Forensic Science International 222 (2012) 259–265

Contents lists available at SciVerse ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Advanced statistical analysis of Raman spectroscopic data for the identification of body fluid traces: Semen and blood mixtures

Vitali Sikirzhytski, Aliaksandra Sikirzhytskaya, Igor K. Lednev *

Department of Chemistry, University at Albany, SUNY, 1400 Washington Ave., Albany, NY 12222, United States

ARTICLE INFO

Article history:
Received 24 September 2011
Received in revised form 7 June 2012
Accepted 3 July 2012
Available online 22 July 2012

Keywords:
Raman spectroscopy
Blood
Semen
Statistical analysis
Forensic science
Body fluid mixtures

ABSTRACT

Conventional confirmatory biochemical tests used in the forensic analysis of body fluid traces found at a crime scene are destructive and not universal. Recently, we reported on the application of near-infrared (NIR) Raman microspectroscopy for non-destructive confirmatory identification of pure blood, saliva, semen, vaginal fluid and sweat. Here we expand the method to include dry mixtures of semen and blood. A classification algorithm was developed for differentiating pure body fluids and their mixtures. The classification methodology is based on an effective combination of Support Vector Machine (SVM) regression (data selection) and SVM Discriminant Analysis of preprocessed experimental Raman spectra collected using an automatic mapping of the sample. This extensive cross-validation of the obtained results demonstrated that the detection limit of the minor contributor is as low as a few percent. The developed methodology can be further expanded to any binary mixture of complex solutions, including but not limited to mixtures of other body fluids.

1. Introduction

Forensic chemistry is an important branch of analytical chemistry that uses the concepts and techniques of chemistry to search for efficient methods of crime scene investigation [1]. Body fluids are a common type of forensic evidence in sexual assault cases, which are among the most difficult crimes to solve [2,3]. A number of recently developed methods to characterize body fluids have been accepted by forensic laboratories [4–23]. Identification of the type of human body fluid, mixtures of such fluids, and even the individuals involved in a crime is now possible. But the destructive character, the extensive sample pre-treatment, the need for expensive chemicals and equipment, and the time-consuming nature of the process are major weaknesses of the conventional methods. Furthermore, the traditional approaches may be significantly impaired by the mixing of body fluids, contamination, and the aging of stains. Here we address the problem of identifying mixed traces of semen and blood as well as the problem of detecting minor contributions and discriminating between pure fluids and their mixtures. The proposed methodology is based on a combination of Raman spectroscopy and advanced statistical analysis, and this method can be extended to characterize various complex mixtures, including but not limited to mixtures of other body fluids. Raman spectroscopy is a rapid, nondestructive and easy-to-use method for the comprehensive characterization of matter. Analysis can be performed with only several femtoliters or picograms of a sample and does not require any special pretreatment [24,25]. The application of the mapping and imaging techniques allows one to obtain detailed spectral information from every point of the sample and to build a complete data image of the studied matter. Moreover, the analysis of the material can be carried out with ~1 μm or less spatial resolution [26].

Recent advances in the field of multivariate calibration methods have led to dramatic increases in the number of their practical applications. Multivariate methods can improve sensitivity and selectivity of the identification when detecting the components of a specific mixture or extracting the information hidden in complex spectra by regressing analyte concentrations on multiple variables simultaneously, as opposed to univariate methods, which only monitor the variation of a single variable [27]. Among the main focuses of multivariate analysis are the nonlinear effects presented in the relationships between the concentration of the mixture component and the mixture spectra, the sample heterogeneity, and the sample-to-sample variability. For example, the application of multivariate methods is extremely effective in the Raman spectroscopic characterization of biomaterials, drugs, food, polymeric materials [26]. These Raman spectra are often characterized by strong fluorescence, overlapping spectral bangs, and nonlinearity in spectral responses because of absorbing components and molecular interactions between components, sample matrix effects, and intrinsic heterogeneity, respectively [24–26]. Here we propose a multistep chemometrical procedure that facilitates

* Corresponding author. Tel.: +1 518 591 8863; fax: +1 518 442 3462.
E-mail address: lednev@albany.edu (I.K. Lednev).

0379-0738/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.forsciint.2012.07.002

the discrimination between pure fluids and their mixtures and the detection of minor mixture contributors. Using Raman spectroscopy, we were able to detect the contributions of blood in semen and of semen in blood stains at all the tested concentrations (5% blood in semen and ~1% semen in blood stains are the smallest concentrations of the minority contributors used in our experiments). Smaller concentrations can be detected, but the probability of mixture detection at smaller concentrations is low. At the reported concentrations of 5% blood in semen and ~1% semen in blood, only a part of spectra was distinguishable from the pure fluids. Our approach is based on the effective combination of regression and classification methods. Regression analysis assists with the selection of binary mixture Raman spectra that have distinct characteristics of both ingredients and can be easily distinguished from the spectra of pure body fluids. Discriminant Analysis based on the classes selected by regression significantly outperforms the direct application of regression or calibration methods alone. For the current study, calibration and classification were obtained using Support Vector Machines (SVM), a chemometrical method invented by Vapnik that can deal with ill-posed calibration problems and produce robust models in the case of spectral variations due to nonlinear interference [28]. SVM has already garnered a strong reputation in the characterization of a wide variety of objects, including biomaterials and forensic evidence [29–32].

