Determining donor age through blood analysis using attenuated total reflection fourier transform-infrared (ATR FT-IR) spectroscopy for forensic purposes

Samantha Leigh Giuliano
*University at Albany, State University of New York*, sgiuliano@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: [https://scholarsarchive.library.albany.edu/legacy-etd](https://scholarsarchive.library.albany.edu/legacy-etd)

Part of the Chemistry Commons

**Recommended Citation**


[https://scholarsarchive.library.albany.edu/legacy-etd/2480](https://scholarsarchive.library.albany.edu/legacy-etd/2480)

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
DETERMINING DONOR AGE THROUGH BLOOD ANALYSIS USING ATTENUATED TOTAL REFLECTION FOURIER TRANSFORM-INFRARED (ATR FT-IR) SPECTROSCOPY FOR FORENSIC PURPOSES

by

Samantha L. Giuliano

A Thesis
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of
the Requirements for the Degree of
Master of Science

College of Arts & Sciences
Department of Chemistry
Spring 2020
Chapter 1 of this thesis is an overview of the use of attenuated total reflection Fourier transform-infrared (ATR FT-IR) spectroscopy in industry and for examination of body fluids with emphasis on examination of blood. It covers recent developments for analysis of samples with ATR FT-IR spectroscopy for forensic purposes.

Chapter 2 of this thesis discusses that results of the experiment conducted using ATR FT-IR spectroscopy as a method to determine donor age. As a part of forensic practices, phenotype profiling is beneficial to help narrow down suspects. The goal of this study is to identify a person's age range using dried bloodstains. Attenuated total reflection Fourier transform-infrared (ATR FT-IR) spectroscopy is the technique used to acquire information about the total (bio)chemical composition of a sample. For the purpose of this study, a diverse pool of donors including those in newborn (<1), adolescent (11-13), and adult (43-68) age ranges was used. Different donor age groups were found to have different levels of lipids, glucose, and proteins in whole blood, although the corresponding spectral differences were minor. Therefore, the collected data set was analyzed using chemometrics to enhance discrepancy and assist in donors’ classification. A partial least squares discriminant analysis (PLSDA) was used to classify ATR FT-IR spectra of blood from newborn, adolescent, and adult donors. The method showed 96% correct classification of spectra in leave-one-out cross-validation (LOOCV) of the model. Overall, ATR FT-IR spectroscopy is nondestructive and can be an infield method that can be used for the variety of forensic applications. In general, the developed approach combining ATR FT-IR spectroscopy and advanced statistics shows the great potential for classifying (bio)chemical samples exhibiting significant intra-class variations.
ACKNOWLEDGMENTS

I would like to start off by thanking everyone who has supported me on my academic journey my family, friends, advisors, and colleagues. It is because of your unwavering support and belief in me that I owe this honor of a master’s degree to.

First, I would like to express my deepest gratitude to my advisor, Professor Igor K. Lednev, for his guidance through my research within his lab. I am always grateful for the time Dr. Lednev was willing to put toward my research projects, papers, and academics. During my time in his lab, I have gained invaluable skills that will help me in career and make me a standout candidate. I cannot thank him enough for all his help and thoughtfulness he has but into making me a better scientist. Working with him is more than I could have ever dreamed for my college career and will be forever grateful to him for providing such an amazing opportunity to expand my skills and intelligence.

I would like to thank my committee member, Professor Jan Halámek, for his impact toward my graduate journey and my research work. I also would like to thank the Chemistry department at the University at Albany for being there through every step of my college education. I would like to note Brian Gabriel for his patience and help, answering each of my questions, and finding solutions to all my concerns. I would also like to thank the Chemistry Graduate Committee for making it possible for me to obtain my master’s degree.

I would like to express my greatest thanks to Ewelina Mistek without your help I would have never been able to be here. You have been such an important person in my research career at Albany. From training me on instruments and computer software, to answering every single one of my questions and reviewing countless numbers of papers for me. You have always been there to help me in any way you can, and I cannot express how much that has meant to me over
the past few years. In addition, I would like to thank all current and past members that have welcomed me into the Lednev lab and for their helpful advice and guidance.

