma R \sim 1/\lambda^4$, thereby enhancing sensitivity.

The absorption properties of blood are not considered to be advantageous in the case of cotton. This is probably because the fluid is not located on top of the substrate but is rather incorporated into it. For this type of blood-stain, near-infrared light performed the best because the interference was in the form of fluorescence from cotton.

These results demonstrate that careful wavelength selection can circumvent the problem of substrate interference altogether and it is a consideration that must be evaluated from the perspective of both the nature of the substrate and the fluid in question. However, even with an unfavorable excitation, the contribution of blood can be extracted by spectral subtraction. Furthermore, with respect to the development of a Raman spectroscopic approach to identifying body fluids, there seems to be a significant difference between detecting a fluid that is located on top of a substrate versus one that is incorporated into a fabric.

3.5 Blood on cotton and blood recovery

Because cotton is a very common fabric and can be used to collect blood evidence, it merited further experimentation to find the limit of detection. A 785 nm laser source was selected for these experiments because it performed the best in detecting blood on cotton (section 3.1.4). First, a series of dilute blood samples was prepared with concentrations (water:blood) of 2:1, 4:1, 6:1, 10:1, 20:1, 40:1, and 60:1. A 100 μL aliquot of diluted blood was placed on a cotton swatch and allowed to dry before analysis.

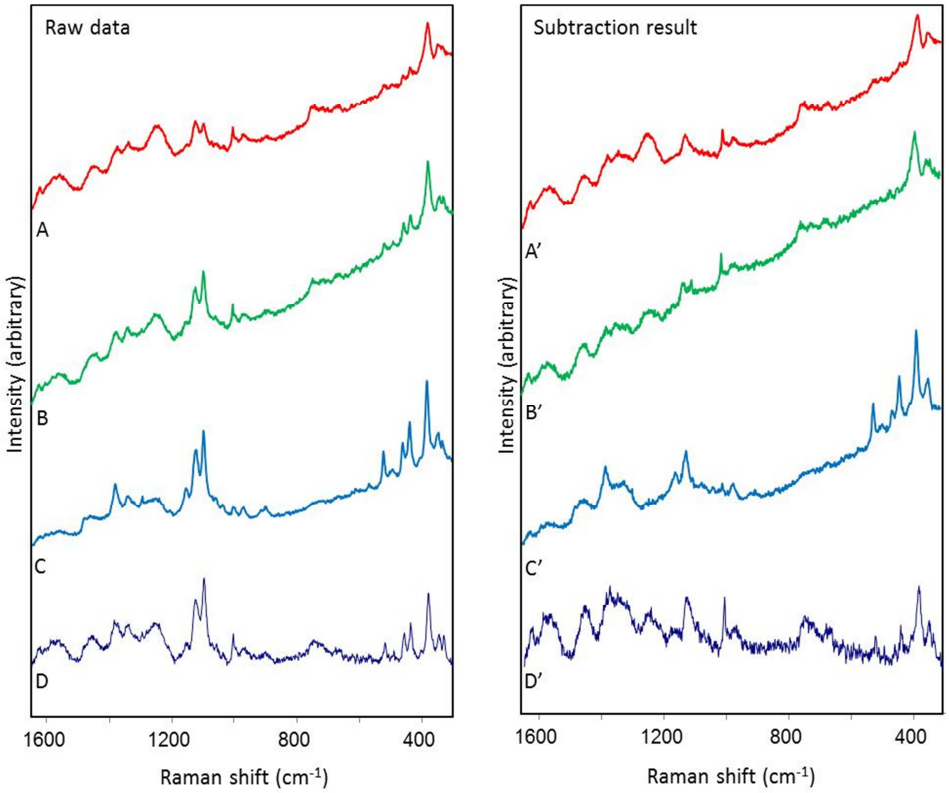

Figure 14. Raman spectra for selected dilute blood and blood recovery experiments. The left panel is the raw spectrum, and the right panel is the result after subtraction of a reference spectrum for cotton. A-A' is 2:1, B-B' is 40:1, C-C' is 60:1, and D-D' is blood recovered from a blood stain on glass. (D and D' spectra were manually baselined for representation).

As displayed in Figure 14, even with considerably diluted blood, the raw spectrum was dominated by blood features. As expected, the blood contribution decreased with dilution. We found that with high levels of dilution, the cotton contribution could be cleanly subtracted. With dilute blood, the blood features are detectable up to a limit of 40:1. The raw and subtracted spectrum of the 60:1 dilute blood stain contained no apparent features of blood.

Because dilute blood on cotton was readably detectable, the detection of recovered blood on a cotton swab should be possible which is important because cotton swabbing is a common evidence collection procedure. This recovery method could feasibly be used to identify blood that cannot be analyzed at the scene. To recover blood, a cotton swab was moistened with water and used to collect blood from stains on tile and glass.

The spectrum of the recovered blood-stain was observed to have contributions from both cotton and blood. As before, the blood spectrum could be obtained by subtraction of the substrates' contribution. Therefore, identifying blood from a swabbing of evidence is possible. We believe this approach is an appropriate analysis method for blood evidence located on top of a substrate that cannot be otherwise collected or immediately analyzed. Therefore, this blood recovery method should not be used for blood-stains discovered on fabric, as this evidence is easily collectable. Blood evidence on fabric should be collected in their entirety and preferably analyzed *in situ*. Further work with blood evidence on dyed fabrics and carpets is ongoing in our laboratory.

3.6 Blood on denim

The identification of a blood-stain on a dyed fabric was attempted using denim as a substrate. Denim is an extremely common dyed cotton fabric and as such, it is reasonable to expect this type of evidence at a crime scene. A 20 μL aliquot of pure blood was placed on blue jeans and allowed to dry. Spectra were obtained from the stain and the neat substrate.

The reference spectrum for denim contained several Raman peaks on top of a strong fluorescent background (Figure 15, B). This spectrum is consistent with the spectrum of indigo dye that is common for this fabric. The most prominent peaks were located at approximately 1573 cm^{-1}, which is assigned to conjugated functional groups in the indigo molecule[11]. No cotton contribution (Figure 15, C) was detected in this spectrum.

Denim was observed to contribute a large fluorescent background in the spectrum of the blood-stain due to the presence of dye (Figure 15, A). The fluorescence was so strong that no Raman peaks from blood or substrate are easily discernible, thus making spectral subtraction very difficult. To solve this problem, a baseline algorithm was applied to the blank and blood-stain spectra before subtraction (Figure 15, E). The Vancouver Raman Algorithm subtracts a fifth-order polynomial, which is a common solution for removing broad fluorescence contributions from a Raman spectrum[10]. After baseline treatment, the subtraction was performed with the baseline-treated reference spectrum. The resulting spectrum (Figure 15, F) matched several peaks from the signature of blood and revealed no apparent residual contribution from denim, thereby demonstrating that even in dyed fabric substrates, the contribution of blood can be extracted.

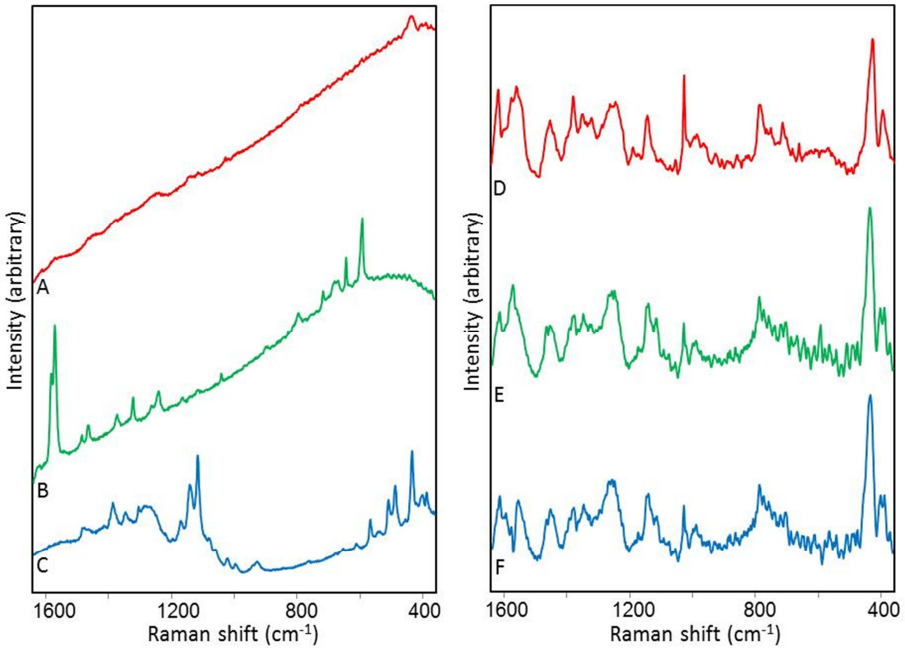

Figure 15. Raman spectra for blood stain on denim experiments. A) Raw blood-stain on denim, B) Blank denim spectrum C) Blank cotton spectrum D) Pure blood reference E) Baselined blood stain spectrum F) Subtraction result after eliminating the denim contribution.

3.7 Statistical analysis

The Raman spectra collected with different excitations from samples of blood stains on different substrates were preprocessed for statistical analysis. Baseline contributions were subtracted with the help of the adaptive iteratively reweighted Penalized Least Squares (airPLS) [ref] and Weighted Least Squares (WLS) [ref] algorithms. The substrate contribution was manually subtracted, where noted, before statistical analysis. The resulting spectra were compared to the reference spectra from the pure blood on aluminum foil collected with the corresponding excitation. Pearson's correlation coefficients (R^2) were used as a quantitative measure of similarity between the spectra. Applied here, it is a measure of the degree that a blood stain spectrum matches a reference blood spectrum. A value of 1 indicates that a spectrum strongly resembles the reference spectrum for blood, and a value of 0 indicates that it does not. Table 4 summarizes the results of this statistical treatment.

Considering all the wavelength excitations, the best results were obtained from glass and tile. In these cases, the correlation coefficients were close to 1 for all instances, demonstrating

that substrate interference can be avoided by the selection of excitation wavelength or by subtraction of a reference spectrum. The 785 nm spectrum of tile scored reasonably high (0.9566), despite the subtraction artifact caused by the substrate heterogeneity.

The lower R^2 values for cotton are a result of substrate interference. The exception to this trend was the spectrum collected at 785 nm for which the substrate contribution was subtracted; this spectrum scored very high (0.9952).

Table 4. Pearson's correlation coefficients calculated to compare Raman spectra collected from blood-stains on different substrates after substrate spectroscopic contribution with the corresponding reference spectra.