The mixture of blood and semen was chosen because of the forensic community's significant interest in its identification. If a victim of sexual assault cannot provide testimony, the ability to accurately reconstruct the crime scene based on the collected evidence is all the more important. Forensic investigators usually have to deal with body fluids and other types of evidence, which are considered important sources of DNA and can be used for serological and toxicological analysis. If a body fluid stain is detected, it is essential to know whether the stain comprises a mixture of fluids. Such a determination becomes more challenging if the crime was not reported for days or even years after the incident. The possibility of body fluid stain detection and identification at a crime scene will be greatly appreciated by the forensic community.

There are several fluorescent tests already available for the detection of biological fluid stains; however, such tests can provide false-positive results when distinguishing between different types of body fluids [4–6], and forensic investigations require a comprehensive analysis that yields trusted results. Laurell immunoelectrophoretic separation of seminal and vaginal acid phosphatases (SAP and VAP) was among the first methods of semen and vaginal fluid mixture characterization [8]. The electrophoretic separation of the spermatozoa-specific lactate dehydrogenase (LDH) isoenzyme was used by Mokashi et al. to differentiate semen from blood and vaginal secretions [9]. Whitehead and co-workers described the technique based on amylase activity, presented as amylase-sensitive test-paper, which was able to detect saliva traces in blood or semen stains [10]. RNA and DNA profiling are the most common methods used for characterization of body fluid stains [1,11–23]. Much attention has been paid to the application of mRNA profiling for body fluid identification [17–23]. Fleming and Harbison recently developed a multiplex polymerase chain reaction (PCR) system, which can identify blood, saliva, semen and menstrual blood in individual stains or in mixtures of body fluids using messenger RNA (mRNA) [14]. Furthermore, they proposed vaginal-specific bacteria (*Lactobacillus crispatus* and *Lactobacillus gasseri*) as promising new markers for the forensic identification of vaginal secretions using a mRNA multiplex assay [15].

Despite recent progress, the development of non-destructive, easy-to-use, fast, and inexpensive methods for the identification of body fluid composition is still anticipated by modern forensic science. We recently reported on the potential application of Raman spectroscopy to identify pure body fluid traces [33–40]. Other research groups have used Raman spectroscopy for analysis of fibers [41], drugs [42,43], and lipsticks [44], as well as ink [45], paint [45], explosives [46–49], bones [50], fingerprints [43,51,52] and condom lubricants [53]. Here we report the application of Raman spectroscopy in the analysis of blood/semen mixtures. The high selectivity and specificity of Raman spectroscopy to chemical and biochemical species coupled with advanced statistics allowed for the identification of pure blood and semen as well as their mixtures. The nondestructive character, high specificity and ability to extract information from small amounts of evidence are the main advantages of this method. The development of portable Raman instruments [54,55] may make it possible to bring these proposed methods directly to the crime scene [48].

2. Materials and methods

2.1. Sample preparation and Raman microspectroscopy

We have already reported the multidimensional Raman spectroscopic signatures of pure blood and semen [35,37]. If two body fluids are not thoroughly mixed and a dry sample contains small spots of pure fluids, the multidimensional Raman spectroscopic signatures can be used to identify the fluids. In the case of thoroughly mixed body fluids, a more complex approach may be required. Therefore, in this study, we focus only on samples prepared by mixing blood and semen thoroughly.

Blood and semen samples were purchased from several companies including Bioreclamation, Inc., Lee Biosolutions, Inc., and Biological Specialty Corp. Blood/semen stains were prepared using samples from two different anonymous individuals (a Caucasian male for semen and a Caucasian female for blood). Both donors were found to be negative for HbsAg, HCV, HIV-1&2, syphilis and HIV-1 antigen. All samples, in volumes of 10 μL each, were placed on microscope slides that were covered with aluminum foil to reduce fluorescence. Mixtures were prepared with different blood/semen ratios (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, 96.875:3.125 and 98.437:1.5625) by thoroughly shaking for 20 s. All samples were allowed to dry completely overnight. A Renishaw inVia confocal Raman spectrometer equipped with a research-grade Leica microscope and 50× long-range objective was used for the spectra acquisition. Raman spectra with 785-nm excitation were measured from 108 points using automatic mapping (Renishaw PRIOR automatic stage) from a sample area of 3.5 × 2.5 mm with a 10-s acquisition at each point. The laser power used on the dried samples was approximately 10 mW, and the spot size of the excitation beam was approximately 10 μm wide for the standard confocal mode. The spectral resolution was 0.8 cm^{-1}, and the CCD camera was calibrated using a silicon standard.

2.2. Data treatment

The Raman spectra obtained using the automatic mapping of all samples were first processed using the GRAMS/AI 7.01 software. After the cosmic ray interference removal, the Raman spectra were imported into MATLAB 7.4.0 for statistical analysis [56]. Normalization to the total area was performed to take into account the varying amount of background interference and the total offset variation. Because the visual inspection revealed the presence of a complex and varying fluorescent background, an adaptive iteratively reweighted penalized least squares (airPLS) algorithm [57] was used for baseline correction. The normalized and corrected spectra were subjected to dimension reduction by principal component analysis (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 [58]. PCA scores were used for SVM regression by compiling the experimental data in to 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 (see Section 3 for details). 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.

3. Results and discussion

Stains of body fluids mixtures are highly heterogeneous [39]. A single stain may have areas with variable composition ranging from practically pure semen to nearly pure blood. Fig. 1 shows selected raw Raman spectra acquired from pure blood and semen