Integration of Raman Spectroscopy and Python-based Data Analysis for Advancing Neurobiological Research

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Natalie Estelle Dunn

December 2023

© by Natalie Estelle Dunn, 2023
All Rights Reserved.

To Frank Vogt, William Foley (III), Beverly 'Buvy' Duyser
and to Mom, Dad, Cathy, Nick, and all my friends and family.

Acknowledgements

I would like to thank every person who has believed in me and encouraged me throughout the years, and shaped me into who I am today. I am indebted to both of advisors for all your support and contributions during my time in graduate school. Dr. Vogt, your kindness and spirit will never be forgotten; you deserved more. You started me on this journey, I wish we could talk about all the code I have written since you left. Dr. Sharma, thank you for being a true advisor and mentor, giving me the freedom to explore new paths and always being available when I have questions.

I want to thank my labmates who let me cover any available space with sticky notes, as well as collected data and proofread for me. Avery and Nick, I don't think I would have survived this without you both.

Thank you to those who raised me: my parents, Cathy, Bill, and Buvy. Not only did you give me a wonderful childhood, but you each inspired me to become a pretty cool adult. Thank you for always believing in me and all of the unconditional love.

I want to thank Mr. Frank for letting me know that it was okay that I was awkward in high school and showed me that I am smart. I think of you when shaping my teaching style.

I want to express my gratitude to Ms. Patsy, Jershiya, and the Moses neighborhood. Thank you for becoming family, you all have filled my time in Tennessee with happiness.

Abstract

Raman spectroscopy, a highly specific and sensitive vibrational spectroscopy, continues to demonstrate great promise for biological applications, particularly for both qualitative and quantitative analysis of molecules related to human health. However, Raman scattering is an inherently weak process, with only 1 in 100 million scattered photons being Raman scattered. This leads to the need for (1) implementation of methods to enhance the Raman scattered signal and (2) implementation of multivariate analysis to extract useful information from low signal to noise spectra. each collected spectra to undergo a preprocessing step to remove noise, background drift, and cosmic rays.

In biological research, large datasets are required due to the large heterogeneity in samples. As datasets grow, any data processing, including pre-processing steps specifically for Raman spectra that include removal of cosmic rays, correction of background, and smoothing, that were previously be carried out through simple algorithms or specialized software that required processing spectra one at a time, become daunting. As Raman spectroscopy enters the realm of "big data", the implementation of computational methods for both pre-processing of Raman spectra, as well as multivariate analysis, need to be developed. Many open-source algorithms exist to assist in spectral processing, however these are designed around 'ideal' homogeneous samples with strong Raman scattering signals and limited complexity. Therefore, there is a need to design open-source programs to accurately and efficiently analyze large Raman datasets in order to advance biological Raman spectroscopy research. The work described in this dissertation

includes development of a Python-based approach for biological Raman data analysis, including molecules in complex biological fluids and living cells.

Table of Contents

1 Introduction **1**

2 Raman Spectroscopy **8**

2.1 Abstract . . . 9

2.2 History of Raman . . . 9

2.3 Raman Theory . . . 10

2.3.1 Raman Effect . . . 10

2.3.2 Classical Theory . . . 12

2.3.3 Surface-Enhanced Raman Spectroscopy (SERS) . . . 15

2.4 Raman Instrumentation . . . 17

2.4.1 Excitation Source . . . 17

2.4.2 Types of Lasers . . . 18

2.4.3 Optics and Sample Stage . . . 19

2.4.4 Dispersion of Raman Scattered Light . . . 20

2.4.5 Detection of Raman Scattered Light . . . 23

2.4.6 Sharma Lab Raman Instrument . . . 24

3 Computational Data Analysis **28**

3.1 Introduction . . . 29

3.2 Considerations for Raman Spectroscopic Data . . . 29

3.2.1 Definition of 'spectra' . . . 29

3.2.2 Spectral Frames . . . 30
3.2.3 Metadata . . . 30
3.2.4 Challenges in Converting Raman Spectra to 2D Matrices . . . 33
3.3 Pre-Processing . . . 33
3.3.1 Cosmic Ray Spikes . . . 34
3.3.2 Baseline Correction . . . 36
3.3.3 Noise . . . 36
3.4 Motivation for *RamanRoot* . . . 37
3.5 Program Framework Architecture . . . 38
3.6 Tier: *Data_Wrangle* . . . 41
3.6.1 Motivation . . . 41
3.6.2 Storage and Retrieval . . . 41
3.6.3 Dataset and Dictionaries . . . 44
3.7 Tier: *Spectracles* . . . 46
3.7.1 Motivation . . . 46
3.7.2 Requirements . . . 47
3.7.3 Initial Figure Object . . . 47
3.7.4 Averaging . . . 48
3.7.5 Generating Figures . . . 48
3.8 Tier: *Analyst* . . . 48
3.8.1 Object Initialization . . . 49
3.8.2 PCA . . . 49
3.8.3 Leave One Out Cross-Validation . . . 50
3.9 Future Development . . . 50

4 Surfaced-Enhanced Raman Spectroscopy Combined with Multivariate Analysis for the Detection and Differentiation of Dopamine and DOPAC in Cerebrospinal Fluid 53

4.1 Abstract . . . 55

4.2 Introduction . . . 55
4.3 Methods . . . 58
4.3.1 Materials . . . 58
4.3.2 Cyclic voltammograms . . . 58
4.3.3 Nanoparticle Synthesis . . . 58
4.3.4 Sample Preparation . . . 59
4.3.5 Raman Instrumentation . . . 59
4.3.6 Raman Spectra Preprocessing . . . 60
4.3.7 Multivariate Analysis . . . 60
4.4 Results and Discussion . . . 61
4.5 Conclusion . . . 74

5 SERS and Multivariate Analysis for Discrimination Between Neurotoxic Arsenic Species in Cerebrospinal Fluid 84

5.1 Abstract . . . 85
5.2 Introduction . . . 85
5.3 Background . . . 87
5.3.1 Water Contamination: Bangladesh . . . 87
5.3.2 Therapeutic Use and Detection in Biofluids . . . 90
5.4 Experimental Section . . . 90
5.4.1 Materials . . . 90
5.4.2 Synthesis of Nanoparticles . . . 91
5.4.3 Sample Preparation . . . 91
5.4.4 Micro-Raman Spectroscopy . . . 91
5.4.5 Raman Data Pre-processing and Multivariate Analysis . . . 92
5.5 Results and Discussion . . . 92
5.5.1 Arsenic Detection with SERS . . . 92
5.5.2 Multivariate Analysis . . . 96
5.6 Conclusions . . . 102

