Forensic Chemistry 1 (2016) 31–38

Contents lists available at ScienceDirect

Forensic Chemistry

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

Full length articles

Forensic body fluid identification and differentiation by Raman spectroscopy

Claire K. Muro, Kyle C. Doty, Luciana de Souza Fernandes, Igor K. Lednev *

Chemistry Department, University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 16 June 2016
Accepted 27 June 2016
Available online 30 June 2016

Keywords:
Raman spectroscopy
Body fluid
Chemometrics
Classification model
Variable selection

ABSTRACT

The ability to identify body fluid traces at crime scenes, and preserve any DNA present, is critically important in forensic science. Identification can be difficult because many of the current techniques are specific to one body fluid, and typical biochemical methods are destructive – preventing any further analysis. To develop a universal, confirmatory, nondestructive, approach that can be used to differentiate and identify body fluids, we combined the specificity of Raman spectroscopy with the analytical power of statistical modeling. Raman spectra were collected from 75 body fluid samples, including peripheral blood, saliva, semen, sweat, and vaginal fluid. After preprocessing, samples were split into calibration and validation datasets. Several chemometric analysis techniques were trained and tested to find the best model. Combining classification modeling with variable selection resulted in a single, robust, technique. This enhanced model accurately predicted the identity of 99.9% of the spectra from the calibration dataset, after cross-validation. More importantly, it correctly predicted the identity of 100% of the spectra in the external validation dataset. All five body fluids were successfully discriminated by coupling Raman spectroscopy and chemometrics. This technique is both reliable and nondestructive, offering substantial advantages over the current techniques used to identify body fluids.

1. Introduction

Traces of body fluids can be discovered in a variety of contexts at a crime scene [1]. The type of body fluid that contributed to a specific stain can provide contextual information to investigators and determine the stain's relevancy in the case. Body fluids are also important because they can be a source of DNA evidence. Along with fingerprints, DNA is one of the few forms of physical evidence that can conclusively identify an individual [2]. Additionally, it is critically important for forensic scientists to be able to determine the source of a DNA sample, so that the individual identified by the DNA profile can be linked to a specific piece of evidence.

Currently, several tests can be employed in the field to identify body fluids. These tests can be categorized as either presumptive or confirmatory, depending on whether their results are likely- but indefinite- or conclusive. While presumptive tests are useful, their results must be supported by further testing because they can produce false positives [3–9]. The two main issues accompanied by many of the current methods are that they are body fluid specific and destructive to the sample. Most of the tests currently in use, such as luminol, starch-iodine, and the Christmas tree stain, are specific to one body fluid, and will only react in the presence of that particular fluid. Tests that can be used on more than one type of sample, such as alternate light sources, often produce non-discriminating responses to different body fluids. Additionally, many biochemical tests damage the sample by consuming it through a chemical reaction or exposure to radiation. At this time, there is no universal, nondestructive, approach that can be used to differentiate body fluids.

Characterizing potential body fluid traces can be difficult for several reasons. Body fluids are heterogeneous in nature, especially when dried [10]. They are complex biochemical mixtures with considerable intra-sample variation. Moreover, body fluids share many individual components, resulting in inter-sample similarities [11]. A comprehensive and selective approach is needed to overcome both of these obstacles.

Several new developments in analytical chemistry and spectroscopy have been recently utilized for forensic purposes [3,12,13]. One such technique is the ability to visualize blood on fabrics, after diluting it 1:100, using thermal IR imaging [14]. By varying fabrics and the spectral region used for thermal IR imaging, the limits of detection for blood on acrylic, cotton, and polyester were estimated to be 2300, 610, and 900 times diluted,

* Corresponding author.
E-mail address: ilednev@albany.edu (I.K. Lednev).

http://dx.doi.org/10.1016/j.forc.2016.06.003
2468-1709/

respectively [15]. This work was expanded, and it was found that traditional thermal imaging could be used to detect blood on fabrics, after diluting it 1:1000, and exposing the fabrics to steam [16]. X-ray fluorescence can presumptively detect semen and blood based on their elemental compositions [17]. Orphanou used attenuated total reflectance Fourier transform infrared spectroscopy to visually differentiate between peripheral blood, saliva, semen, and vaginal fluid [18]. However, this particular study did not include sweat, nor did they propose a quantitative method for differentiation. ^{1}H NMR has been shown to discriminate between serum, saliva, seminal fluid, and urine with the use of principal component analysis [9]. Yet, this study did not include sweat or vaginal fluid, and used serum and seminal fluid in place of whole blood and semen.

Raman spectroscopy, in particular, has proven to be a very promising analytical technique for several forensic applications [19–21]. Raman mapping can be used to probe across a sample's surface, instead of a single point. Meanwhile, the selective nature of Raman spectroscopy enables it to discriminate between chemically analogous species [11]. Its power is further enhanced when chemometric analyses are applied to spectroscopic datasets [19]. Consequently, Raman spectroscopy has been used to study pure body fluids [22–26], cells [27], mixtures [28], and contaminated traces [29]. Raman spectroscopy can also discriminate between human and non-human animal blood [30–32], as well as peripheral and menstrual blood [33]. Sikirzhytski et al. reported the first statistical model for differentiating peripheral blood, semen, and saliva based on their Raman spectra, and demonstrated great potential for the development of a "universal" forensic method for body fluid identification [34].