Substrate	Excitation wavelength	Pearson's correlation coefficient	Selected region	Baseline correction algorithm/ input parameters
Glass	406	0.9991	100-3200 cm^{-1}	airPLS, 300
	457	0.9929	100-3200 cm^{-1}	airPLS, 300
	488	0.9384	820-2001 cm^{-1}	airPLS, 100
	514	0.9926	850-2000 cm^{-1}	airPLS, 30
	647	0.9822	600-1700 cm^{-1}	airPLS, 300
	785*	0.9566	800-1800 cm^{-1}	airPLS, 300
Tile	406	0.9989	100-3200 cm^{-1}	airPLS, 300
	457	0.9968	100-3200 cm^{-1}	airPLS, 100
	488	0.9593	820-2001 cm^{-1}	airPLS, 100
	514	0.9903	850-2000 cm^{-1}	airPLS, 30
	647	0.9832	600-1700 cm^{-1}	airPLS, 300
	785*	0.9760	900-1800 cm^{-1}	WLS
Cotton	406	0.0627	100-3200 cm^{-1}	airPLS, 300
	457	0.7309	100-3200 cm^{-1}	airPLS, 100
	488	0.1237	820-2001 cm^{-1}	airPLS, 100
	514	0.5484	850-2000 cm^{-1}	airPLS, 30
	647	0.9466	600-1700 cm^{-1}	airPLS, 300
	785*	0.9952	850-1800 cm^{-1}	VRA

*Indicates that the substrate contribution was subtracted before the spectrum was compared to the blood reference.

An additional complication in blood identification is the intrinsic heterogeneity of human blood. Our previous studies demonstrated donor variations in the Raman signal of blood in addition to variations in different spots of a particular blood-stain (spatial heterogeneity). Therefore, a single reference spectrum cannot completely characterize a pure blood-stain. For this reason, a multidimensional fitting using iterative component fitting is a better identification method for body fluids. This approach has been developed and applied extensively in our lab for Raman spectra acquired using 785 nm excitation.

Here, we used the recently developed multidimensional signature of blood to fit the collected Raman spectra of blood (785 nm excitation) after the subtraction of substrate

contributions (Figure 16). As determined by visual inspection, the reconstructed by fitting procedure and the analyzed spectra overlap well for blood-stain spectra obtained from glass and cotton. As was expected from the direct comparison with a single reference spectrum (Table 4), Raman spectra obtained from blood on tile were not well reproduced by this fitting procedure. The SSE, R^2, and RMSE statistical parameters were used as quantitative tests of the fitting (Table 5). A coefficient of determination (R^2) of 1.0 indicates that the fit explains the data perfectly. RMSE is a measure of the differences between the values predicted by the fit and the values actually observed. RMSE and SSE values close to 0 indicate a good quality of fitting. For comparison, these statistical parameters were calculated for the Raman spectrum of blood on aluminum foil. Because aluminum foil does not contribute to the Raman spectra, deviations from the ideal SSE, R^2, and RMSE values are only produced by noise contributions.

From visual inspection, it was expected that the blood stain spectra from glass and cotton would fit to the multidimensional signature better than the blood stain on tile (due to the subtraction artifact). The scores in Table 5 reflect this effect.

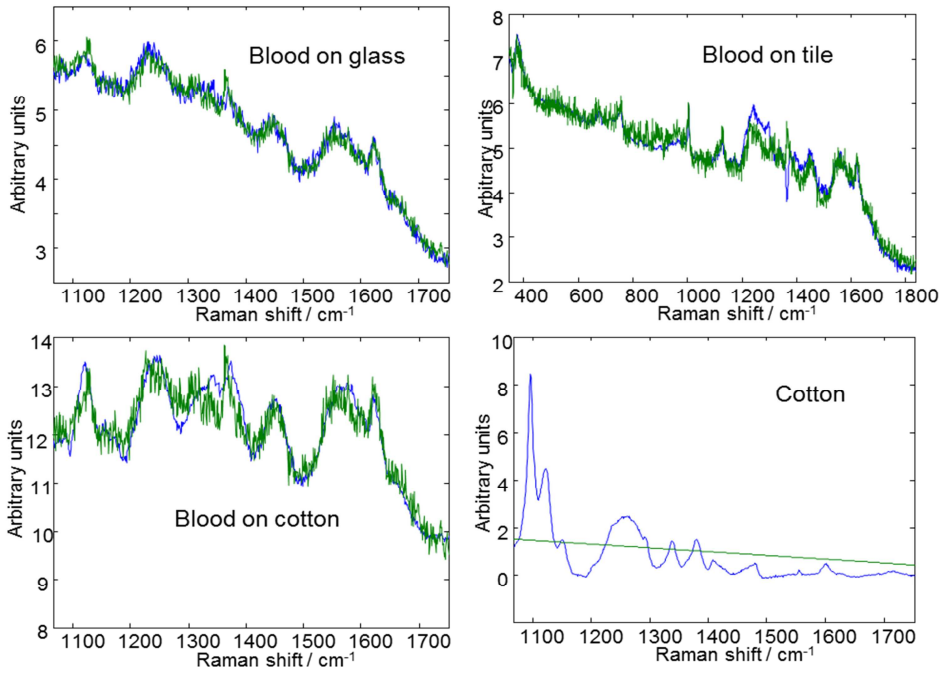

Figure 16. The Raman spectra of blood on different substrates with the fitted spectroscopic signature of blood (green lines). For comparison the same fitting procedure was performed for the Raman spectrum for pure cotton.

Table 5. Goodness-of-fit statistical results for blood, semen, and saliva signature fitting. The fitting procedure was performed for the Raman spectra collected from blood-stains on different substrates after subtraction of substrate contribution.

Signature	#	Sample	SSE	R^2	RMSE
Blood	1	Blood on glass	1.525	0.986	0.043
	2	Blood on tile	10.01	0.508	0.109
	3	Blood on cotton	1.893	0.981	0.048
	4	Cotton	14.16	0.639	0.131
	5	Tile	32.25	0.366	0.197
	6	Glass	52.61	0.364	0.256
	7	Blood on aluminum foil	0.231	0.992	0.015
Semen	8	Blood on glass	6.461	0.938	0.088
	9	Blood on tile	10.76	0.450	0.114
	10	Blood on cotton	27.83	0.686	0.183
Saliva	11	Blood on glass	8.891	0.901	0.104
	12	Blood on tile	10.31	0.493	0.417
	13	Blood on cotton	22.33	0.758	0.164

*All blood stain spectra had the substrate manually subtracted before fitting.

Note that fitting according to the multidimensional signature demonstrated superior results. Artifacts of spectral subtraction in the case of tile impaired the fitting procedure (Table 5. line 2). Line 7 of Table 5 contains the SSE, R^2, and RMSE statistical parameters calculated for the fitting of the multidimensional signature of blood to the averaged Raman spectra of pure blood collected from different donors and dried on aluminum foil. The SSE, R^2, and RMSE values in the case of blood on glass (Table 5, line 1) and cotton (Table 5, line 3) substrates are close to those of pure blood (Table 5, line 7), indicating that blood can be readily identified on these substrates with the help of the multidimensional signature. The specificity of the proposed approach was tested by using multidimensional signatures of other body fluids. Their application yielded worse results regardless of the substrate (Table 5, lines 8-13).

Blood recovery can be used as an alternative approach when the direct identification of blood stains on the substrate is not feasible. In this case, blood should be transferred to a cotton swab and analyzed using Raman spectroscopy. For our experiment, different dilutions of blood were deposited and tested (Figure 16). Table 6 contains the corresponding SSE, R^2, and RMSE values. The expected trend of the statistical criteria was worse with greater dilutions, although this trend is distorted due to the variable quality of data acquisition. However, preliminary results demonstrated that blood identification is possible up to a dilution of 1:40. Next, similar

analyses were performed with the Raman spectra of recovered blood (Table 6; see section 3.2 for more information about the recovering procedure). The validation of the identification thresholds will be addressed in our future studies.

Table 6. Goodness-of-fit statistical results of blood signature fitting. This fitting procedure was performed for the Raman spectra collected from diluted blood deposited on cotton swatches after the subtraction of substrate contribution. The shaded rows correspond to the Raman spectra that can be identified as blood.

#	Dilution coefficient	SSE	R^2	RMSE
1	2	2.38	0.971	0.058
2	4	11.16	0.821	0.126
3	6	2.49	0.968	0.060
4	8	4.25	0.950	0.078
5	10	20.68	0.684	0.172
6	20	6.42	0.919	0.096
7	40	2.10	0.976	0.055
8	60	12.3	0.813	0.133
9	N/A*	2.05	0.971	0.054

N/A – not applicable. Spectrum was collected from blood on a cotton swab recovered from a dried stain on a glass substrate.

This approach was successfully tested for other body fluid/substrate combinations including semen on snake skin. Our preliminary tests show that the Raman signature of semen can be easily obtained despite the fact that spectra of skin and semen share many similar features (Figure 17). More work is required to complete this study on a quantitative level. Also, this study needs to be expanded for human skin to make it most valuable for practical forensic science.

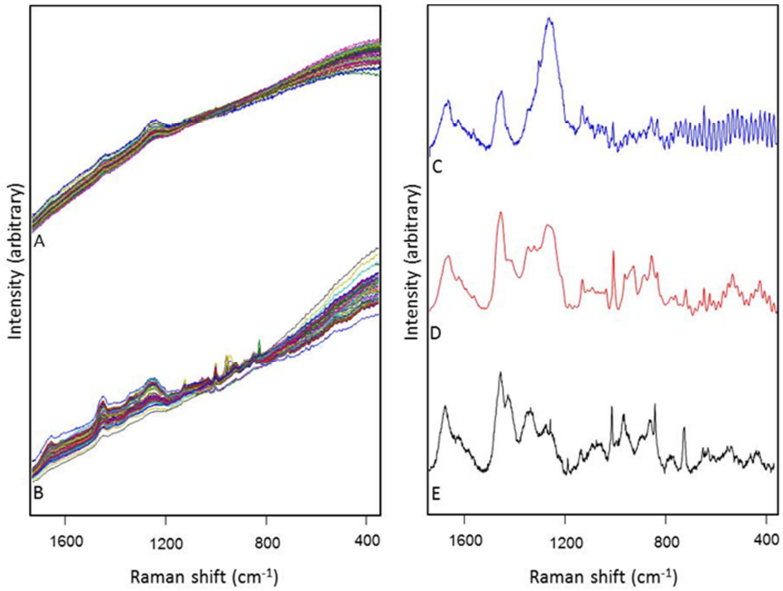

Figure 17. Raw Raman spectra of snake skin (A) and semen stain (B) on the skin. Baselined and averaged spectra of skin (C), semen stain on skin (D) and a reference spectrum of semen (E).

3.8 Conclusions

Simulated blood stain evidence was examined using Raman spectroscopy with special consideration to the issue of substrate interference. A range of substrates and instrumental parameters were compared to find the best method for eliminating interference to identify the blood trace on a sample. Although 785 nm laser excitation was expected to be the most promising, no single excitation laser source was ideal for all of the surveyed scenarios. Considering the raw spectra, the NIR laser source performed the worst with blood stains on glass and tile. Conversely, the NIR laser performed significantly better than all of the others with blood-stains on cotton. This finding may be a result of the light absorption properties of both the blood and the substrate at each excitation. Most importantly, these experiments demonstrate that in some circumstances, substrate interference can be avoided with the proper selection of the laser source. In instances where substrate interference is high, the contribution can be cleanly subtracted if the substrate is homogeneous.

The most problematic substrates were found to be those that contribute both Raman peaks and a strong fluorescent background and those that are heterogeneous in nature. The former was an issue with denim, which required both background correction and manual

subtraction of the substrate reference spectrum. The difficulty of heterogeneous substrates is in obtaining a reference spectrum, as was observed to be the case for bathroom tile.