6 Conclusions **110**

A Chapter 4 Supplemental Information **113**

Vita **120**

List of Tables

4.1 SERS peak assignments for DA (orange), DOPAC (green), DA/DOPAC mixtures (purple), and blanks (AuNPs in aCSF, gray). 63

4.2 SERS peak positions from the loadings of PCs 1-3 with colors indicating the SERS spectra from which they arise: DA (orange), DOPAC (green), DA/DOPAC mixtures (purple), blanks (AuNPs in aCSF, gray). 70

5.1 Structures of inorganic arsenic species and reported lethal amounts values. 86

5.2 A summary of neuropathy in APL patients treated with ATP in literature. Severe neurpathy is defined as severity classification grades of 3 or 4 (affecting daily living or incapacitating) , with reported side effects being dizziness, moodiness, musculoskeletal pain, and seizures. Adapted from Kühn et al.[12]. 88

5.3 SERS peak assignments (cm^{-1}) for As(III), As(III)/As(V) mixtures, As(V) and blanks (AuNPs in aCSF) 95

5.4 SERS peak positions from the loadings of PCs 1-4 with symbols indicating the SERS spectra from which they arise. 100

A.1 Healthy Brain Concentrations for DA, DOPAC, HVA, 3-OMD, and 3-MT 114

A.2 Summary of the samples prepared for DA/DOPAC in aCSF study. For each sample group, five samples were prepared at each concentration or mixture 115

List of Figures

1.1 Visualization of the rate of scientific publications referencing Raman spectroscopy, according to search engine Web of Science, February 2023 . . 3

2.1 Jablonski diagram demonstrating the energy changes experienced by a photon during scattering. Where h is Planck's constant, E_0 is the ground state, E_1 is the first excited state, f_0 is the frequency of the ground vibrational state and f_m is a vibrational mode. 11

2.2 Polarized electromagnetic radiation depicted as a wave, where E represents the electric component and H represents the magnetic component. In terms of Raman scattering, only the electric component is considered. Adapted from Ferraro et al. 13

2.3 Localized Surface Plasmon Resonance (LSPR). Excitation of the metal surface conduction electrons by a laser results in the collective oscillation of electrons around a nanoparticle and generates an electric field. The generated electric field enhances the signal of the Raman scattered light. . 16

2.4 Diagram of a reflective diffraction grating showing how the incident light (In) strikes a groove and returns diffracted (Di) and reflected (Re) light. d is the groove spacing of the grating, α is the incident light angle with respect to the grating normal, and β is the angle of the diffracted light. Adapted from Vandenabeele 22

2.5 The Micro-Raman System in the Sharma Lab. The Raman system consists of four lasers, a Nikon inverted microscope, a spectrometer, and a CCD camera detector. 25

3.1 Raman data consists of information that falls into three categories: spectra, experimental, or instrumental. Spectral data, the Raman spectrum itself, consists of an array of Raman shifts and the Raman intensity recorded at that shift. Instrumental data represents parameters specific to the spectrometer. All information regarding the sample itself falls into the broad category of experimental data. Data collected for any singular spectra falls into three categories: instrumental parameters, sample information, and the Raman spectra itself 31

3.2 Illustration of the day-to-day variability in starting wavenumber position when compiling Raman spectra from multiple sources. 32

3.3 Example spectra from microalgal cells. The raw spectra are shown in blue, the approximated baseline of the raw spectra is calculated with an ASLS Spline function and shown in red. The final spectra are shown in green after cosmic ray removal, baseline removal, followed by smoothing via a Savitzky-Golay filter. 35

3.4 Broad overview of the *RamanRoot* architecture. The three tiers *Data_Wrangle*, *Spectracles*, and *Analyst*, are designed to be independent. There is a 'utils' object included in the package, which offers Python shortcuts for tasks frequency used when using the package in Python. The tiers use and produce standardized objects, providing functionality of the overall package to grow without needing to rewrite large sections of code. The user controls the analysis through a Jupyter Notebook. A summary of input and output objects is provided for each tier to interface interaction. 39

3.5 Overview of the *Data_Wrangle* tier. A database object must be initialized with a path to the database and a table name. Once initialized, the user can interact with the database to summaries of the data, compile a dataset for further analysis, or generate labels for spectra in the database using values from the MySQL columns. Functions within these modes produce standardized datasets for analysis, filtered and labeled based on the values contained with their metadata. 40

3.6 A flowchart diagram depicting the 'Build Dataset' module in *Data_Wrangle*. This pipeline grabs Raman spectra and metadata from a MySQL database can gets them to a standardized dataset that can be used as an input into *Spectracles* and *Analyst*. A database object is created when a user provdies the path and name of a SQL database table. 42

3.7 An example of a MySQL database table used to store RS data. Each csv file imported into the database is treated as an unique spectrum and stored as a row, with an unique sampleID assigned to it. The first and second columns within the csv files are expected to be the 'Raman shifts' and the 'original intensity' values, respectively. Both columns are converted into arrays and then stored within the database as blob-type data. Metadata can be stored within the database when the csv files are imported. After initial import, metadata can be viewed, added to, and edited using functions within *Data_Wrangle* or by directly accessing the database. 43

3.8 Summary of two types of dictionaries, utilized within *RamanRoot* 45

4.1 Cyclic voltammogram of 100 µM 3-MT (black), 3-OMD (green), DOPAC (red), DOPA (blue), HVA (pink) in 0.1 M $HClO_4$ and PBS. The potential was scanned from -0.80 V to 0.75 V at a 100 mV/s scan rate. 62

4.2 Averaged SERS spectra of DA, DOPAC, and DA/DOPAC mixtures and blanks. Pure DA SERS spectra are shaded orange and pure DOPAC SERS spectra are shaded green, both with the darkest color representing the greatest analyte concentration. Mixture SERS spectra are colored on a red-blue gradient based on the ratio of DA to DOPAC in the mixture: greater ratio being red and lowest ratio being blue. λ_{ex} = 785 nm, P= 10 mW, t= 30 sec. 66

4.3 3-dimensional PCA plot of PC 1 vs. PC 2 vs. PC3. Pure DA SERS spectra are represented by the orange shades and pure DOPAC SERS spectra are represented by the green shades, both with the darkest color representing the greatest analyte concentration. DA/DOPAC mixture SERS spectra are colored on a red-blue gradient based on the ratio of DA to DOPAC in the mixture; greater ratio being red and lowest ratio being blue. The SERS spectra for the blanks are shown as gray crosses. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec. 68