I would like to thank my parents, Marlene and Steven Giuliano, and my sisters, Madison and Karissa. You all have been there through all my ups and downs. Whether it was calming me down about an exam, helping me find clarity in the mists of worrying, or helping me celebrate small milestones. You have always helped me find perspective and happiness during this journey. Thank you for visiting me, driving all the way to bring me home, and calling me every night. Lastly, I would like to thank all my friends I have made along my journey at UAlbany and the friends that I have had for year. You all have a special place in my heart and your word of encouragement have pushed me through every obstacle. Again, thank you for everyone whom I have meet along this journey I greatly appreciate all the guidance and support.
TABLE OF CONTENTS

CHAPTER 1: Review of Literature ................................................................. 1-4
  1.1 FT-IR Spectroscopy ........................................................................ 1
  1.2 FT-IR Spectroscopy in Industry ....................................................... 2
  1.3 FT-IR Spectroscopy for Blood Analysis ........................................... 3

CHAPTER 2: Research Studies ................................................................. 5-9
  2.1 Introduction ..................................................................................... 5
  2.2 Materials and Methods ................................................................. 8-9
    2.2.1 Blood Samples ........................................................................ 8
    2.2.2 Instrumentation and Spectral Collection .................................. 8
    2.2.3 Data Processing ..................................................................... 9
    2.2.4 Validation .............................................................................. 9
  2.3 Results and Discussion ................................................................. 9-17
    2.3.1 Spectral Analysis ................................................................... 9
    2.3.2 Statistical Modeling of Spectra ............................................... 13
    2.3.3 Model Validation .................................................................. 16
  2.4 Conclusions .................................................................................. 17-18

APPENDIX:

  Funding ................................................................................................ 18
  Permission .......................................................................................... 18
  References .......................................................................................... 18-20
CHAPTER 1: Review of Literature

1.1 FT-IR Spectroscopy

Spectroscopy is the study of how matter interacts with electromagnetic radiation. There are different forms of spectroscopy, such as absorption spectroscopy, which will be covered in this paper. Infrared (IR) spectroscopy is the analysis of IR light interacting with the molecule. Absorption, emission, and reflection are three different types of analysis that can be performed.[1] Upon irradiation with IR light, certain bonds respond by vibrating faster or slower which can be detected. It is then translated into what is called a spectrum, a visual representation of the vibrations.

The main uses are to study organic and inorganic chemistry. When a chemical sample is exposed to the action of IR light, it can absorb some frequencies, transmit unabsorbed light and reflect it back onto the source. The detector recognizes the transmitted frequencies, and can produce values of the absorbed frequencies.[2] This does this by determining functional groups in a molecule based on their vibrational modes. The energy in these vibrational modes are determined by length of the bond and the mass of the atom classifying the functional group. Dipole moments are an important part of determining the length of a bond and if it will be detected by the IR instrument. A dipole moment is a separation of charge between atoms in a molecule and it occurs when electrons are shared unequally. This can produce the magnitude of the charge and the distance between the atoms. For a molecule to be IR active is must have a dipole moment. When a IR active molecule is irradiated with the light source, the dipole moment interacts. It is changed momentarily this change is produces a readable vibrational mode.[1]

These different vibrational modes can be detected using an IR spectrometer which produces an IR spectrum of the component. The near-IR region consists of wavenumbers
including 12,500-4,000 cm\(^{-1}\).[2] The mid-IR region includes 4,000-400 cm\(^{-1}\) and the far-IR region is 400-10 cm\(^{-1}\). The most used range is a range of 4000-400 cm\(^{-1}\). Atoms of each molecule are bound and these bonds vibrate in different IR regions, what is then used for characterization purposes. The most informative region of IR spectrum is called a fingerprint region. It consists of wavenumbers between 1500 and 500 cm\(^{-1}\) on the spectrum.