Statistical fitting was performed using two methods to confirm the results of the visual analysis and to provide a measure of spectral quality. All experimental spectra, after substrate contribution subtraction, were compared to a blood reference spectrum. Pearson's correlation coefficients were used as a quantitative measure of the similarity between the spectra. We found that direct subtraction of substrate contribution works very well with particular combinations of substrates and excitations. Experimental Raman spectra (785 nm) were reproduced with even greater quality with the help of a multidimensional spectroscopic signature for pure blood (785 nm) designed to cover the intrinsic heterogeneity of dried blood-stains. The SSE, R^2, and RMSE statistical parameters of fitting indicated that the Raman spectra of blood can be easily distinguished from blank substrates. Control tests, particularly those performed using multidimensional signatures of semen and saliva, demonstrated the great selectivity of the proposed approach. This approach was successfully applied to other body fluid/substrate combinations.

4 Extend the application of the developed method to the dry body fluid mixtures contaminated with non-biological components (Objective 4)

The developed method of body fluid identification was extended for the dry traces, contaminated with non-biological components such as sand, dust, soil and all-purpose bleach.

4.1 Main approach

Our approach is based on a multistep, chemometrical procedure for distinguishing among pure fluids and fluids contaminated with dust, sand and soil. More than 500 spectra were obtained during the experimental procedure from contaminated samples as well as from pure contaminants. First, experimental spectra are subjected to fluorescent baseline subtraction and cosmic ray removal. The next step is the removal of uninformative, noisy spectra. This is accomplished by an algorithm, which uses the proximity of the recorded signal to Gaussian function as a criterion. Prepared spectra are trhen subjected to fitting by an alternating least squares (ALS) algorithm. Quantitative statistical analysis based on SSE, R^2, and RMSE was used for confirming a satisfactory fitting of experimental spectra by their corresponding body fluid signature.

4.2 Analysis of blood and semen, contaminated with sand, soil and dust

Raman spectra of sand, dust, soil and contaminated blood

The Raman spectra of sand, dust and soil are characterized by multiple peaks and diverse fluorescent backgrounds. The Raman spectrum of sand is dominated by intense narrow bands within the spectral ranges of 450-460, 700-710 and 1080-1090 cm^{-1} (Figure 18.A). Peaks at 700-710 and 1080-1090 cm^{-1} are characteristic calcite minerals[154,155] . A contribution from quartz, the main constituent of sand, is reflected by the Raman band at 460 cm^{-1} [156]. The Raman band at 1084 cm^{-1} was attributed to gypsum[154]. The Raman spectra of dust and soil, which are mixtures of inorganic and organic constituents, demonstrate a greater variability than the Raman spectra of sand (Figure 18.B-C).

Initial statistical analysis revealed the necessity of tens, if not hundreds, of spectral components to account for the spectral variability of sand, dust and soil. Therefore, the development of a complete library of all possible contaminants is impractical. We propose an alternative approach to address the interference of different contaminants (see *Fitting experimental Raman spectra according to the multidimensional spectroscopic signature of blood*). As expected, the Raman spectra of contaminated blood samples are very heterogeneous with varying contributions of contaminants, which cause varying fluorescent backgrounds (Figure 18. D-F). Certain spectra have very low signal-to-noise ratio significant contribution from fluorescence, cosmic rays, and artifacts originating from the saturation of the CCD detector (plateau). These spectra were not excluded from the following analysis in order to test the robustness of the proposed methodology.

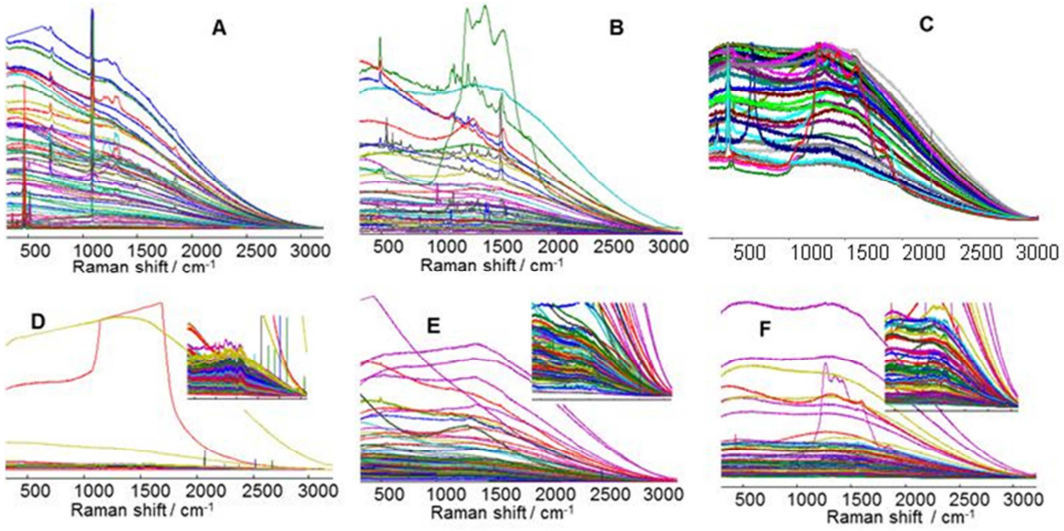

Figure 18. Raw Raman spectra of sand (A), dust (B), soil (C), and blood contaminated with sand (D), dust (E) and soil (F). All data sets are highly heterogeneous. Inserts are the original plots (D-F) scaled to demonstrate the variability of the less intense Raman spectra.

Fitting experimental Raman spectra according to the multidimensional spectroscopic signature of blood

During preliminary studies, we found that each new sample of sand, soil or dust demonstrated essentially unique Raman profiles. Because it is unfeasible to create a spectroscopic reference library of all possible contaminants, body fluid detection and identification should be based on the spectroscopic properties of the body fluid in question. We hypothesized that the multidimensional Raman signatures, which were originally developed for the identification of pure body fluid traces[157], could be used for the contaminated samples as well.

The identification process involves the determination of the experimental data, which can be efficiently fitted by a predefined set of spectra (signature). The basic idea is to build a model that will score each spectrum on a criterion whereby a point dominated by blood will result in a high score and a point dominated by a contaminant will score low.

Figure 19 illustrates a fitting of the blood signature to the experimental spectra of pure and contaminated blood (sand, dust, and soil) and the experimental spectra of contaminants. The fitting was accomplished using ALS statistics hard constrained to the spectral profiles of the blood signature. The ALS algorithm (PLS toolbox, Eigenvector Research Inc.) attempts to fit

each experimental spectrum using a linear combination of blood signature components. Figure 19 illustrates several typical fitting cases from more than one thousand calculations. The blue lines represent the experimental spectra, while the green lines are the result of the fitting. A good fitting quality was obtained for the experimental spectrum of pure blood (Figure 19.A). The original and calculated spectra closely resemble each other, while visual inspection reveals disagreements between the experimental and calculated spectra of the contaminants (Figure 19.B-D). For example, narrow Raman bands of crystalline sand particles are very different from those of blood (Figure 19.B)[143]. Despite of the high level of heterogeneity (Figure 18), the Raman spectra of dust and soil are distinct and cannot be described as a linear combination of blood's signature spectral components (Figures 19.C and 19.D).

Figures 19.E-F display the fitting of two selected Raman spectra obtained from the blood contaminated with sand. One spectrum (Figure 19.E) is dominated by sand contribution, while another (Figure 19.F) is typical for pure blood. Approximately two-thirds of the experimental spectra (Figure 19) have obvious similarities (by visual inspection) with the spectra of pure blood, revealing that Raman microspectroscopy of even thoroughly contaminated samples may provide enough evidence for the identification of blood. Similar results were observed in the case of dust (Figures 19.G and 19.H) and soil (Figures 19.I and 19.J). These observations suggest that it is possible to use the quality of fitting as a statistical criterion for the presence of blood within a contaminated bloodstain.

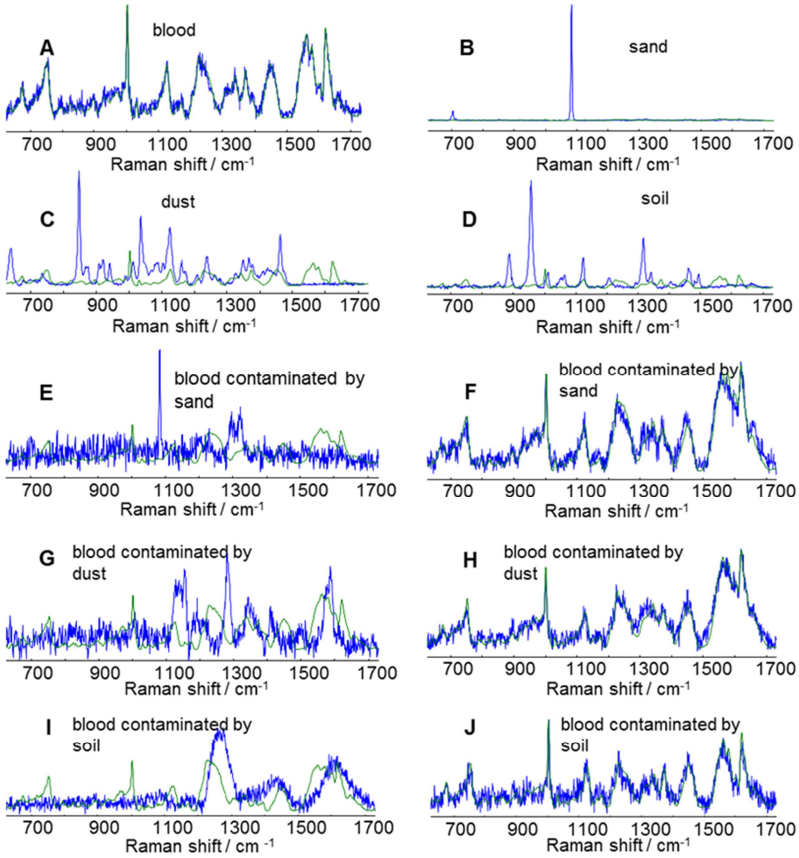

Figure 19. Selected Raman spectra of pure blood (A), sand (B), dust (C), soil (D), and blood contaminated by sand (E-F), dust (G-H), and soil (I-J). The green lines on plots (A-J) represent the result of fitting using the multidimensional Raman signature of blood.

Statistical criteria of fitting

Using MATLAB plotting tools, the fitted Raman intensity values were plotted with the experimental intensity values (Figure 19). The deviation of the experimental spectrum from the result of the fitting can be used for the determination of how close an experimental spectrum and signature match. SSE**,** R^2, and RMSE were used as quantitative parameters of the fitting quality (Table 2). The results of the SSE, R^2, and RMSE calculations are summarized in Figure 20, in which each colored symbol corresponds to a single experimental Raman spectrum. The SSE values for all of the pure blood spectra (red triangles) are very close and are found in a narrow range, while the SSE values for neat contaminants and contaminated blood samples span several orders of magnitude (Figure 20.A). The highest SSE values were observed for the Raman spectra of sand (green asterisks), dust (blue squares) and soil (light blue crosses). The Raman spectra

from the contaminated samples of blood are distributed between these two extreme levels. Figure 20.A was scaled for a better visual comparison of the variations in the SSE values (Figure 20.B).