4.4 2D PCA plot of PC 1 vs. PC 2 of the SERS spectra of DA, DOPAC, DA/DOPAC and blank solutions. Pure dopamine solutions are marked with an orange square (■), pure DOPAC solutions are marked with a green diamond (♦), mixtures are marked with a circle (●) and the blanks are marked with grey crosses (+). A red to blue gradient was applied to the mixtures depending on the ratio of DA to DOPAC in the solution, with red having the highest ratio and blue the lowest. Pink represents two mixtures that are not naturally found in the brain 10 µM DA + 10 µM DOPAC (light pink) and 7 µM DA + 3 µM DOPAC (raspberry pink). The loadings for PC 1 and PC 2 are included. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec. 72

4.5 2D PCA plot of PC 1 vs PC 2 zoomed in to highlight the concentration dependence of the pure DA samples (top, orange) and DA/DOPAC mixtures (bottom, pink to blue). In both cases, the samples separate along a gradient from highest DA concentration to lowest DA concentration. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec. 73

5.1 SERS spectra of arsenic samples in aCSF 94

5.2 2D PCA plot of PC1 vs. PC 2 for SERS spectra of arsenic species in aCSF. The loadings for PC 1 and PC 2 are included. λ_{ex} = 785 nm, P = 10 mW, t = 120s. 97

5.3 PCA plot of PC 2 vs PC 3, with loadings included. There is limited spread of the pure As(III) samples along the PC 3, especially compared to the spread seen in the mixtures. λ_{ex} = 785 nm, P = 10 mW, t = 120s 98

5.4 PCA plot of PC 2 vs PC 4, with loadings included. λ_{ex} = 785 nm, P = 10 mW, t = 120s. 99

5.5 A flow chart summarizing the conclusions which can be drawn from the PCA analysis of Arsenic species in aCSF. 103

A.1 SERS spectra of 10 nM DA (blue), 3-OMD (red), 3-MT (green), HVA (purple), DOPAC (pink) and a spectrum of the blank (AuNPs in 0.1 M HClO4 and PBS). λ_{ex} = 785 nm, P= 10 mW, t= 30 s. Figure produced by P. Pimento. 116

A.2 PC 1 vs PC 3 of DA and DOPAC solutions. Both PC 1 and 3 strongly differentiate DA/DOPAC solutions from the blank. λ_{ex} = 785nm, P= 10 mW, t= 30s. 117

A.3 PC 2 vs PC 3 of DA and DOPAC solutions. From this perspective, there is clear distinction between DOPAC and its P mixtures. Pure DA and high concentration DA mixtures cluster tightly compared to the pure DOPAC and low concentration DA mixtures. This difference in clustering is attributed to the smaller concentration range for the DA in mixtures (1-3 µM) vs the large concentration range (7-18 µM) for DOPAC. λ_{ex} = 785nm, P= 10 mW, t= 30s. 118

Chapter 1

Introduction

Raman spectroscopy (RS), a powerful analytical technique, is applied across multiple disciplines, due to its unique ability to quantitatively analyze substances in their native state with little to no sample preparation (Figure 1.1).[1,2] RS provides excellent chemical specificity, allowing researchers to obtain unique spectra specific to their molecules of interest. The high specificity of Raman spectra led it to be considered molecular 'fingerprinting' technique.[3,4] RS has broad applicability that spans diverse fields such as art, archaeology, space exploration, medicine, and environmental research. RS has proven to be an invaluable tool for probing the molecular composition of materials and gaining critical insights into the chemical structure.

RS is widely considered to be a non-destructive technique that is used to study delicate and fragile samples across various fields. RS has been used to examine ancient and modern artwork, including paintings found within Pompeii, revealing the chemical composition of the pigments used. The knowledge gained through this analysis aids in understanding ancient civilizations and aids in conservation efforts to protect and preserve our cultural heritage.[5] Mineral composition can be determined with RS, aiding in geological surveys and soil characterization[2,6,7] The development of portable Raman spectrometers have added to its applications and versatility.[3,7] Raman spectroscopy is a vital component of the Mars 2020 rover with two instruments on the Perseverance rover, SHERLOC (Scanning

Habitable Environments with Raman and Luminescence for Organics and Chemicals), and SuperCam. SHERLOC is a deep-UV Raman spectroscopy system that allows for the detection of biologically relevant molecules and organic matter, while SuperCam is a combined stand-off Raman spectroscopy and laser-induced breakdown spectroscopy (LIBS) system focused on determining the chemical composition of rocks and soil. Both Raman-based systems on the Perseverance rover are focused on determining the chemical makeup of the surface of the plant and searching for evidence of life beyond our planet.[8,9]

RS is of particular interest for human health monitoring, as it provides the ability to probe biological systems, often with limited sample preparation, and extract chemical information.[10] RS offers distinct advantages over traditional methods for studying biomolecules in cells and tissues, including high sensitivity and spatial resolution, without the addition of exogenous markers, as is needed in fluorescence spectroscopy. RS is growing as a biological and medical technique, due to the ability to directly visualize the chemical responses of molecules within living cells and tissues.[11] Additionally, multiple molecules can be monitored simultaneously, without impacting cellular activity, as can sometimes be found with fluorescence.[11]

However, due to the inherent weakness of Raman scattering, signal amplification is often necessary, particularly for trace analyte detection. Surface-enhanced Raman spectroscopy (SERS) is used to overcome this inherent weakness through implementation of a phenomenon associated with noble metal nanoparticles known as the localized surface plasmon resonance (LSPR). The LSPR is an electric field generated at the surface of nanostructures through excitation of the surface conduction electrons of the metal. Excitation of the LSPR results in enhancement of the Raman scattering signals from samples near (within 1-2 nm) or adsorbed to the surface of the metal nanoparticle. Enhancement of the signal through SERS is powerful enough to detect trace chemicals down to concentrations of a few ppb[3,8] and has been applied to distinguish between species of bacteria and fungal cells.[12]