There are known wavenumber (cm\(^{-1}\)) values for absorbance of different functional groups. The assignments of peaks and bands allow for proper description and finding a structure of a molecule.[1] This is very useful in a multitude of fields. IR spectroscopy can identify an unknown substance, it can study the progress of a reaction, and it can detect impurities when tested against a standard sample. IR spectroscopy is also able to detect mixtures of different components. It is a very important technique because it is so versatile, it can be used with dried liquids, pastes, powder, fiber, and gases. It is fairly inexpensive compared to many analytical methods and it is easy to use.[3]

FT-IR stands for “Fourier transform-infrared” and it is the most common form of IR spectroscopy. FT-IR spectroscopy is the preferred method because it does not destroy the sample, it is faster than older models, it is more sensitive and precise, and it can also be portable. The FT-IR spectrometer uses an interferometer to control the wavelength from the IR source. A detector measures the intensity of the reflected light as a function of its wavelength. The signal obtained from the detector is an interferogram. This signal is then translated into readable spectrum using Fourier transform software.[2]

1.2 FT-IR Spectroscopy in Industry

FT-IR spectroscopy is used by many industries such as organic synthesis, biological research, the pharmaceutical, and food sciences. In biological and organic research, FT-IR
spectroscopy can monitor reactions of the amino acids, ligands, and proteins. Environmental scientist use FT-IR spectroscopy to analyze soil samples, and to monitor air and water quality. It can test contamination levels quickly and accurately in the field to try to combat the ever-growing environment crises.[4]

The food industry uses FT-IR spectroscopy to make sure that the required government standards are met. Food properties can be monitored allowing the production of food to run efficiently. Foodborne pathogens can also be identified by FT-IR spectroscopy.[5] In the forensics field, FT-IR spectroscopy is used to quickly identify illegal drugs and materials, test crime scene evidence, and counterfeit goods. This technique can provide fast, easy, and accurate results.[6]

1.2 FT-IR Spectroscopy for Blood Analysis

Body fluid analysis contributes significantly to scientific community. In particular, blood analysis is mainly used in medical and forensics fields. Blood can be used to find and test for a number of disorders and diseases. It is one of the most common body fluids to be tested in the medical field. IR spectroscopy has been used by many research groups to determine glucose concentration in blood and for diagnostic measures of cancer. As one of the recent developments in the medical field, IR spectroscopy has been used to study blood oxygenation in kidneys.[3, 7] Forensics is another major industry that investigates blood. Blood is one of the most common body fluids found at a crime scene and can be extremely helpful in solving the case. A major part of blood analysis is the examination of DNA but it does have disadvantages: it is destructive and can only provide a match if one is available.[8]

These drawbacks have pushed forensic scientists to start looking outside the box for solutions to make the most out of the blood evidence. Many laboratories are working on
identification and age estimation of bloodstains.[9] The estimation of how long blood has been present at a crime scene is being studied under different environmental conditions to see how it degrades over time. IR spectroscopy has allowed for the advancements in examination of body fluid traces for forensics. Research groups are working on a database library of body fluids, including blood, to allow for proper identification.[10] The Lednev group has also included phenotype profiling to a number of papers to allow for more identifiable information about a donor when a DNA match is not available. Thus far, species identification as well as biological sex and race determination models have been published.[11, 12] These models allow for evidence to be helpful in a multitude of ways to make conclusions as accurate and descriptive as possible.
CHAPTER 2: Research Studies

2.1 Introduction

Analytical methods are becoming extremely useful in the growing field of forensic science. A variety of emerging analytical and bioanalytical approaches, including spectroscopy, have been developed to revolutionize the world of forensics.\[13-15\] As analytical chemistry becomes more relied upon in court cases, the need arises for universal and statistically confident methods. Advanced statistics allows investigators to obtain more reliable results with reduced human bias, which is of high importance in the courtroom.\[16\]