A similar pattern was observed for the R^2 and RMSE values (Figures 20.C and 20.D). Goodness-of-fit values for all contaminants are substantially different from those for pure blood, for which variation is caused primarily by noise contribution. The fitting of sand spectra resulted in very large SSE and RMSE values and a very small R^2 parameter, reflecting the significant differences between the blood and sand spectra. The presence of significant concentrations of organic matter in soil and dust leads to some similarities between their Raman spectra and that of pure blood. However, regardless of the type of contaminants examined, they can all be easily distinguished from pure blood based on the statistical criteria.

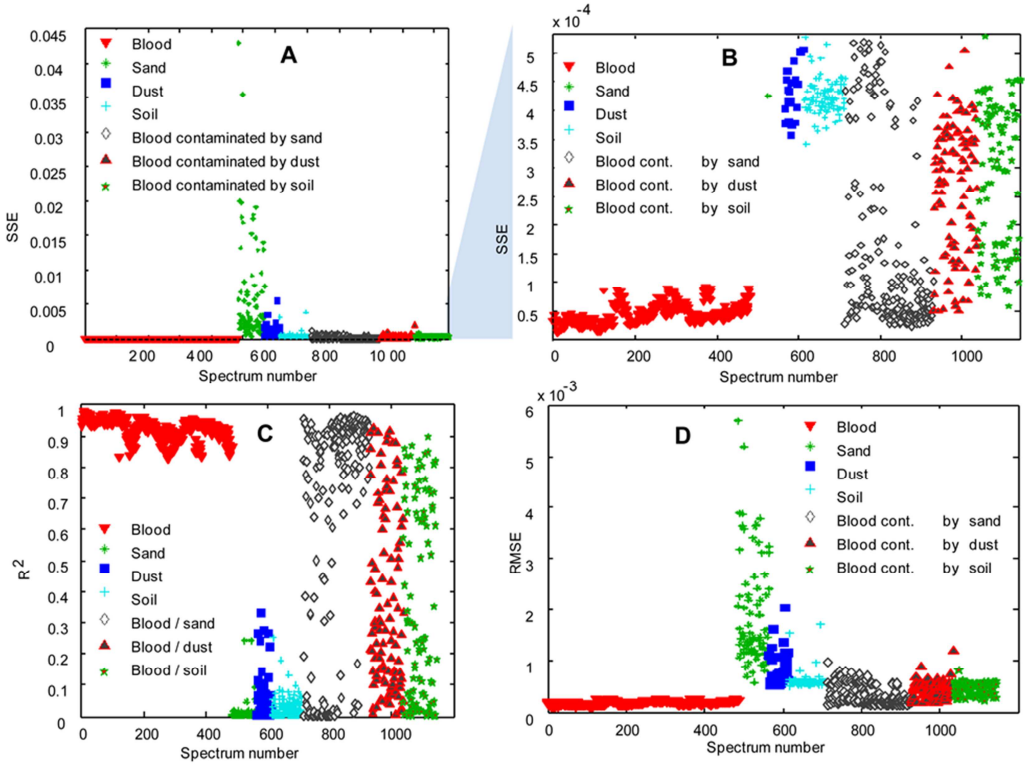

Figure 20. Statistical parameters of fitting using the multidimensional Raman signature of blood. (A) SSE values, (B) scaled SSE values, (C) R^2, (D) RMSE. Each colored symbol corresponds to a single experimental spectrum. The pure blood spectra are well fitted by blood signature (the SSE values are close to zero), while neat contaminants and contaminated blood samples demonstrate much greater dissimilarity (A and B). Similar behavior was observed for the R^2 and RMSE values (C and D). Goodness-of-fit values for all contaminants are substantially different from those for pure blood, for which variation is caused primarily by noise contribution.

Raman microspectroscopy of blood traces with heterogeneous contamination

Statistical analysis of the Raman spectra acquired from contaminated blood samples demonstrated a very high level of spectral heterogeneity. Spectra dominated by the contribution of a contaminant or blood only were found for each contaminated sample. Based on this result, we propose the following ***approach for the identification of body fluid traces in stains with heterogeneous contamination.*** Raman microspectroscopy in an automatic mapping mode is used to generate a spectroscopic data set representing a ***local*** chemical composition of the biological stain. Automatic software could be developed for selecting spectra with dominating contributions of body fluids based on the multidimensional Raman signatures. The detection and identification of body fluid traces would be based on the analysis of these selected spectra.

Using multivariate curve resolution (MCR), an estimation of the blood contribution was obtained for each spectrum acquired from blood samples contaminated with sand, soil or dust. Figure 21 illustrates the example of such an estimation in the case of blood contaminated with domestic dust. Here, multiple spectra with a significant contribution of blood were detected. Visual inspection of the Raman spectra that correspond to maximum values on this plot reveals obvious similarities to a reference blood spectrum. The blue and green arrows in Figure 21.A indicate two representative spectra. Corresponding spectral profiles are given in Figure 21.B. Moreover, the residuals of fitting (not shown) for such spectra (blue and green) demonstrate essentially normal distribution around zero, evidence for nearly pure blood spectra altered only by noise contribution. The Raman spectra that correspond to low values on the blood contribution plot reveal characteristic profiles for dust (red arrow/spectrum, Figure 21). Treatment of other blood samples contaminated with sand, dust and soil demonstrated similar results, consistently detecting “hot” spots of nearly pure blood.

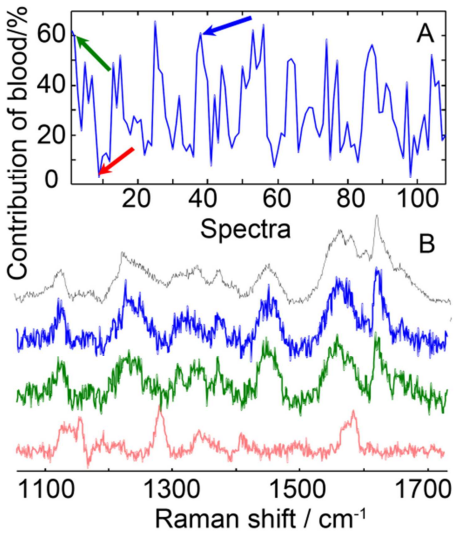

Figure 21. The result of fitting a blood signature to Raman spectra of a dried bloodstain contaminated with domestic dust. A) The Y-axis represents the estimated blood contribution, while numbers on the X-axis represent the number of experimental spectra recorded. The lower plot (B) shows the experimental spectra, which corresponds to the points in the upper plot marked by arrows with the corresponding colors. The black line is the spectrum of a dried stain of pure human blood.

Statistical analysis of individual Raman spectra was used to detect and identify blood within the contaminated samples. Figure 22.A displays the results of PLS-DA analysis using SSE, R^2 and RMSE values as input data. Training of the PLS-DA classification model was performed using data for four classes: pure blood, dust, sand and soil. Data obtained from the contaminated samples of blood were used as a test data set.

The Y-axis in Figure 22.A corresponds to the predicted “class values” of a particular Raman spectrum, or the propensity of that spectrum to be identified as blood. A value of 1 indicates that a spectrum belongs to a class (blood), and a value of 0 indicates that it does not. Error bars indicate the uncertainty in the reported values. The horizontal red dashed line represents the prediction threshold estimated using Bayes' theorem and the available data in order to minimize total errors. The threshold is selected at the Y-value, at which the number of false positives and false negatives are minimized (PLS toolbox, Eigenvector Research, Inc., Wenatchee). In Figure 22.A, it can be observed that all pure blood spectra were correctly classified and that there are no spectra for contaminants that are close to the threshold line.

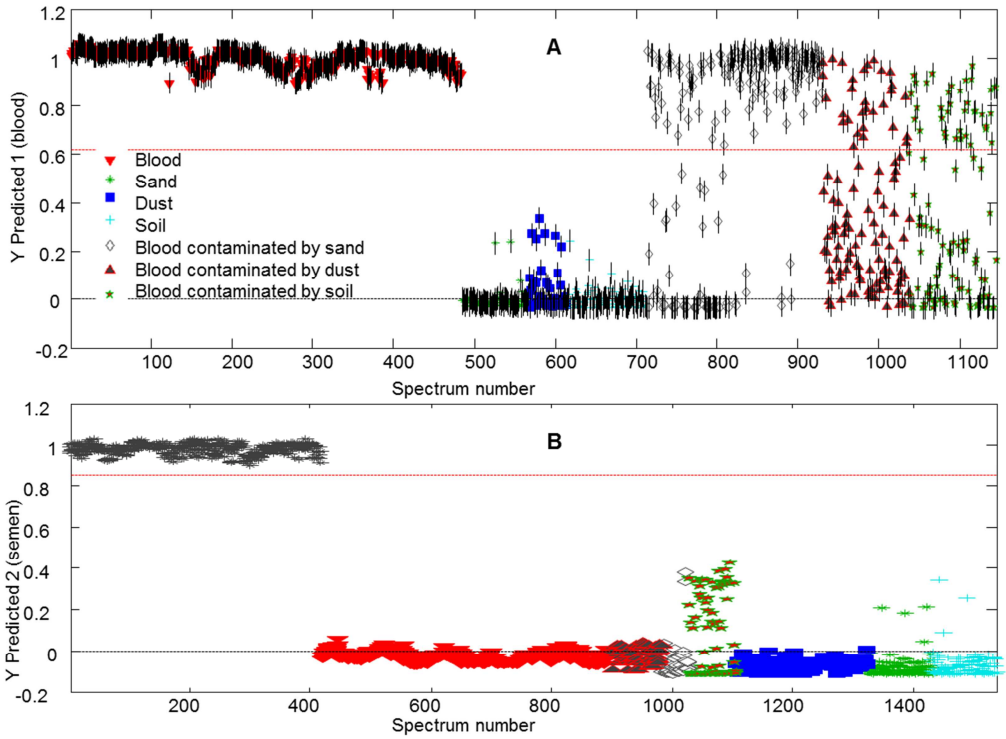

Figure 22. A) PLS-DA analysis of statistical parameters (SSE, R^2, RMSE) of blood signature fitting. The Y-axis corresponds to the predicted "class values" of Raman spectra, or the propensity of a spectrum to be identified as blood. The horizontal red dashed line is the threshold of blood identification estimated using Bayes' theorem. Each colored symbol corresponds to a single experimental spectrum. B) PLS-DA analysis of the statistical parameters of a semen signature fitting. Here, the Y-axis represents the propensity of a spectrum to be identified as semen. The horizontal red dashed line is the threshold of semen identification estimated using Bayes' theorem. The gray crosses correspond to the experimental spectra of a pure semen stain. Errors bars, comparable with those in (A), were omitted to simplify the visual examination. The results were verified by leave-one-out cross-validation.

To compare how accurately Raman spectra of contaminants and contaminated blood can be distinguished from other body fluids, we fitted them to the multidimensional spectroscopic signature of semen (5). Figure 22.B displays the results of the PLS-DA analysis of the statistical parameters of semen signature fitting. For this experiment, training of the PLS-DA classification model was performed using data from four classes: pure semen, dust, sand and soil. Here, a 1 indicates that a spectrum belongs to a class (semen), and a 0 indicates that it does not. All blood,

contaminants and contaminated blood spectra are not assigned to semen, indicating the specificity of the model. We would like to note that the Raman spectra of pure blood and semen were identified with zero false positive and false negative (Figure 22). The percentage of captured pure blood spectra depends on the type (Figure 22) and level of contamination. Nevertheless, the results reported for highly contaminated samples clearly demonstrate the great potential of Raman spectroscopy in combination with advanced statistical analysis for the nondestructive and confirmatory identification of body fluid stains.