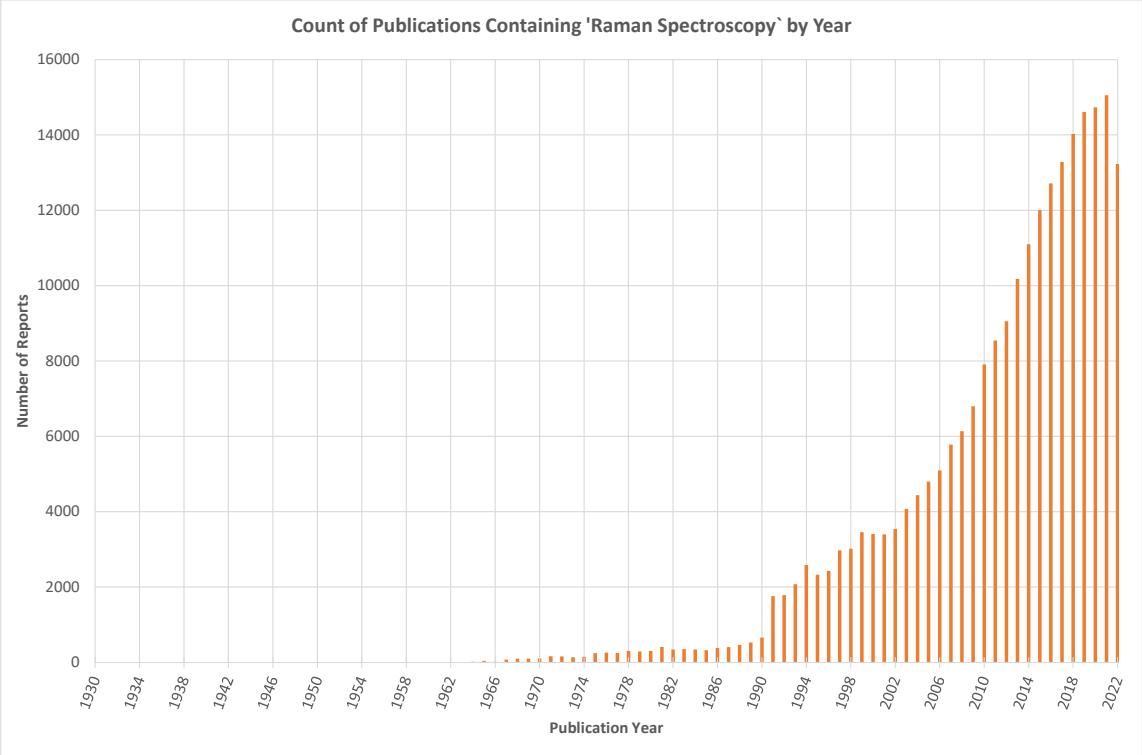


Figure 1.1: Visualization of the rate of scientific publications referencing Raman spectroscopy, according to search engine Web of Science, February 2023

Throughout this dissertation, Raman spectroscopy is highlighted as the core analytical technique and is discussed in detail in Chapter 2. Chapter 3 details an interactive tool for Raman spectral data management, pre-processing, visualization, and multivariate analysis. While software tools for data processing and analysis exist, they are either not designed specifically for Raman spectroscopy, or they are designed for analysis of a single spectrum at a time, which hinders their usability when dealing with large datasets. Addressing these needs, the RamanRoot Python-based package described in this dissertation has been developed to streamline the Raman spectral data organization, visualization, and analysis processes. Chapters 4 and 5 highlight the power of SERS in discrimination between molecules with similar chemical composition, dopamine and its metabolites (4) and between neurotoxic and non-toxic arsenic species (5). The implementation of the interactive Raman spectral data management tool described in chapter 3 is also presented in chapters 4 and 5.

References

[1] Antonio Candeias and Juan M. Madariaga. Applications of raman spectroscopy in art and archaeology. *Journal of Raman Spectroscopy*, 50(2):137–142, 2019. doi: doi.org/10.1002/jrs.5571. URL https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/jrs.5571. 1

[2] Guo L-Yu J Zeng P Wang A, Han J. Database of standard raman spectra of minerals and related inorganic crystals. *Applied Spectroscopy*, 48(8):959–968, 1994. doi: 10.1366/0003702944029640. 1

[3] Salvatore Almaviva, Florinda Artuso, Isabella Giardina, Antonia Lai, and Alessandra Pasquo. Fast detection of different water contaminants by raman spectroscopy and surface-enhanced raman spectroscopy. *Sensors*, 22(21), 2022. ISSN 1424-8220. doi: 10.3390/s22218338. URL https://www.mdpi.com/1424-8220/22/21/8338. 1, 2

[4] Shuxia Guo, Jürgen Popp, and Thomas Bocklitz. Chemometric analysis in raman spectroscopy from experimental design to machine learning–based modeling. *Nature Protocols*, 16:5426–5459, 2021. URL https://doi.org/10.1038/s41596-021-00620-3. 1

[5] Irantzu Martínez-Arkarazo Anastasia Giakoumaki Kepa Castro Juan M. Madariaga Maite Maguregui, Ulla Knuutinen. Raman spectroscopy as a tool to diagnose the impact and conservation state of pompeian second and fourth style wall paintings exposed to diverse environments (house of marcus lucretius. *Journal of Raman Spectroscopy*, January 2011. doi: https://doi-org.utk.idm.oclc.org/10.1002/jrs.2671. URL https://analyticalsciencejournals-onlinelibrary-wiley-com.utk.idm.oclc.org/doi/full/10.1002/jrs.2671. 1

[6] Aderval S. Luna, Igor C. A. Lima, Werickson F. C. Rocha, Joyce R. Araújo, Alexei Kuznetsov, Erlon H. Martins Ferreira, Ricard Boqué, and Joan Ferré. Classification

of soil samples based on raman spectroscopy and x-ray fluorescence spectrometry combined with chemometric methods and variable selection. *Anal. Methods*, 6(22): 8930–8939, 2014. doi: 10.1039/C4AY01967A. URL http://dx.doi.org/10.1039/C4AY01967A. 1

[7] Delano G. Henry, Ian Jarvis, Gavin Gillmore, and Michael Stephenson. Raman spectroscopy as a tool to determine the thermal maturity of organic matter: Application to sedimentary, metamorphic and structural geology. *Earth-Science Reviews*, 198:102936, 2019. ISSN 0012-8252. doi: https://doi.org/10.1016/j.earscirev.2019.102936. URL https://www.sciencedirect.com/science/article/pii/S0012825219303824. 1

[8] Jeffrey T. Osterhout, J. William Schopf, Anatoliy B. Kudryavtsev, Andrew D. Czaja, and Kenneth H. Williford. Deep-uv raman spectroscopy of carbonaceous precambrian microfossils: Insights into the search for past life on mars. *Astrobiology*, (10):1239–1254, October 2022. doi: doi.org/10.1089/ast.2021.0135. URL https://pubmed.ncbi.nlm.nih.gov/36194869/. 2