Critical characteristics of the human phenotype profile can significantly narrow down the search for potential suspects or missing persons.\[17\] Biological traces can be recovered at a crime scene as evidence and can be very valuable for phenotype profiling. Phenotype profiling can categorize people into different groups based on their biological sex, race, and age.\[18\] For this study, age groups of donors will be the main focus. Body fluids that are typically found at crime scenes are blood, semen, sweat, saliva, urine, and vaginal fluid. The examination of these body fluids can help narrow down suspects. The presence of these fluids is very common; however, blood is the body fluid most frequently found at crime scenes.\[19\] The presence of blood results mostly from violence such as homicide, assaults, and robberies. Current forensics techniques, such as forensics serology and DNA profiling, have become very important. The most powerful forensic analysis is DNA analysis since it can identify a person. DNA profiling has a multitude of steps and needs time to process from start to finish.\[8\] The whole process could take days. With the limitations of current technology, there is a very large backlog at the crime laboratories and some evidence samples are never processed. To process the samples, extraction, quantification amplification, separation, analysis and then database searching need to
be completed. Currently, there is the database for DNA that is known as the combined DNA index system (CODIS). Samples from known and unknown criminals are stored to see if there is a match from any crimes. If there is no match, it is impossible to know much more about the criminal. This is a limitation for investigators; therefore, creating an analytical method such as human phenotype profiling by ATR FT-IR spectroscopy and chemometrics can alleviate this problem. This can provide invaluable information with statistical confidence even if there is no profile currently in the system.

Since blood is the most-commonly found body fluid at a crime scene, in this study, it will be tested and correlated with the chronological age of a donor. Forensic anthropologists can determine a person’s age by studying the human body or skeleton.[17] However, with the lack of a physical body or eyewitnesses, it is difficult to determine the age of a victim or their assailant. Donor biological sex, chronological age, and health status all have an effect on human blood.[17, 20] Levels of alkaline phosphatase (ALP) have been found to differ for young and old donors.[9] The comparison can be done using a biocatalytic assay. The major difference in donor age based on blood can be found in hemoglobin–newborn whole blood samples that contain fetal hemoglobin (HbF) which differs from adult hemoglobin (HbA).[21-23] The volume of red blood cells in blood is higher in newborns than in adults, and the structure of hemoglobin is slightly different. Both hemoglobin molecules, HbF and HbA, contain four subunits. HbA consists of $\alpha_2\beta_2$, while HbF is made of $\alpha_2\gamma_2$ which affects the amino acid composition of both molecules. This change in structure, in addition to many other discrepancies found between concentrations of components, such as other proteins, lipids, and carbohydrates, can be detected by the attenuated total reflection Fourier transform-infrared (ATR FT-IR) spectroscopy.
For different developmental processes, levels of the (bio)chemical components in blood can vary with the age of an individual.[24] This study focuses on chronological age of a donor compared to biological age because biological age can be greatly affected by health status, diet, and physical fitness.[17] To account for such discrepancies, only healthy donors from each category who showed normal growth rate specific to their age were selected for this proof-of-concept study. In general, it is expected that newborn and adolescent subjects have different (bio)chemical composition of blood than a fully developed adult. These quantitative differences can help distinguish between the age categories.

Recently, Doty and Lednev reported on determining the chronological age of donors based on Raman microspectroscopy of dry bloodstains.[17] They were able to conclude that there are differences between age groups and that Raman spectroscopy in combination with chemometrics was efficient in detecting these differences. In addition, Fujihara et al. conducted an experiment discriminating between infant and adult bloodstains using microRaman spectroscopy.[25] This method has advantages and disadvantages in the field.[26] The main limitation of near-IR Raman spectroscopy when used for blood analysis is that the 785-nm excitation Raman spectra of blood are dominated by the hemoglobin contribution with limited input from other components. To overcome this limitation, a different technique was used to probe the (bio)chemical composition of blood in this study—ATR FT-IR spectroscopy. ATR FT-IR spectroscopy is a vibrational spectroscopic technique that is complementary to Raman spectroscopy. This technique has already been used in the examination of body fluids and trace evidence in the forensics field.[10, 13] ATR FT-IR spectroscopy has been proven to successfully determine donor sex and race from a bloodstain.[12]
In this study, ATR FT-IR spectroscopy and chemometrics were used to analyze blood from 20 human donors and to differentiate between them based on their chronological age. The donors were split into three classes consisting of newborn, adolescent, and adult samples. The spectra variations between each group were studied using chemometrics, and a partial least squares discriminant analysis (PLSDA) model was constructed. For age determination, each donor was left out of the training model once and used for predictions in the model.