4.3 Identification of contaminated mixtures

Raman spectra acquired from contaminated mixtures were subjected to the analysis using approach developed for characterization of pure mixtures. An automatic scanning allowed for detecting sports dominated only by body fluid contribution even within highly contaminated body fluid stains (see Chapter 4.2 *Analysis of blood and semen, contaminated with sand, soil and dust*). Raman spectra collected from such sports can be identified. Raman spectra dominated by contaminants or noise contribution were not assigned to any class. Visual inspection revealed that all non-assigned spectra exhibited either the characteristic of contaminants, cosmic rays or low signal-to-noise ratio. The mapping procedure should provide more reliable information about the real crime scene evidence and determine a right sport for further DNA/RNA profiling.

4.4 Effect of cleaning materials on the identification of body fluid stains

Semen and blood were contaminated with various concentrations of all-purpose bleach. Raman spectra of mixtures have characteristic peaks of bleach and the corresponding body fluid. The active ingredients of bleach are sodium carbonate anhydrous, sodium hypochorite, and water. As previously stated, water displays very little interference (in a Raman spectrum). The major peaks present in the Raman spectrum of bleach are located at approximately 1060 and 936 cm^{-1}, and they are from carbonate[158], and symmetrical ClO_3^- stretching [159] respectively. Using the approach developed for other contaminants (dust, sand and soil) we were able to detect the presence of body fluid within the stain in their mixtures with all-purpose bleach (up to 10x dilutions). For example, Figure 23 demonstrates the Raman spectra of 2:1 and 4:1 bleach/semen mixtures.

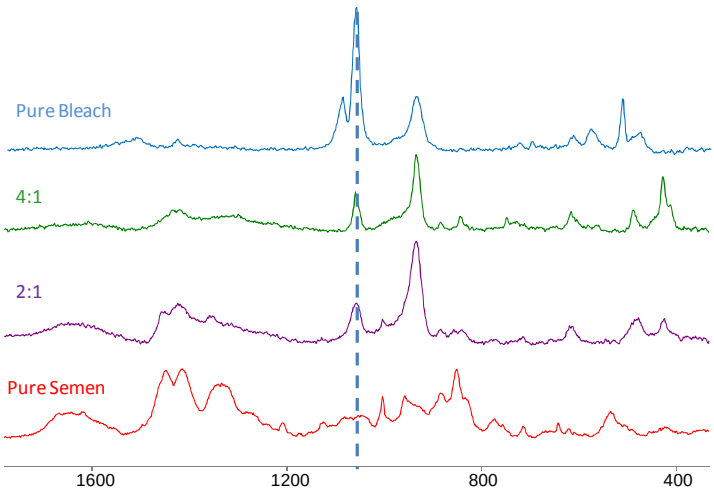

Figure 23. Raman spectra of bleach and semen mixtures. Vertical line represents Raman shift at 1060cm^{-1}.

4.5 Conclusion

It was discovered that body fluid stains can be identified, even when the contaminant contribution dominated the spectrum. The proposed approach is based on the analysis of spectra recorded via automatic mapping, comparison with the library of multidimensional spectroscopic signatures, and identification of spectra dominated by the contribution from an individual body fluid.

Preliminary analysis allows us to conclude, that Raman spectroscopy has the potential to identify body fluid stains altered by cleaning materials (all-purpose bleach). However, our approach successfully probed with chemically inactive contaminants will require further modification to take into account the possibility of chemical reactions with active ingredients of cleaning materials.

5 Determine the optimum wavelength range for Raman spectroscopic measurements from the viewpoint of the method selectivity, sensitivity, and reduction of fluorescence interference (Objective 5)

The optimal wavelength range was determined from the viewpoint of substrate interference. Six laser excitations (including 785 nm) were evaluated using blood and semen. The excitation wavelength was not that critical for blood identification on various substrates. In the case of semen, 785-nm excitation was preferable. At all other laser excitations, the spectrum was dominated by substrate fluorescence. We demonstrated that the substrate interference can be reduced using an optimal laser excitation. Additionally, we found that no single laser excitation was ideal in all possible scenarios. This is discussed in more detail in section 3.1-3.4. To get reliable representation of narrow Raman bands within the Raman spectra of body fluids spectroscopic data were acquired with ~1 cm^{-1} spectral resolution.

IV. Conclusions

1. Discussion of findings

<u>Objective 1</u>. Obtain statistically confident Raman spectroscopic characteristics for various body fluids including blood, semen, vaginal fluid, saliva, and sweat

Multidimensional Raman spectroscopic characteristics (spectroscopic signatures) were obtained for blood, semen, vaginal fluid, saliva, and sweat. Raman spectroscopy coupled with advanced statistical analysis demonstrates great potential for the non-destructive and rapid confirmatory identification of body fluids at crime scenes. Raman spectroscopic signatures were tested in multiple samples from a variety of pure body fluids obtained from donors of different races, genders, and ages. These signatures were fitted to the experimental Raman spectra of various dried body fluid stains. High goodness-of-fit statistical results (SSE, R^2, and RMSE) were obtained for consistent body fluid-Raman signature pairs. We demonstrated that the use of spectroscopic signatures could be applied to any human body fluid sample for its potential identification and in distinguishing it from other body fluids and substances found at a crime scene. Spectral features of the Raman spectral components were assigned to individual biochemical constituents.

The identification of pure body fluids can also be achieved by such chemometrical methods as PCA, linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA), support vector machines (SVM), and artificial neural networks (ANN). For example, we demonstrated that human, feline and canine blood samples can be distinguished using PCA. Also, LDA, SIMCA and partial least squares discriminant analysis (PLS-DA) were tested as alternative methods for the discrimination of semen, blood, and saliva traces. The application of multivariate classification methods will be significantly impaired in cases of contamination, body fluid mixing, aging and spectral contributions from substrates. We believe that the effective identification of real-life samples can be achieved using the combination of multidimensional signature fitting with multivariate classification and regression analyses.

Objective 2. Develop a statistical program for automatic identification of body fluids in the dry mixtures. Evaluate major performance characteristics and limitations of the developed method as presumptive and confirmatory tests

The developed software uses the concept of multidimensional signatures combined with advanced statistical analysis for the identification of body fluids with a high degree of accuracy. The program analyzes a set of recorded spectra and compares them with the library of multidimensional spectroscopic signatures built for five body fluids. The application of multidimensional spectroscopic signatures overcomes the intrinsic heterogeneity of dry samples and possible variations with donors. The obtained results were tested by extensive cross-validation. The novel statistical approach based on the combination of classification and regression SVM methods was proposed to examine the composition of body fluid mixtures. The detection limit can be as low as only several percents of minor contributors. The proposed methodology was successfully probed with semen-blood mixtures and is under extension to the characterization of other complex mixtures.

Objective 3. Extend the application of the developed method to body fluid stains on various substrates including human skin, fabrics, carpets, hard surfaces, etc.

The application of the developed method was extended for various substrates including cotton, denim, glass and tile. It was discovered that the substrate interference could be cleanly subtracted from spectra of blood stains, even when the substrate contribution dominated the spectrum. In addition, un-dyed cotton was seen to produce very low interference such that to the

extent that dilute blood, up to 1:40, was still identifiable. Suspected blood stain cotton swabs could be suitable for identification with Raman spectroscopy in cases where evidence cannot otherwise be analyzed or collected *in situ*.

The substrates considered for this research was fairly homogeneous. It is imperative that this research continues with heterogeneous substrates and transparent body fluids, such as semen. We expect semen to have a smaller Raman cross section, and since it is mostly transparent, substrate interference would be a much larger factor. This assumption was seen to be accurate. The disparity of signal between luminescent substrates and semen was to the degree that no single collection point was useful for classification; subtraction results were very noisy and not clearly identifiable. For this reason, a Raman mapping approach was utilized. We demonstrate that with semen on tile, the signal from the fluid can be extracted and classified. This method also accounts for substrate heterogeneity.

Objective 4. Extend the application of the developed method to the dry body fluid mixtures contaminated with non-biological components

The concept of multidimensional signatures was also extended to the dry body fluid mixtures contaminated with non-biological components. The approach is based on the analysis of recorded spectra, comparing them with the library of multidimensional spectroscopic signatures of body fluids, and identification of spectra dominated by the contribution from an individual body fluid. The proposed methodology was successfully probed with semen and blood contaminated by sand, dust and soil.

We declined the idea to create a full library of the contaminants because of the huge variability of Raman spectra of contaminants. For example, at least one hundred characteristic spectra are required to account the intrinsic heterogeneity of dust components. The proposed approach is based on analyzing data from the “hot spots”, where body fluid contributions are more obvious.

Mixtures of all-purpose bleach with blood and semen were studied by Raman microspectroscopy. Analysis of the obtained data showed that the characteristic peaks of body fluids presented in their respective Raman spectra allow us to conclude that Raman spectroscopy has the potential to identify even altered body fluid stains. However, preliminary studies of semen and blood samples diluted with large amounts of bleach revealed that body fluid

constituents are involved in chemical reactions with active ingredients of bleach. This approach successfully probed with chemically stable samples requires further modification to take into account the described complications.

Objective 5. Determine the optimum wavelength range for Raman spectroscopic measurements from the viewpoint of the method selectivity, sensitivity, and reduction of fluorescence interference

For the development of spectroscopic signatures, 785 nm laser excitation was used. The optimal wavelength range was determined from the viewpoint of substrate interference. Six laser excitations (including 785 nm) were evaluated using blood and semen. It was found that ~~only~~ 785 nm light was preferable in the case of semen. At all other laser excitations, the spectrum was dominated by fluorescence.

2. Implications for policy and practice

The obtained results clearly demonstrate a great potential of Raman spectroscopy and establish a methodological basis for the development of a portable instrument for an easy-to-use, on-field, rapid, nondestructive, confirmatory identification of body fluid traces discovered at a crime scene. This instrument could also be used in a forensic laboratory for assessing the probative value of evidentiary samples. There is every indication that forensic community appreciates the results of our studies. Our 2009 paper is the most downloaded article from *Forensic Science International*, the top journal in the field, in 2010-2011. Total of 10 articles were published or submitted for publication in peer-reviewed journals, more than 15 presentations were made at meetings conferences. Several scientific and general news agencies published press coverage about our research:

BioOptics World press coverage. SPECTROSCOPY/FORENSICS: CSI: Multidimensional Raman spectroscopy. This feature article made the magazine cover.