[9] Caudill C. Lalla-E. A. Newman J. Daly M. Lymer E. Freemantle J. et al. Cloutis, E. A. Lunar: Overview of a versatile raman spectrometer for lunar exploration. *Front. Astron. Space Sci.*, 9, October 2022. doi: https://doi.org/10.3389/fspas.2022.1016359. URL https://www.frontiersin.org/articles/10.3389/fspas.2022.1016359/full. 2

[10] Jean-Francois Masson. Surface plasmon resonance clinical biosensors for medical diagnostics. *ACS Sens.*, 2(1):16–30, 2016. URL https://doi-org.utk.idm.oclc.org/10.1021/acssensors.6b00763. 2

[11] Katsumasa Fujita Kosuke Dodo and Mikiko Sodeoka. Raman spectroscopy for chemical biology research. *Journal of the American Chemical Society*, 144(43):19651–19667, 2022. doi: 10.1021/jacs.2c05359. URL https://pubs.acs.org/doi/full/

10.1021/jacs.2c05359. 2

[12] Roger M. Jarvis, Alan Brooker, and Royston Goodacre. Surface-enhanced raman scattering for the rapid discrimination of bacteria. *Royal Society of Chemistry*, 132: 281,292, 2006. doi: 10.1039/B506413A. URL https://pubs.rsc.org/en/content/articlelanding/2006/fd/b506413a/unauth. 2

Chapter 2

Raman Spectroscopy

2.1 Abstract

Raman spectroscopy (RS), a highly specific and sensitive vibrational spectroscopy, is applied to harness the information contained within the molecular vibrations of a sample after it has been excited by a light source. From the molecular vibrations, the composition, structure, and even concentration of a sample can be inferred. RS is an important analytical technique, as it allows for the identification of complex chemical compounds with little sample preparation and without the need for exogeneous labeling.

2.2 History of Raman

RS is a relatively young analytical technique; it was theoretically predicted by Smekal in 1923,[1] and was first observed by Sir Chandrasekhra Venkata (C.V.) Raman and K.S. Krishnan in 1928.[2] In the first Raman spectroscopy experiment, the sun was implemented as a light source, passed through simple light filters and lenses, and directed onto the sample,[3] a 600 ml beaker of pure liquid. It was noted that the light striking the sample was a different color than the resulting scattered light.[1] The Raman spectra of the scattered light was recorded on photographic plates, over a period of 24 hours.[1,2] Raman and Chandrashekar continued development of the technique, moving from sunlight as an excitation source to the use of lamps.

While the use of lamps improved the applicability of the technique, Raman scattering was still considered a curiosity, without any real-world applications. To utilize Raman scattering in practical applications, a powerful, stable, excitation source and a highly sensitive detector would be needed. During the 1950s, some powerful light sources were developed, but it was not until the 1960s and the development of lasers (light amplification by stimulated emission of radiation) that Raman scattering and its applications were truly able to be explored.[1]

2.3 Raman Theory

2.3.1 Raman Effect

Scattering refers to the process that happens when a photon interacts with a molecule. After the interaction, most photons have no change in frequency, simply traveling in a new direction after the interaction (referred to as Rayleigh scattering). However, some photons experience a change in frequency, which is referred to as Raman scattering.

If you consider a wave of light as a series of photons being directed towards a substance, some of the photons pass through the substance (transmission), while others will collide into a molecule with enough force to cause an electronic transition (absorption). However, a few photons will strike a molecule without enough force to push an electron into a new orbital. Due to this temporary interaction, an electron can be excited to what is referred to as a *virtual energy state*.

As shown in Figure 2.1, once a photon is excited to a virtual energy state, it undergoes one of three processes as it relaxes back to its ground state. Most photons will undergo *Rayleigh* or *elastic* scattering; the photon will return to the same vibrational frequency (f) it was at before the scattering event. These Rayleigh scattered photons will experience no permanent change in energy, therefore no chemical information can be discerned from detecting these photons.

However, 1 in every 10^8 photons will experience permanent change in energy called *Raman* or *inelastic* scattering. The change in the frequency of the photon can be a gain of energy or a loss of energy. *Anti-Stokes* describes a Raman scattered photon that has a net increase in energy, while photons that lose energy undergo *Stokes* scattering. At room temperature, there is an increaesed probability of electrons existing in the ground state, which results in the Stokes scattering being the most common form of inelastic scattering.

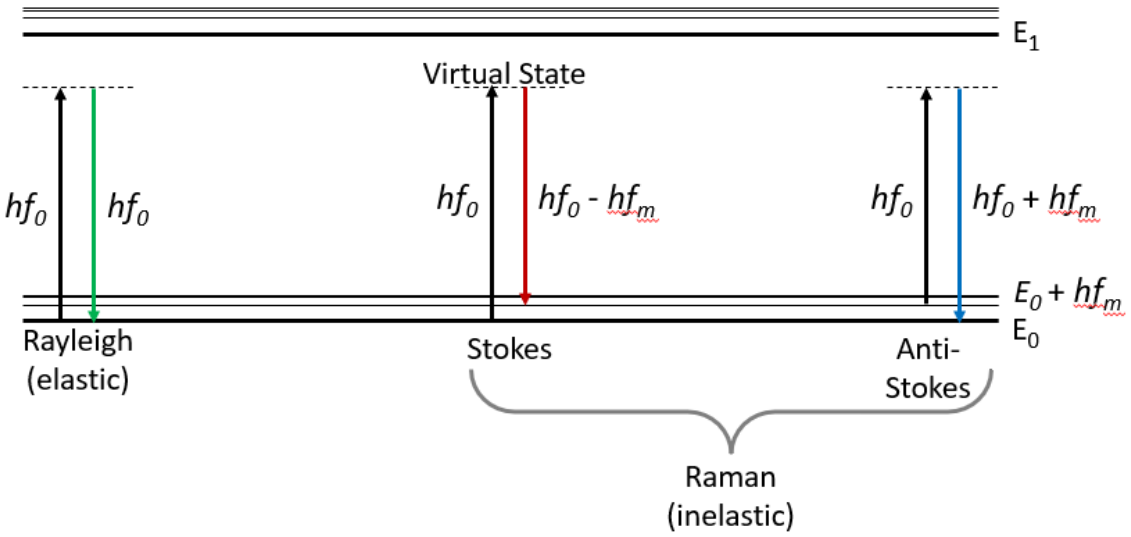


Figure 2.1: Jablonski diagram demonstrating the energy changes experienced by a photon during scattering. Where h is Planck's constant, E_0 is the ground state, E_1 is the first excited state, f_0 is the frequency of the ground vibrational state and f_m is a vibrational mode.