2.2 Materials and Methods

2.2.1 Blood Samples. The experiment was performed on blood samples obtained from 20 donors provided to us by BioIVT (Westbury, NY). Samples were divided into age classes of newborn (<1 year old), adolescent (11-13 years old), and adult (43-68 years old). The pool of donors included 5 newborns, 4 adolescents, and 11 adults. Each whole blood sample contained disodium EDTA. The samples were defrosted and vortexed before deposition. Samples were prepared by placing 30 µL of whole blood on a microscope slide and drying overnight.

2.2.2 Instrumentation and Spectra Collection. Spectra were recorded using a PerkinElmer Spectrum 100 FT-IR spectrometer (PerkinElmer, Inc., Waltham, MA) equipped with a diamond/ZnSe ATR crystal. Spectrum software version 6.0.2.0025 (PerkinElmer, Inc., Waltham, MA) was used for collecting spectra. The crystal was cleaned with water and acetone before each sample was placed on its surface. A solution of 10% bleach was used for safety purposes to clean the crystal after measurements. A background check was run prior to placing the sample on the crystal. Ten spectra were recorded from different spots of each sample in a spectral range of 4000-600 cm⁻¹. The spectral resolution was set to 4 cm⁻¹, and 3 scans were taken per spectrum.
2.2.3 Data Analysis. Data set preparation and statistical analysis were performed using PLS Toolbox (Eigenvector Research, Inc., Wenatchee, WA) operating in MATLAB R2017b (MathWorks, Inc., Natick, MA).[27] The spectral range of 2669-1711 cm\(^{-1}\) demonstrated enhanced contribution from the ATR crystal; therefore, this region was excluded from the analysis as shown in an earlier study provided by our group.[11] Truncated spectra were transformed into absorption by applying a log(1/T) function, smoothed and baseline corrected by a 1st order derivative, normalized by total area, and mean-centered. After these preprocessing steps, a genetic algorithm (GA) was used to select the most significant spectral regions for classifying the studied groups.[28] A supervised statistical method, PLSDA, was employed to distinguish between age groups of blood donors. The venetian blinds method of cross-validation (CV) with ten splits was used for internal validation.

2.2.4 Validation Tests. Because of the limited sample population size, the technique was validated by subject-wise leave-one-out cross-validation (LOOCV). The spectra from one donor were removed from the training data set, and the PLSDA model was refitted to the remaining training data. Removed spectra were then used for prediction purposes. This process was repeated until all subjects were removed and predicted. The subject-wise LOOCV results are reported as the performance over all test sets, providing an estimate of model performance and confirming the classification of predictions performed for this particular training data set.

2.3 Results and Discussion

2.3.1 Spectral Analysis. The goal of this study was to use ATR FT-IR spectroscopy combined with chemometrics to classify blood samples based on donor age and differentiate blood samples of newborn, adolescent, and adult donors. There were 20 donors and 200 spectra total to be classified. An average spectrum of each group of donors can be seen in Figure 1. The
10

infrared spectrum of biological samples illustrates signal from the following: lipids (3000-2800 cm$^{-1}$), proteins (1700-1500 cm$^{-1}$), nucleic acids (1250-1000 cm$^{-1}$), and carbohydrates (1000-800 cm$^{-1}$).[29, 30] Small discrepancies between average spectra of different classes can be seen in Figure 1. The average difference spectra between all three categories were calculated and plotted against the corresponding spectral standard deviations in Figure 2 and Figure 3. Most of the standard deviations were larger than the corresponding difference spectra that indicated that the average spectral differences are statistically insignificant. The peaks in the difference spectra that were larger than the standard deviation peaks were analyzed separately including those at 1538 cm$^{-1}$, 1659 cm$^{-1}$, 2871 cm$^{-1}$, and 2909 cm$^{-1}$ (Figure 3). It was concluded that the changes in these peak intensities are too small to make a reliable class identification. This result differs from Raman spectroscopic results reported earlier.[17] The conclusion was that at 375 cm$^{-1}$ there is a distinct statistical difference in the Raman peak intensity to differentiate newborns from adolescent and adult donors. The 375-cm$^{-1}$ peak is not evident in the ATR FT-IR spectrum of dried blood. Therefore, chemometrics was used to build a statistical model and classify the samples.
Figure 1. Averaged ATR FT-IR spectra of blood from each age group: newborn (red), adolescent (green), and adult (blue).
Figure 2. Raw ATR FT-IR spectra of blood standard deviations (SDs) for newborn vs. adolescent (a), newborn vs. adult (b), and adolescent vs. adult (c) compared against the differences between each age group.
Figure 3. Plots with mean ±1 standard deviation for the following peaks 1538 cm\(^{-1}\) (a), 1659 cm\(^{-1}\) (b), 2871 cm\(^{-1}\) (c), and 2909 cm\(^{-1}\) (d).