January 1, 2012

http://www.bioopticsworld.com/articles/print/volume-5/issue-1/features/multidimensional-raman-spectroscopy.html

Forensic science: CSI Raman

Raman Newsletter from SpectroscopyNOW.com, November 1, 2011

http://www.spectroscopynow.com/coi/cda/detail.cda?chId=6&id=26318&type=Feature&page=1

Albany school on cutting edge of forensics

The Daily Gazette. January 30, 2011 by Sara Foss

Seat at forensic table

A University at Albany professor has been appointed by the White House to help shape the field of forensic science for the next two decades

Times Union newspaper, January 20, 2011

http://www.timesunion.com/default/article/Opera-offers-vision-of-a-peaceful-Mideast-966144.php

UAlbany professor develops CSI Raman technique

Press coverage in *spectroscopyNOW.com*, August, 2010

http://www.spectroscopynow.com/coi/cda/detail.cda?id=23983&type=News&chId=2

University at Albany Researcher Develops CSI Technique

Press coverage in *Forensic* Magazine, July 16, 2010

http://www.forensicmag.com/news/university-albany-researcher-develops-csi-technique

UAlbany professor develops CSI technique

Times Union newspaper, July 14, 2010 by Carol DeMare

http://www.timesunion.com/AspStories/story.asp?storyID=950434&category=REGION

UAlbany Researcher Identifies Novel Method to Improve Crime Scene Investigations

Press release at the University at Albany website. July 12, 2010

http://www.albany.edu/news/release_9789.php?WT.svl=headline

Kelly Virkler, '09. Discovering New Frontiers of Forensic Science

Press release at the University at Albany website. 2010

http://www.albany.edu/news/profile_8195.php

Non-destructive spit test

Raman Newsletter from SpectroscopyNOW.com, 1 March, 2010

http://www.spectroscopynow.com/coi/cda/detail.cda?id=23147&type=Feature&chId=6&page=1

Forensic science calls woman

FLUID ANALYSIS EXPERT: Watertown High School grad making waves

Watertown Daily Times, Thursday, December 31, 2009, by Jamie Munks

http://www.watertowndailytimes.com/article/20091231/NEWS03/312319970

Blood will tell

By Hank Hogan, Biophotonics, November/December 2009, 16-17

Raman Spectroscopy: As Seen in CSI? *October 11, 2009*:

http://calvinus.wordpress.com/2009/10/11/raman-spectroscopy-as-seen-in-csi/

Species in a snap: Raman analysis of blood

A news story published by the Analytic Chemistry, number one journal in the field

(by Erika Gebel, Anal. Chem., 2009, 81, 7862):

http://pubs.acs.org/doi/pdf/10.1021/ac901827u

http://pubs.acs.org/journal/ancham

3. Implications for further research

The completed research clearly indicates the great potential and the necessity of the development of a portable Raman instrument for body fluid identification at a crime scene. The developed methodology and software can be adopted for such an instrument.

The practical application of our new method will require understanding of how aging affects the spectral response from body fluid traces. This question is beyond the scope of the currently funded program. Our preliminary observations indicate that the Raman signature of dry blood evolves with time in a reproducible way, which suggests that Raman spectroscopy can be potentially used for determining the age of a blood stain. One can envision that blood stain aging should depend on the environmental conditions. We also hypothesize that traces of other body fluids should exhibit predictable changes with time as well. The development of a portable instrument, which would allow for determining the age of a body fluid stain on a crime scene, will have a major impact on the efficiency of crime scene investigation in at least two major aspects. First, determining the stain age could help to establish the time of the crime. Secondly, determining the age of numerous body fluid stains recovered at a crime scene should allow CSIs to prioritize the collected evidence to those, which are time-related to the crime. Provided that the age variation of Raman spectroscopic signatures of body fluid traces is experimentally established, it would be straight forward to incorporate these data to the software, which was developed during this study for the identification of body fluid traces.

In addition, our preliminary results clearly indicate that Raman spectroscopy combined with advanced statistical analysis is capable of differentiating species based on blood traces[160]. Further study is required to implement this capability to the method developed here. The ability to differentiate human blood stains at a crime scene from those of animal origin will have a

decisive impact on the efficiency of crime scene investigation since there are many potential sources of blood related to food preparation and domestic animals. Expanding the capability of Raman spectroscopic method to differentiate species to other body fluids will further increase the power of the developed method.

While working with body fluid samples collected from diverse group of donors we have observed that the Raman spectroscopic signature of an individual's fluids depends on race and gender. Again, expanding the developed methodology to genetic profiling, determining the gender and race, will significantly enhance the power of the method.

V. References

(1) Greenfield, A.; Sloan, M. A. In *Forensic science: an introduction to scientific and investigative techniques*; James, S. H., Nordby, J. J., Eds.; CRC Press: Boca Raton, 2003, p 203.

(2) Spalding, R. P. In *Forensic science: an introduction to scientific and investigative techniques*; James, S. H., Nordby, J. J., Eds.; CRC Press: Boca Raton, 2003, p 181.

(3) Watson, N. In *Crime scene to court; the essentials of forensic science*; Royal Society of Chemist.: Cambridge, UK, 2004, p 377.

(4) Jones Jr., E. L. In *Forensic science handbook*; Saferstein, R., Ed.; Prentice Hall: Upper Saddle River, NJ, 2005; Vol. 2, p 329.

(5) Shaler, R. C. In *Forensic science handbook*; Saferstein, R., Ed.; Prentice Hall: Upper Saddle River, NJ, 2002; Vol. 1, p 529.

(6) Virkler, K.; Lednev, I. K. *Forensic Sci Int* 2009, *188*, 1.

(7) Vandenberg, N.; van Oorshot, R. A. H. *J. Forensic Sci.* 2006, *51*, 361.

(8) Barni, F.; Lewis, S. W.; Berti, A.; Miskelly, G. M.; Lago, G. *Talanta* 2007, *72*, 896.

(9) Webb, J. L.; Creamer, J. I.; Quickenden, T. I. *Luminescence* 2006, *21*, 214.

(10) Quinones, I.; Sheppard, D.; Harbison, S.; Elliot, D. *Can. Soc. Forens. Sci. J.* 2006, *40*, 53.

(11) Luczak, S.; Wozniak, M.; Papuga, M.; Stopiniska, K.; Sliwka, K. *Arch. Med. Sadowej. Kryminol.* 2006, *LVI*, 239.

(12) Castello, A.; Alvarez, M.; Verdu, F. *Can. Soc. Forens. Sci. J.* 2002, *35*, 113.

(13) Garofano, L.; Pizzamiglio, M.; Marino, A.; Brighenti, A.; Romani, F. *Int. Con. Ser.* 2006, *1288*, 657.

(14) Blum, L. J.; Esperanca, P.; Rocquefelte, S. *Can. Soc. Forensic Sci. J.* 2006, *39*, 81.

(15) Johnston, E.; Ames, C. E.; Dagnall, K. E.; Foster, J.; Daniel, B. E. *J. Forensic Sci.* 2008, *53*, 687.

(16) Li, R. *Forensic Biology*; CRC Press: Boca Raton, 2008.

(17) Gaensslen, R. E. *Sourcebook in forensic serology, immunology, and biochemistry*; U.S. Department of Justice: Washington, D.C., 1983.

(18) Dixon, T. R.; Samudra, A. V.; Stewart Jr., W. D.; Johari, O. *J. Forensic Sci.* 1976, *21*, 797.

(19) Sottolano, S.; de Forest, P. R. *The Microscope* 1980, *28*, 41.

(20) Kotowski, T. M.; Grieve, M. C. *J. Forensic Sci.* 1986, *31*, 1079.

(21) Kashyap, V. K. *J. Immunoassy* 1989, *10*, 315.

(22) Sensabaugh, G. F. In *Isozymes: current topics in biological and medical research*; Rattazzi, M. C., Scandalios, J. G., Whitt, G. S., Eds.; Alan R. Liss, Inc.: New York, 1982; Vol. 6, p 247.

(23) Komuro, T.; Mukoyama, R.; Mukoyama, H. *Jpn. J. Clin. Med.* 1995, *53*, 2322.

(24) Santucci, K. A.; Nelson, D. G.; McQuillen, K. K.; Duffy, S. J.; Linakis, J. G. *Pediatrics* 1999, *104*, 1342.

(25) Nelson, D. G.; Santucci, K. A. *Acad. Emerg. Med.* 2002, *9*, 1045.

(26) Seiden, H.; Duncan, G. T. *Assoc. Off. Anal. Chem.* 1983, *66*, 207.

(27) Banerjee, P. K. *J. Inst. Chemists (India)* 1983, *55*, 23.

(28) Toates, P. *Forensic Sci. Int.* 1979, *14*, 191.

(29) Stolorow, M.; Hauncher, J. D.; Stuver, W. C. *J. AOAC* 1976, *59*, 1352.

(30) Kido, A.; Oya, M.; Katsumata, Y.; Suzuki, O.; Fujisawa, K. *Act. Crim. Japan* 1979, *45*, 127.

(31) Suzuki, O.; Hattori, H.; Katsumata, Y.; Oya, M.; Mitzutani, S. *Jpn. J. Legal Med.* 1981, *35*, 356.

(32) Kosa, F.; Rengei, B.; Szendrenyi, J. *Krim. Forens. Wiss.* 1987, *67-68*, 123.

(33) Suzuki, O.; Oya, M.; Kido, A.; Katsumata, Y.; Asano, M. *Z. Rechtsmed* 1980, *86*, 35.

(34) Suzuki, O.; Asano, M.; Kido, A.; Oya, M. *Forensic Sci. Int.* 1983, *22*, 231.

(35) Hooft, P. J.; van de Voorde, H. P. *Am. J. Forens. Med. and Path.* 1997, *18*, 45.

(36) Noppinger, K.; Morrison, R.; Jones, N. H.; Hopkins, H. *J. Forensic Sci.* 1987, *32*, 1069.

(37) Suzuki, O.; Matsumoto, T.; Oya, M.; Matsumata, Y.; Asano, M. *J. Forensic Sci.* 1981, *26*, 410.

(38) Manabe, F.; Tsutsumi, A.; Yamamoto, Y.; Hashimoto, Y.; Ishizu, H. *Jpn. J. Legal Med.* 1991, *45*, 205.

(39) Oki, Y.; Tsutsumi, A.; Ishizu, H. *Jpn. J. Legal Med.* 1987, *42*, 1.

(40) Gonmori, K.; Yoshioka, N.; Chattopadhyay, P. K. *Act. Crim. Japon.* 1994, *60*, 149.

(41) Tsutsumi, A. *Jpn. J. Legal Med.* 1987, *41*, 294.

(42) Suzuki, O.; Oya, M.; Katsumata, Y.; Matsumoto, T.; Yada, S. *J. Forensic Sci.* 1980, *25*, 99.

(43) Sato, S.; Moriya, F.; Yamamoto, Y.; Ishizu, H. In *Current topics in forensic science: proceedings of the 14th meeting of the International Association of Forensic Sciences*; Shunderson Communications: Tokyo, Japan, 1996, p 40.

(44) Seta, S. *Int. Crim. Police Rev.* 1977, *307*, 119.

(45) Healy, D. A.; Hayes, C. J.; Leonard, P.; McKenna, L.; O'Kennedy, R. *Trends in Biotech.* 2007, *25*, 125.

(46) Tsuda, R.; Hara, M.; Sagawa, K. *Jpn. J. Legal Med.* 1984, *38*, 83.

(47) Rao, D. V.; Kashyap, V. K. *J. Immunoassay* 1992, *13*, 537.

(48) Matsuzawa, S.; Itoh, Y.; Kimura, H.; Kobayashi, R.; Nakagawa, T.; Ohno, S. *Forens. Sci. Int.* 1994, *64*, 119.