Molecular Information from Raman Scattering

Information on the structure and composition of a molecule can be derived from the change in frequency a photon undergoes after interactions with the molecule. This frequency change is a function of the difference in energy between the initial and final molecular vibrational energy levels. The difference in frequency between the initial and final energy levels is equal to the frequency of the molecular bond vibration.

2.3.2 Classical Theory

Polarizability describes the probability of a the electron cloud of a molecule to be deformed by an external electric field. A change in the polariability is a key factor in the intensity of the Raman scattered light and the efficiency of the Raman Effect. A Raman active molecule can undergo a temporary change in polarizability (α) when it is exposed to an electromagnetic field (E). This results in the molecule experiencing an induced dipole moment (P), as show in equation 2.1.[4]

$$P = \alpha_m E \tag{2.1}$$

The light incident on a sample that induces the Raman scattering is an oscillating electrical field, which can be described as:

$$E = E_0 \cdot cos(2\pi \cdot f_0 \cdot t) \tag{2.2}$$

where (f_0) and (E_0) represent the frequency of the photons and equilibrium electromagnetic field strength, respectively, at a given time (t).[2]

Combining equations 2.1 and 2.2 results in the description of the induced dipole moment as a function of the incident photon:

$$P = \alpha_m \cdot E_0 \cdot cos(2\pi \cdot f_0 \cdot t) \tag{2.3}$$

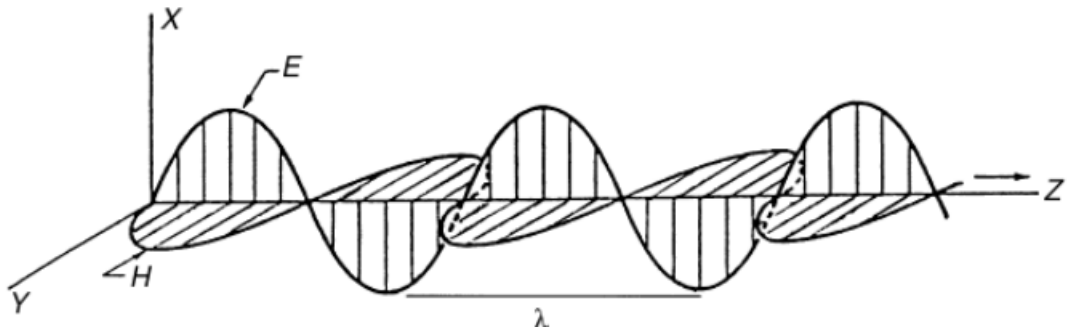


Figure 2.2: Polarized electromagnetic radiation depicted as a wave, where E represents the electric component and H represents the magnetic component. In terms of Raman scattering, only the electric component is considered. Adapted from Ferraro et al.

The manner in which a molecule vibrates is described by its normal coordinate (q_m), a vector that summarizes how the molecule is moving in space.[4] Polarizability of a molecule can be described in terms of the normal coordinate when the amplitude of q_m is small:

$$\alpha = \alpha_0 + \left(\frac{\delta\alpha}{\delta q_m}\right)_0 q_m \tag{2.4}$$

Equation 2.5 expresses the normal coordinate as a linear equation:

$$q_m = q_0 \cdot cos(2\pi \cdot f_m \cdot t) \tag{2.5}$$

Combining equations 2.4 and 2.5 gives:

$$\alpha = \alpha_0 + \left(\frac{\delta\alpha}{\delta q_m}\right) \cdot q_0 \cdot cos(2\pi \cdot f_m \cdot t) \tag{2.6}$$

Equations 2.1, 2.2, and 2.6 can be combined to express P as a function of frequency, polarizability, and the electric field (E) of the incident light:

$$\begin{aligned} P &= \alpha E \\ &= \left(\alpha + \left(\frac{\delta\alpha}{\delta q}\right)_0 \cdot q_0 \cdot cos(2\pi f_m t)\right) \cdot (E_0 \cdot cos(2\pi f_0 t)) \\ &= \alpha_0 \cdot E_0 \cdot cos(2\pi f_0 t) + \left(\frac{\delta\alpha}{\delta q}\right)_0 \cdot q_0 \cdot E_0 \cdot cos(2\pi f_m f \cdot cos(2\pi f_0 t)) \end{aligned} \tag{2.7}$$

The trigonometry identity, shown here as 2.8, can be applied to equation 2.7.

$$cos(\alpha)cos(\beta) = \frac{1}{2}cos(\alpha + \beta) + cos(\alpha - \beta) \tag{2.8}$$

The result of this can be seen in equation 2.9:

$$P = \alpha_0 \cdot E_0 \cdot cos(f_0 t) + \frac{1}{2}\left(\frac{\delta\alpha}{\delta q_m}\right)_0 q_0 E_0[cos(f_0 - f_m)\cdot t] + \frac{1}{2}\left(\frac{\delta\alpha}{\delta q_m}\right)_0 q_0 E_0[cos(f_0 + f_m)\cdot t] \tag{2.9}$$

A series of mathematical manipulations allow the induced dipole moment to be written as a function of the vibrational frequencies of the molecule (f_m) and incident radiation (f_0):

$$P = P(f_0) + P(f_0 - f_m) + P(f_0 + f_m) \tag{2.10}$$

In equation 2.10 the first term represents Rayleigh scattering while the second and third terms represent Raman scattering. The component $f_0 + f_m$ (anti-Stokes scattering) shows a increase in energy while $f_0 - f_m$ shows energy decrease in energy (Stokes scattering).

2.3.3 Surface-Enhanced Raman Spectroscopy (SERS)

In 1974, Fleischmann et al., noted that when exciting the Raman spectra of pyridine on the surface of silver electrodes, there was an enhancement of the Raman scattering.[5] They, however, offered no explanation for the enhanced signal. In 1977, Jeanmaire and Van Duyne explained the mechanism for the increased intensity and signal to noise ratio of the Raman scattering on the roughened electrode surface, which was also confirmed by Albrecht and Creighton.[6,7] This enhanced Raman scattering on the roughened electrode surface was named surface-enhanced Raman spectroscopy (SERS). Continued study of the SERS phenomenon has led to a theoretical understanding of the mechanism of action.