2.3.2 Statistical Modeling of Spectra. Statistical modeling was used to study the small discrepancies in spectra that correspond to known functional groups shown in Table 1. The main components found in whole human blood are glucose, hemoglobin, proteins, and lipids. These components vary between different age groups due to human body development and can be studied using ATR FT-IR spectroscopy.

A total of 200 spectra were used to build a PLSDA model. The model included 50 spectra of newborns, 40 spectra of adolescents, and 110 spectra of adults. The results of the model can be seen in Figure 4, which shows the strict class prediction for newborn (red), adolescent (green),
and adult (blue) samples. Each symbol on the plot is a result of an individual spectrum collected with ten spectra collected per sample. Based on internal predictions of the model, 196 spectra were correctly classified, and only four spectra were unassigned (Figure 4). For these four spectra, two spectra belonged to the adolescent group and two to the adult group. None of the spectra were misclassified. Moreover, 100% correct classification was observed at the donor level predictions with the training data set.

![Graph](image)

**Figure 4.** Strict class prediction scores plot of PLSDA model of each classification group: newborn (red), adolescent (green), and adult (blue) with spectral regions of 4000-2670 cm\(^{-1}\) and 1710-600 cm\(^{-1}\). A value 1 corresponds to newborns, value 2 to adolescents, and value 3 to adults. A score of 0 indicates unassigned predictions.
Table 1. Wavenumber assignments of ATR FT-IR spectra of blood to functional groups.

<table>
<thead>
<tr>
<th>Wavenumber (cm⁻¹)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>3500-3200</td>
<td>Water and hydroxyl (O-H stretching)</td>
</tr>
<tr>
<td>3284</td>
<td>Amide A (N-H stretching)</td>
</tr>
<tr>
<td>2958</td>
<td>Lipids (asymmetric stretching of CH₃)</td>
</tr>
<tr>
<td>2872</td>
<td>Lipids (symmetric stretching of CH₃)</td>
</tr>
<tr>
<td>1700-1600</td>
<td>Amide I (C=O stretching)</td>
</tr>
<tr>
<td>1560-1500</td>
<td>Amide II (N-H bending, C-N stretching)</td>
</tr>
<tr>
<td>1390</td>
<td>Lipids and proteins (symmetric bending of CH₃)</td>
</tr>
<tr>
<td>1239</td>
<td>Amide III (C-N stretching)</td>
</tr>
<tr>
<td>1082</td>
<td>Glucose (C-O stretching)</td>
</tr>
<tr>
<td>698</td>
<td>Amide IV (C-H bending)</td>
</tr>
</tbody>
</table>

A GA was performed to detect the most informative peaks in the spectra for distinguishing between the three age groups. As shown in Table 2, the major chemical components for determination of donor age are lipids (1390 cm⁻¹), proteins (1390 cm⁻¹ and 1239 cm⁻¹), and glucose (1082 cm⁻¹). Newborns have a different quantity of protein because they are rapidly growing at such a young age compared to adolescents and adults. Adults tend to have a decrease in muscle mass which correlates to the amount of protein in their system, especially as they age. These quantities are similar in healthy people in the same age bracket but can shift due to medical problems. This discrepancy was not considered in the current model.
Table 2. Spectral ranges identified by genetic algorithm that are statically significant for differentiation between donor age groups from bloodstains.