(49) Masibay, A. S.; Lappas, N. T. *J. Forensic Sci.* 1984, *29*, 1173.

(50) Maher, J.; Vintiner, S.; Elliot, D.; Melia, L. *New Zeal. Med. J.* 2002, *115*, 48.

(51) Hochmeister, M. N.; Budowle, B.; Rudin, O.; Gehrig, C.; Borer, U.; Thali, M.; Dirnhofer, R. *J. Forensic Sci.* 1999, *44*, 1057.

(52) Kearsey, J.; Louie, H.; Poon, H. *Can. Soc. Forens. Sci. J.* 2001, *34*, 63.

(53) Denison, S. J.; Lopes, E. M.; D'costa, L.; Newman, J. C. *Can. Soc. Forensic Sci. J.* 2004, *37*, 197.

(54) Sato, I.; Sagi, M.; Ishiwari, A.; Nishijima, H.; Ito, E.; Mukai, T. *Forensic Sci. Int.* 2002, *127*, 71.

(55) Yokota, M.; Mitani, T.; Tsujita, H.; Kobayashi, T.; Higuchi, T.; Akane, A.; Nasu, M. *Legal Medicine* 2001, *3*, 171.

(56) Simich, J. P.; Morris, S. L.; Klick, R. L.; Rittenhouse-Diakun, K. *J. Forensic Sci.* 1999, *44*, 1229.

(57) Herr, J. C.; Summers, T. A.; McGee, R. S.; Sutherland, W. M.; Sigman, M.; Evans, R. J. *Biology of reproduction* 1986, *35*, 773.

(58) Sato, I.; Yoshiike, M.; Yamasaki, T.; Yoshida, K.; Takano, S.; Mukai, T.; Iwamoto, T. *Forens. Sci. Int.* 2001, *122*, 27.

(59) Sato, I.; Kojima, K.; Yamasaki, T.; Yoshida, K.; Yoshiike, M.; Takano, S.; Mukai, T.; Iwamoto, T. *J. Immunol. Methods* 2004, *287*, 137.

(60) King, S. J.; Sutton, J. G.; Trewsdale, L. A. *Forens. Sci. Int.* 1989, *40*, 221.

(61) Lolov, S. R.; Yomtova, V. M.; Tsankov, Y.; Kehayov, I. R.; Kyurkcheniev, S. D. *Forens. Sci. Int.* 1992, *54*, 39.

(62) Allen, S. M. *J. Immunoassay* 1995, *16*, 297.

(63) Tsutsumi, A.; Yamamoto, Y.; Okamura, K.; Oki, Y.; Hamada, K.; Ishizu, H. *Act. Crim. Japan* 1984, *50*, 1.

(64) Mokashi, R. H.; Malwankar, A. G.; Madiwale, M. S. *J. Indian Acad. Forens. Sci.* 1975, *14*, 1.

(65) Auvdel, M. J. *J. Forensic Sci.* 1988, *33*, 929.

(66) Quarino, L.; Hess, J.; Shenouda, M.; Ristenbatt, R. R.; Gold, J.; Shaler, R. C. *J. Forensic Sci. Soc.* 1993, *33*, 87.

(67) Martin, N. C.; Clayson, N. J.; Scrimger, D. G. *Sci. and Just.* 2006, *46*, 97.

(68) Troeger, H. D.; Schuck, M.; Tutsch-Bauer, E. *Forens. Sci. Int.* 1984, *25*, 143.

(69) Ohya, I.; Iwasa, M.; Komoriya, H.; Bunai, Y.; Sagisaka, K. *Tohoku J. Exp. Med.* 1986, *150*, 309.

(70) Kozu, T.; Ichinose, T.; Komatsu, Y.; Kakegawa, K. *Rep Natl Res Inst Police Sci* 1977, *30*, 18.

(71) Bedrosian, J. L.; Stolorow, M. D.; Tahir, M. A. *J. Forensic Sci.* 1984, *29*, 601.

(72) Kimura, A.; Nakashima, R.; Yasuda, S.; Takata, H.; Ikemoto, E.; Nosaka, M.; Tsuji, T. In *Current topics in forensic science: proceedings of the 14th meeting of the International Association of Forensic Sciences*; Shunderson Communications: Tokyo, Japan, 1996, p 34.

(73) Tsutsumi, H.; Okajima, H.; Sato, K.; Katsumata, Y. *J. Forensic Sci.* 1988, *33*, 237.

(74) Taylor, M. C.; Hunt, J. S. *J. Forensic Sci. Soc.* 1983, *23*, 67.

(75) Sato, K.; Tsutsumi, H.; Htay, H. H.; Tamaki, K.; Okajima, H.; Katsumata, Y. *Forensic Sci. Int.* 1990, *45*, 27.

(76) Hayashiba, Y.; Kashimura, S.; Kageura, M.; Kimura, K.; Nakazono, T. In *Current topics in forensic science, proceedings of the meeting of the international association of forensic sciences*; Takatori, T., Takasu, A., Eds.; Shunderson Communications: Tokyo, Japan, 1996; Vol. 2, p 19.

(77) Nakazono, T.; Kashimura, S.; Hayashiba, Y.; Hisatomi, T.; Hara, K. *J. Forensic Sci.* 2002, *47*, 568.

(78) Hisatomi, T.; Nakazono, T.; Hayashiba, Y.; Fukushima, S. *Nippon Iyo Masu Supekutoru Gakkai Koenshu* 1998, *23*, 145.

(79) Divall, G. B. *Forens. Sci. Int.* 1984, *24*, 239.

(80) Hausmann, R.; Baltzer, M.; Schellmann, B. *Int. J. Legal Med.* 1996, *109*, 10.

(81) Martin, P. D.; Cheshire, S. K. *Adv. in Forens. Haemogen.* 1986, *1*, 299.

(82) Bauer, M. *Forensic Sci. Int. Gen.* 2007, *1*, 69.

(83) Alvarez, M.; Juusola, J.; Ballantyne, J. *Anal. Biochem.* 2004, *335*, 289.

(84) Juusola, J.; Ballantyne, J. *Forensic Sci. Int.* 2003, *135*, 85.

(85) Ballantyne, J.; Juusola, J.; Research Foundation of The University of Central Florida, Inc.: USA, 2007; Vol. 7270983.

(86) Juusola, J.; Ballantyne, J. *Forensic Sci. Int.* 2005, *152*, 1.

(87) Bauer, M.; Patzelt, D. *J. Forensic Sci.* 2002, *47*, 1278.

(88) Bauer, M.; Patzelt, D. *Forens. Sci. Int.* 2008, *174*, 54.

(89) Nussbaumer, C.; Gharehbaghi-Schnell, E.; Korschineck, I. *Forensic Sci. Int.* 2006, *157*, 181.

(90) Juusola, J.; Ballantyne, J. *J. Forensic Sci.* 2007, *52*, 1252.

(91) Zubakov, D.; Hanekamp, E.; Kokshoorn, M.; van IJcken, W.; Kayser, M. *Int. J. Legal Med.* 2008, *122*, 135.

(92) Reich, K.; Independent Forensics, Inc.: USA, 2008; Vol. US Patent 0171399.

(93) Trombka, J. I.; Schweitzer, J.; Selavka, C.; Dale, M.; Gahn, N.; Floyd, S.; Marie, J.; Hobson, M.; Zeosky, J.; Martin, K.; McClannahan, T.; Solomon, P.; Gottschang, E. *Forens. Sci. Int.* 2002, *129*, 1.

(94) Chun-Yen Lin, A.; Hsieh, H. M.; Tsai, L. C.; Linacre, A.; Chun-I Lee, J. *J. Forensic Sci.* 2007, *52*, 1148.

(95) Haas, C.; Klesser, B.; Kratzer, A.; Bar, W. *Forensic Sci. Int. Gen.* 2008, *1*, 37.

(96) Pang, B. C. M.; Cheung, B. K. K. *Forens. Sci. Int.* 2007, *169*, 27.

(97) Sato, I.; Barni, F.; Yoshiiki, M.; Rapone, C.; Berti, A.; Nakaki, S.; Yamazaki, K.; Ishikawa, F.; Iwamoto, T. *Int. J. Legal Med.* 2007, *121*, 315.

(98) Fielder, A.; Rehdorf, J.; Hilbers, F.; Johrdan, L.; Stribl, C.; Benecke, M. *Open Forensic Sci. J.* 2008, *1*, 12.

(99) Quarino, L.; Dang, Q.; Hartmann, J.; Moynihan, N. *J. Forensic Sci.* 2005, *50*, 873.

(100) Reich, K. USA, 2007; Vol. 2007048084.

(101) Myers, J. R.; Adkins, W. K. *J. Forensic Sci.* 2008, *53*, 862.

(102) Pang, B. C.; Cheung, B. K. K. *J. Forensic Sci.* 2008, *53*, 1117.

(103) Soukos, N. S.; Crowley, K.; Bamberg, M. P.; Gillies, R.; Doukas, A. G.; Evans, R.; Kollias, N. *Forens. Sci. Int.* 2000, *114*, 133.

(104) Sakurada, K.; Motani, H.; Akutsu, T.; Ikegaya, H.; Iwase, H. *Can. Soc. Forensic Sci. J.* 2008, *41*, 13.

(105) Mann, C. K.; Vickers, T. J. In *Handbook of Raman Spectrscopy: From the Research Laboratory to the Process Line*; Lewis, I. R., Edwards, H. G. M., Eds.; Marcel Dekker, Inc.: New York, 2001, p 251.

(106) Williams, T. L.; Collette, T. W. In *Handbook of Raman spectroscopy: From the Research Laboratory to the Process Line.* ; Lewis, I. R., Edwards, H. G. M., Eds.; Marcel Dekker: New York 2001.

(107) Ali, E. M.; Edwards, H. G.; Hargreaves, M. D.; Scowen, I. J. *Analytica chimica acta* 2008, *615*, 63.

(108) Thomas, J.; Buzzini, P.; Massonnet, G.; Reedy, B.; Roux, C. *Forensic Sci Int* 2005, *152*, 189.

(109) Mazzella, W. D.; Buzzini, P. *Forensic Sci Int* 2005, *152*, 241.

(110) Tramini, P.; Bonnet, B.; Sabatier, R.; Maury, L. *Forensic Sci Int* 2001, *118*, 1.

(111) West, M. J.; Went, M. J. *Forensic Sci Int* 2008, *174*, 1.

(112) Eckenrode, A.; Bartick, E. G.; Harvey, S.; Vucelick, M. E.; Wright, B. W.; Huff, R. A. *Forensic Sci. Comm.* 2001, *3*.

(113) *Crime scene to court: The Essentials of Forensic Science*; Royal Society of Chemistry, The, 2010.

(114) Roux, C.; Maynard, P.; Dawson, M. *Chem Aust* 1999, *66*, 11.

(115) *Encyclopedia of Soil Science*; Springer-Verlag New York, LLC, 2008.

(116) Boyd, S.; Bertino, M. F.; Seashols, S. J. *Forensic Sci Int* 2011, *208*, 124.

(117) *Physical Evidence in Forensic Science*; 2 ed.; Henry C. Lee, H. A. H., Ed., 2000.

(118) Castelló, A.; Francés, F.; Verdú, F. *Talanta* 2009, *77*, 1555.