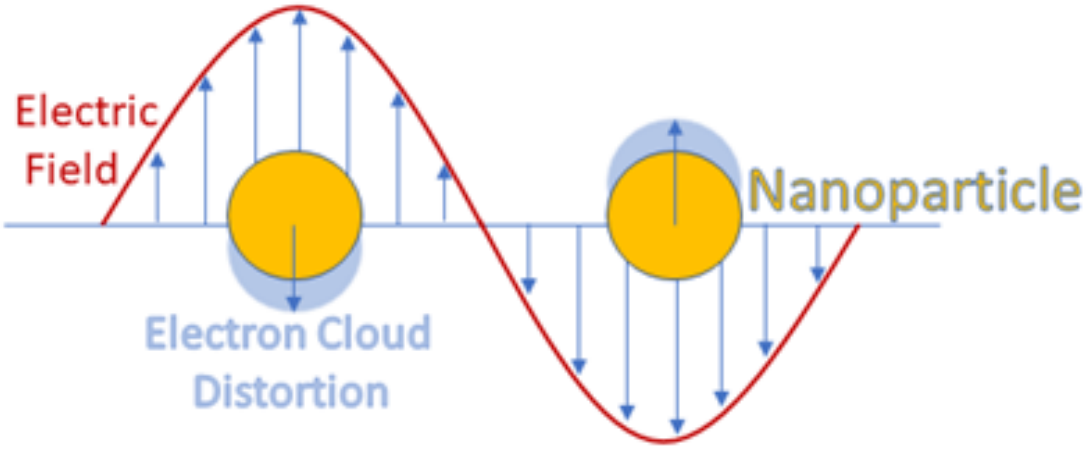


Figure 2.3: Localized Surface Plasmon Resonance (LSPR). Excitation of the metal surface conduction electrons by a laser results in the collective oscillation of electrons around a nanoparticle and generates an electric field. The generated electric field enhances the signal of the Raman scattered light.

In SERS, noble metal nanoparticles (NPs), usually gold or silver, are synthesized. The NPs are then combined with the sample. When the sample+NPs are excited by a laser, an electric field, called the localized surface plasmon resonance (LSPR) (Figure 2.3), is generated at the surface of the NPs. Molecules within 1-2 nm or adsorbed to the NP surface are excited by the laser and result in Raman scattering, which is then enhanced by the LSPR, generally by an enhancement factor of up to $10^{8\text{-}10}$.

2.4 Raman Instrumentation

Numerous companies provide a diverse array of Raman instruments meeting various specifications. However, those seeking greater control over the functionalities of a Raman instrument may build their own. Instrumentation for Raman spectroscopy includes several key components, including various optics for directing and focusing light; an excitation source, which is generally a laser; a spectrometer, for collection of the Raman scattered light; and a detector, which detects the dispersed Raman scattered light from the spectrometer. Raman instruments are set-up in various geometries and can be designed as either macro-illumination systems or integrated with a microscope as a micro-illumination system. The work described in this dissertation completed on a micro-Raman system, which will be described in detail below (2.4.6).

2.4.1 Excitation Source

Exciting Raman scattering requires an excitation source with the following characteristics. The light will need to be a monochromatic (single wavelength), collimated, and coherent. Monochromatic light allows for selective excitiation of Raman-active vibrational modes in the target. Collimation insures that the rays of light are parallel and do not diverge over large distances. Coherent waves are in the same phase temporally and spatially, in which the emitted light interacts constructively, strengthening the signal, and ensuring that the detected frequency shifts are consistent. The sensitivity of the

system is increased with a coherent light source, due to constructive interference. Finally, the light source will need to be consistent, stable, and widely available with different wavelengths. Additionally, narrow bandwidths are also required.

Lasers fulfill all of these characteristics, providing stable, coherent, and monochromatic light. Today, all Raman spectroscopy instrumentation uses lasers as the excitation light sources. Lasers produce the high intensity light that is required to excite Raman scattering.

2.4.2 Types of Lasers

Lasers fall into two main categories: pulsed and continuous-wave (CW). The primary lasers used to excite Raman spectra were pulsed Nd:YAG or continuous wave noble gas lasers. Most Raman instrumentation use CW lasers, as they are simpler to operate, relatively inexpensive, and offer continuous excitation of the sample.[8] Recently, CW gas lasers have been replaced by lighter, more stable, more reliable, and cost-effective diode lasers.[9]

Raman spectroscopy can be excited by laser wavelengths from the UV (<200 nm) to the near-IR (1064 nm). Common excitation laser wavelengths include 488, 532, 633, and 785 nm. Fourier-transform (FT) Raman spectrometers typically use longer wavelengths, such as 1064 nm.

Laser Safety

There are multiple functions of lasers, from laser pointers, to scientific research, to medical grade lasers. These lasers can be harmless or are capable of causing skin or eye damage. Government standards, such as the ANSI Z136.1 in the United States, have been established to classify lasers. *Class 1* lasers are eye and skin safe lasers, with low laser outputs in the micro-Watt range. *Class 2* lasers are have low intensities (∼1 mW) and are generally in the visible wavelength range. *Class 2* lasers are fairly safe, but will cause damage if directed into the eye for longer than 0.25 seconds. *Class 3b* lasers are only safe

with diffuse reflections. It is not safe to have direct contact with these lasers. *Class 4* lasers are they most dangerous class of lasers. They have the highest power (Watts) and are capable of inducing skin and eye damage, even from specularly reflected light.

When operating a laser, it is important to limit reflective surfaces in the area, which can inadvertently reflect light randomly throughout the room. Lasers beams should be blocked when not actively in use. When a beam is being adjusted, laser goggles, which work by filtering light at the specific wavelength of the laser, should be used.[2]

2.4.3 Optics and Sample Stage

The sample is the central component of Raman instrumentation. The sample can be in any phase, solid, liquid, or gas, which results in the need for a multifunctional sample stage. Optics are used to direct, focus, and capture light. The optics are versatile and increase the flexibility of a Raman system to handle different sample types.

Excitation

A standard micro-Raman set up will consist of a laser, microscope, optics including lenses, mirrors, irises, filters, a spectrometer and detector. Mirrors are used to direct the excitation wavelength from the laser toward the sample. The irises are used to insure that the laser is parallel to the laser table and does not deviate. Lenses are used to focus the beam. Filters, including a laser line clean-up filter, which insures a single wavelength is incident on the sample, and a long-pass filter, which removes the Rayleigh scattered light and allows Raman scatterd light to enter the spectrometer, are also used.