<table>
<thead>
<tr>
<th>Important spectral ranges identified by GA (cm⁻¹)</th>
<th>Functional groups within spectral ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000-3800</td>
<td>Water and hydroxyl (O-H stretching) [6, 29]</td>
</tr>
<tr>
<td>3100-3000</td>
<td>Water and hydroxyl (O-H stretching) [6, 29]</td>
</tr>
<tr>
<td>1500-1350</td>
<td>Lipids and proteins (symmetric bending of CH₃) [6, 31]</td>
</tr>
<tr>
<td>1300-1200</td>
<td>Amide III (C-N stretching) [6, 31, 32]</td>
</tr>
<tr>
<td>1110-1000</td>
<td>Glucose (C-O stretching) [31, 33]</td>
</tr>
</tbody>
</table>

2.3.3 Model Validation. After the model was built with the training data set, it was validated using the subject-wise LOOCV method. In this method, spectra from each donor at the time were excluded from the training data set and used for predictions. CV is a technique that can be used to validate the stability of a model and how well a model will perform with unknown samples. The samples are used for both training and testing which is beneficial especially with small sample size. This process provides more metrics of the model and helps to draw important conclusion about the algorithm and data set. The goal of CV is to limit problems like overfitting by capturing noise or patterns that do not generalize well to unseen data set. In order to build the final model for the prediction of future cases, the learning algorithm would be applied to the entire learning set. The purpose of CV is to provide an estimate for the performance of this final model on independent data set. LOOCV is a good choice for high-dimensional data giving a low bias.[34] The major advantage of CV is the universality of the data splitting which only assumes that data are identically distributed and the training and test samples are independent. Thus, it is important for study involving subjects/donors which contain the subject-to-subject variation to perform subject-wise LOOCV which eliminates the subject bias.[35] Based on a subject-wise
LOOCV in this study it was found that 96% of the spectra were correctly classified, and 95% of the donors were correctly assigned. Table 2 shows the results of the classification. All 50 spectra of newborn donors were assigned correctly. The adolescent class had only one unassigned spectrum out of 40 spectra total, and the adult class had two spectra unassigned and five misclassified out of 110 spectra total. The results of the subject-wise LOOCV per donor can be seen in Table S2. These results illustrated high accuracy within the model.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Newborn</th>
<th>Predicted Adolescent</th>
<th>Predicted Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newborn</td>
<td>50</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Adolescent</td>
<td>0</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Adult</td>
<td>0</td>
<td>0</td>
<td>103</td>
</tr>
<tr>
<td>Unclassified</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

2.4 Conclusions

Phenotype profiling is a tool that can help narrow down suspects and the victim pool. Using the biological composition in the blood, one can study the differences in samples and categorize samples based on classification groups. In this study, age was examined in three groups: newborn (<1 year old), adolescent (11-13 years old), and adult (43-68 years old). With the use of ATR FT-IR spectroscopy and chemometrics, the dried blood, which exhibited significant intra-class variations, was analyzed and classified into each prospective class of human age groups. Specifically, the PLSDA model was used to enhance spectral differences to create a model for
distinguishing between newborn, adolescent, and adult samples. This approach was evaluated by subject-wise LOOCV, resulting in 96% correct spectral classification and 95% correct donor classification. Further studies based on larger cohort of donors is necessary to validate the reported approach before it can be used in practice. Overall, this proof-of-concept study demonstrated the great potential of ATR FT-IR spectroscopy and chemometrics for phenotype profiling for forensic purposes based on dry bloodstains. This methodology is specifically advantageous in the forensic field since it is nondestructive and could be potentially used at a crime scene. Future research is required to validate the developed methodology with respect to potential environmental contaminations and the presence of other body fluids as well as substrate interference and the limit of detection of the portable instruments.

APPENDIX:

Funding

This project was supported by Awards No. 2017-R2-CX-0006 and 2017-DN-BX-0135 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the U.S. Department of Justice.

Permissions

Chapter 2 of this project is under review at the journal of American Chemical Society Omega.

References