(119) Harris, K. A.; Thacker, C. R.; Ballard, D.; Court, D. S. *International Congress Series* 2006, *1288*, 589.

(120) Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, J. S. *PLS Toolbox 3.5 for use with Matlab*; Eigenvector Research Inc: Manson, 2005; Vol. 17.

(121) Malinowski, E. R. *Factor Analysis in Chemistry*; 3 ed.; A John Wiley & Sons, Inc.: New York, 2002.

(122) Spiro, T. G.; Strekas, T. C. *Proc Natl Acad Sci U S A* 1972, *69*, 2622.

(123) Hu, X.; Spiro, T. G. *Biochemistry* 1997, *36*, 15701.

(124) Dugaiczyk, A.; Law, S. W.; Dennison, O. E. *Proc Natl Acad Sci U S A* 1982, *79*, 71.

(125) Topilina, N. I.; Ermolenkov, V. V.; Higashiya, S.; Welch, J. T.; Lednev, I. K. *Biopolymers* 2007, *86*, 261.

(126) Tuma, R.; Prevelige, P. E., Jr.; Thomas, G. J., Jr. *Proc Natl Acad Sci U S A* 1998, *95*, 9885.

(127) Owen, D. H.; Katz, D. F. *J Androl* 2005, *26*, 459.

(128) Chen, M. C.; Lord, R. C.; Mendelsohn, R. *J Am Chem Soc* 1974, *96*, 3038.

(129) Remmele, R. L., Jr.; McMillan, P.; Bieber, A. *J Protein Chem* 1990, *9*, 475.

(130) Stein, E. A.; Junge, J. M.; Fischer, E. H. *J Biol Chem* 1960, *235*, 371.

(131) Bianchetta, J. D.; Bidaud, J.; Guidoni, A. A.; Bonicel, J. J.; Rovery, M. *Eur J Biochem* 1979, *97*, 395.

(132) Sikirzhytskaya, A.; Sikirzhytski, V.; Lednev, I. K. *Forens Sci Int* 2011, *In press*.

(133) Mickelsen, O.; Keys, A. *J. Biol. Chem* 1943, *149*, 479.

(134) Keuleers, R.; Desseyn, H. O.; Rousseau, B.; Van Alsenoy, C. *J Phys Chem A* 1999, *103*, 4621.

(135) Pecul, M.; Rizzo, A.; Leszczynski, J. *J Phys Chem A* 2002, *106*, 11008.

(136) Patrick, K. L.; Luz, P. M.; Ruan, J. P.; Shi, H.; Ullu, E.; Tschudi, C. *Mol Microbiol* 2008, *67*, 435.

(137) Socrates, G. *Infrared and Raman characteristic group frequencies: Tables and charts*; 3rd ed.; Johh Wiley & Sons Ltd: Chichester, 2001.

(138) Rajan, N.; Cao, Q.; Anderson, B. E.; Pruden, D. L.; Sensibar, J.; Duncan, J. L.; Schaeffer, A. J. *Infect Immun* 1999, *67*, 5027.

(139) Wiberg-Itzel, E.; Pettersson, H.; Cnattingius, S.; Nordstrom, L. *BJOG* 2006, *113*, 1426.

(140) Caspers, P. J.; Lucassen, G. W.; Puppels, G. J. *Biophys J* 2003, *85*, 572.

(141) Bell, W.; Booksh, K.; Myrick, M. L. *Anal. Chem* 1998, *70*, 332.

(142) Virkler, K.; Lednev, I. K. *Forensic Sci Int* 2008, *181*, e1.

(143) Virkler, K.; Lednev, I. K. *Anal Bioanal Chem* 2010, *396*, 525.

(144) Shaller, J.; Gerber, S.; Kampfer, U.; Lejon, S.; Trachsel, C. *Human blood plasma proteins: structure and function*; John Wiley & Sons Ltd.: Chichester, 2008.

(145) Mikhonin, A. V.; Bykov, S. V.; Myshakina, N. S.; Asher, S. A. *J Phys Chem B* 2006, *110*, 1928.

(146) Berjot, M.; Marx, J.; Alix, A. J. P. *J Raman Spectrosc* 2005, *18*, 289.

(147) Liang, J.; Cheng, Y.; Han, H. *J Mol Struct* 2008, *892*, 116.

(148) Ivanov, A. I.; Zhbankov, R. G.; Korolenko, E. A.; Korolik, E. V.; Meleshchenko, L. A.; Marchewka, M.; Ratajczak, H. *J Appl Spectrosc* 1994, *60*, 305.

(149) Edsall, J. T. *J Am Chem Soc* 1943, *65*, 1767.

(150) Sigma-Aldrich Co: 2009.

(151) Sigma-Aldrich Co: 2009.

(152) Sigma-Aldrich Co: 2009.

(153) Zhao, J.; Lui, H.; McLean, D. I.; Zeng, H. *Appl. Spectrosc.* 2007, *61*, 1225.

(154) Vitek, P.; Edwards, H. G.; Jehlicka, J.; Ascaso, C.; De los Rios, A.; Valea, S.; Jorge-Villar, S. E.; Davila, A. F.; Wierzchos, J. *Philos Transact A Math Phys Eng Sci* 2010, *368*, 3205.

(155) Jorge-Villar, S. E.; Benning, L. G.; Edwards, H. G. *Geochem Trans* 2007, *8*, 8.

(156) Edwards, H. G.; Wilson, A. S.; Nik Hassan, N. F.; Davidson, A.; Burnett, A. *Anal Bioanal Chem* 2007, *387*, 821.

(157) Sikirzhytski, V.; Sikirzhytskaya, A.; Lednev, I. K. *Appl Spectrosc* 2011, *65*, 1223.

(158) Dutta, P. K.; Gallagher, P. K.; Twu, J. *Chem. Mater.* 1992, *4*, 847.

(159) Moros, J.; Lorenzo, J. A.; Lucena, P.; Tobaria, L. M.; Laserna, J. J. *Anal. Chem* 2010, *82*, 1389.

(160) Virkler, K.; Lednev, I. K. *Anal Bioanal Chem* 2009, *396*, 525.

VI. Dissemination of Research Findings

1) Sikirzhytskaya A.; Sikirzhytski, V.; McLaughlin, G.; Lednev, I.K. Forensic identification of blood in the presence of contaminations using Raman microspectroscopy coupled with advanced statistics: effect of sand, dust, and soil, *J Forensic Sci*., In preparation.
2) McLaughlin, G.; Sikirzhytski, V.; Lednev, I.K. Forensic identification of blood: effect of substrates, *Forensic Sci. Int.*, In preparation
3) Sikirzhytski, V.; Sikirzhytskaya A.; Lednev, I.K. Non-destructive, confirmatory, on-field identification of body fluid stains using Raman spectroscopy, *Analyst*, in preparation.
4) Sikirzhytski, V.; Sikirzhytskaya A.; Lednev, I.K. Multidimensional Raman spectroscopic signatures as a tool for forensic identification of body fluid traces: a review, *Applied Spectroscopy*, Anal. Chem 2011, 65(11), 1223-1232.
5) Sikirzhytski, V.; Sikirzhytskaya A.; Lednev, I.K. Multidimensional Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification, *Anal. Chim. Acta 2012, 718(9), 78-83.*
6) Sikirzhytskaya A.; Sikirzhytski, V.; Lednev, Raman spectroscopic signature of vaginal fluid and its potential application to forensic body fluid identification, *Forensic Sci. Int. 2012*, 216(1-3), 44-48.
7) Sikirzhytski, V.; Sikirzhytskaya A.; Lednev, I.K. Advanced statistical analysis of Raman spectroscopic data for the identification of body fluid traces: semen and blood mixtures, *Forensic Sci. Int.*, 2012, submitted.

8) Ozek, N.S.; Afnan, A.; Bueno, J,; Mclaughlin, G.; Ralbovsky, M,; Sikirzhytskaya A.; Sikirzhytski, V.; Severcan, F.; Lednev, I. K. Forensic Applications of Vibrational Spectroscopy on Screening and Characterization of Tissues. Invited chapter for book Application of Vibrational Spectroscopy in Diagnosis and Screening. In press.
9) Sikirzhytski, V.; Virkler, K; Lednev, I.K. Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes, Sensors 2010, 10(4), 2869-2884.
10) Virkler, K.; Lednev, I.K. Raman spectroscopic signature of blood and its potential application to forensic body fluid identification, Anal. Bioanal. Chem. 2010, 396 (1) 525 -524.
11) Virkler, K.; Lednev, I.K. Forensic Body Fluid Identification: The Raman Spectroscopic Signature of Saliva, Analyst, 2010, 135, 512-517. Made a cover.
12) Virkler, K.; Lednev, I.K. Raman spectroscopic signature of semen and its potential application to forensic body fluid identification, Forensic Sci. Int. 2009, 193, 56–62.
13) Virkler, K.; Lednev, I.K. Blood Species Identification for Forensic Purposes Using Raman Spectroscopy Combined with Advanced Statistical Analysis, Anal. Chem. 2009, 81 (18), 7773-7777. Made a news story
14) Virkler, K.; Lednev, I.K. Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids, Forensic Sci. Int. 2008, (Rapid Communication) Published on web: doi:10.1016/j.forsciint.2008.08.004
15) V. Sikirzhytski, Igor K. Lednev, Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes, XXII International Conference of Raman Spectroscopy., August 8-13, 2010, Boston, oral presentation.
16) V. Sikirzhytski, Sikirzhytskaya A.; Igor K. Lednev, Forensic applications of Raman spectroscopy, Life science Research Symposium II, University at Albany, SUNY, September 17, 2010, Albany, oral presentation.
17) A. Sikirzhytskaya, V. Sikirzhytski, I. K. Lednev, Application of Raman Spectroscopy for an Easy-to-Use, on-Field, Rapid, Nondestructive, Confirmatory Identification of Body Fluids,Second International Conference on Vibrational Optical Activity (VOA-2) and Bio-Medical Applications of Raman Spectroscopy (BMARS), August 5-7, 2010, Albany, poster.
18) V. Sikirzhytski, A. Sikirzhytskaya, I. K. Lednev, Forensic Applications of NIR Raman Spectroscopy: automatic, on-field, rapid and non-destructive identification of human body fluid

stains, Third Annual Summer Research Event, University at Albany, SUNY, August 25, 2010, Albany, poster.

19) A. Sikirzhytskaya, K. Virkler, V. Sikirzhytski, I. K. Lednev, Raman spectroscopic signatures of human body fluids and its potential application to forensic body fluid identification, Third Annual Summer Research Event, University at Albany, SUNY, August 25, 2010, Albany, poster.
20) A. Sikirzhytskaya, V. Sikirzhytski, I. K. Lednev, Identification of body fluids by Raman spectroscopy, Life science Research Symposium II, University at Albany, SUNY, September 17, 2010, Albany, poster.
21) V. Sikirzhytski, A. Sikirzhytskaya, I. K. Lednev, Body fluid identification using Near IR Raman spectroscopy, PITTCON Conference&Expo, Georgia World Congress Center, March 13-18, 2011, Atlanta, poster.
22) A. Sikirzhytskaya, V. Sikirzhytski, I. K. Lednev, Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification, PITTCON Conference&Expo, Georgia World Congress Center, March 13-18, 2011, Atlanta, poster.