The objective of this study was to construct a new method to automatically discriminate between human body fluids in a single step. To develop a universal, nondestructive, noncontact, approach that can be used to differentiate and identify the five main body fluids, we combined the specificity of Raman spectroscopy with the analytical power of statistical modeling. Peripheral blood, saliva, semen, sweat, and vaginal fluid samples were analyzed by Raman spectroscopic mapping to account for intra-sample variations. Classification models using the entire spectral range collected during sample analysis were built using two different approaches: Partial Least Squares Discriminant Analysis (PLSDA) and Support Vector Machine Discriminant Analysis (SVMDA). Next, interval PLSDA (iPLSDA) and Genetic Algorithm (GA) were independently employed prior to PLSDA and SVMDA modeling to determine whether initial variable selection might enhance differentiation. All of the models were externally validated with samples that were excluded from model calibration. The final classification model performed with 100% accuracy during external validation. While this study uses "ideal" laboratory conditions, the method can be expanded to contaminated samples [29] and biological stains on common substrates [35,36].

2. Materials and methods

2.1. Sample analysis

A total of 75 samples were purchased from Bioreclamation IVT, Inc. (Westbury, NY) and Lee Biosolutions, Inc. (Maryland Heights, MO). The sample population included peripheral blood, saliva, semen, sweat, and vaginal fluid donors (n = 15 each). Peripheral blood samples were prepared in 30 μL aliquots, while saliva, semen, sweat, and vaginal fluid were all prepared with only 10 μL. Samples were deposited onto individual microscope slides, which had been covered with aluminum foil to avoid fluorescence interference [37], and allowed to dry completely. The dried traces were approximately 15 mm^2 in area. Spectra were collected using an inVia Raman spectrometer (Renishaw, Inc., Hoffman Estates, IL) operated with WiRE 3.2 software. All samples were excited with a 785 nm wavelength laser and mapped using a PRIOR automatic mapping stage and a 50× microscope objective (except peripheral blood, which was collected with a 20× objective). The sampling area for each point is controlled by the numerical aperture of the microscope objective used (20× = 5.88 μm^2, 50× = 2.88 μm^2). Laser power varied from 4 mW (peripheral blood), to 65 mW (semen), and to 130 mW (saliva, sweat, vaginal fluid) to assure sample integrity and a high signal-to-noise ratio in the acquired Raman spectra.

Each sample was mapped to collect several spectra at various points across the surface of the traces. The points were selected by automatic mapping in a grid-like fashion, except for vaginal fluid samples. These samples had such a low concentration of biochemical material that many points selected by automatic mapping would have irradiated the bare aluminum foil substrate instead of the dried sample. Consequently, the optical image was used to manually select points for mapping in order to ensure that the spectra were in fact acquired from the physical sample. The acquisition parameters selected were based on prior studies of individual body fluids [22–26]. The number of points collected per map varied from one body fluid to another (peripheral blood = 35, saliva = 25, semen = 64, sweat = 115, vaginal fluid = 15). The number of spectra accumulated at each point in the maps ranged from 1 (saliva and vaginal fluid) to 20 (peripheral blood), with some body fluids in between (semen = 7 accumulations, sweat = 3 accumulations). The acquisition time was set to 10 s for all body fluids, except saliva and vaginal fluid (30 s).

2.2. Data analysis

The spectral data collected were formatted and preprocessed prior to being used for modeling. A single dataset containing all experimental spectra (n = 3926) was created in MATLAB version 2012b (MathWorks, Inc., Natick, MA). PLS_Toolbox version 8.0.2 (Eigenvector Research, Inc., Wenatchee, WA) was used for analysis and processing. The spectra were baseline corrected using automatic weighted least squares (polynomial order: 5) [38] and normalized by total area. The preprocessed dataset was then split into training and testing data. Fifteen donors (three from each body fluid) were randomly chosen for validation, and the remaining 60 donors were used for calibration. All models were trained using only the spectra included in the calibration dataset. The spectra selected for validation were set aside until the models were built, after which the body fluid class for each spectrum in the validation dataset was predicted by the trained models.

All models were set to mean center the calibration dataset prior to building. Two classification models were built using the entire calibration dataset. First, a PLSDA model was constructed with five latent variables and venetian blinds cross-validation. Venetian blinds split the calibration dataset into n subsets, with one subset used for validation after $n - 1$ subsets are used for calibration. This process is repeated until each subject has been used for cross-validation. The model was then externally validated with the 15 donors reserved for testing. Second, an SVMDA model was built. The X-block data were compressed by partial least squares with five latent variables, and cross-validated by venetian blinds. This second model was also externally validated.

Lastly, iPLSDA and GA were applied to the calibration dataset for variable selection. For iPLSDA, the model was executed in forward mode and instructed to automatically choose the interval step size and the number of intervals (each set to be 30 variables wide). For the GA, the population size was set to 32, the window width to 25 variables, and 30% of the variables (terms) were included in the initial variable subsets. Subsequent breeding was set to double cross over, with the default mutation rate of 0.005, and allowed to continue for a maximum of 100 generations. Partial Least Squares (PLS) models were selected to automatically evaluate the GA, using a maximum of 5 latent variables. These models were cross-validated by contiguous blocks with 5 splits. The entire GA was set to repeat 10 times. The original PLSDA and SVMDA models were rebuilt, first using the variables selected by iPLSDA, and then with those selected by GA. These four additional models were all tested with the external validation dataset.

3. Theory

Chemometrics, the application of advanced statistics to chemical systems, has been used to solve several other complex problems in spectroscopy [39]. Readers who are not familiar with the field of chemometrics are directed to Wold et al. [40]. One of the main benefits of using chemometric models, even when they are not necessarily needed for differentiation, is the ability to quantitatively and conclusively interpret spectral data. Visual analysis of spectra with the naked eye is ultimately qualitative in nature, and can introduce human error or bias. Conversely, statistical models provide objective results, which are often accompanied by confidence intervals. This particular study required a classification model to group spectra according to their class, or body fluid. Two classification schemes were explored with the dataset: PLSDA and SVMDA.