In micro-Raman systems, the sample is placed on the stage of a microscope, which can be either an upright or inversted microscope. In the work here, an inverted microscope is included in the micro-Raman system. In an inverted microscope, the excitation light is directed into the microscope, through the microscope objective, and onto the sample from below. The objective functions the same as a standard light microscope: the light is

focused on the sample by adjusting the space between sample and objective, irradiating a specific, controlled, and powerful laser spot.

Sample

Depending on the type of sample, it may be placed on the sample stage directly, in a vial or cuvette, or other surface such as a microscope slide. Care should be used in selecting the substrate the sample is placed on because depending on the material of the substrate, it can either contribute background signal or have a Raman spectrum itself . Thin glass, such as microscope coverslips offer limited noise and signal interference, while most plastics will add to the background signal of the sample with the Raman spectrum of the polymer, which usually consists of sharp narrow peaks. Quartz has essentially no Raman signal, while thicker glass usually contributes a broad peak in the 1400 wavenumber regions.

When using Raman spectroscopy for quantification, generally a concentration curve can be developed that utilizes a relationship between the integrated intensity of Raman peaks and sample concentration. As samples become more complicated, this simple relationship no longer holds. Biological samples, for example, are comprised of numerous dynamic components, that are not static, and instead, interact with one another and change over time. The Raman spectrum of such samples reflect the complex nature of the sample, as the overall Raman spectrum is a sum of the Raman spectra of the individual components. In this case, multivariate analysis and/or methods for deconvolution of the Raman spectra are necessary to understand the complexity of the RS spectrum.

2.4.4 Dispersion of Raman Scattered Light

Once light is directed into the Raman spectrometer, it strikes a dispersion element which separates the light into discrete wavelengths. The simplest dispersion element is a prism (either glass or quartz), which is able to separate wavelengths based on the diffraction angle of each wavelength as they pass through a triangular prism.

Prisms are still used in Raman systems today as they are simple, widely available, inexpensive, and have high spectral resolution. However, it is more common for modern Raman systems to use diffraction gratings, which are comprised of a series of grooves, where each groove serves as a reflective component, as shown in figure 2.4.

A grating is selected based on the distance between grooves so that the distance is similar to the incident wavelength. The advantage of a diffraction grating is its utilization of the constructive and destructive interference. The incident light strikes a series of reflective elements, with each one causing diffraction. Wavelengths that are not similar to the optimized wavelength for the grating will destructively interfere, while wavelengths that are similar to the optimized wavelength of the grating will constructively interfere. The constructive interference allows for reduced noise and increased Raman scattering signal.

The following equation describes the relationship between constructive interference and diffraction angles (as shown in figure 2.4).[2]

$$\lambda = d(sin\alpha - sin\beta) \tag{2.11}$$

where λ is the wavelength of light, d is the groove spacing of the grating, α is the angle of light incident on the grating with respect to the grating normal, and β is the angle of diffracted light.

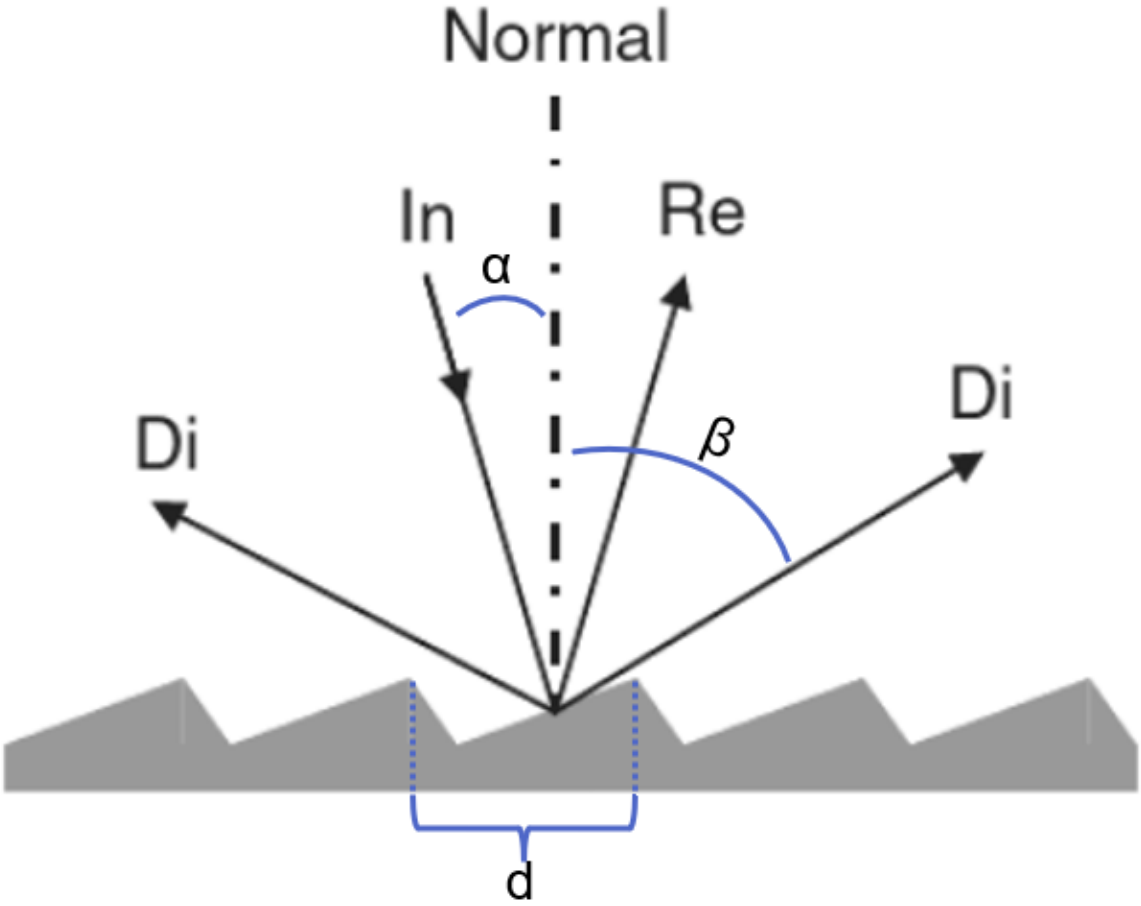


Figure 2.4: Diagram of a reflective diffraction grating showing how the incident light (In) strikes a groove and returns diffracted (Di) and reflected (Re) light. d is the groove spacing of the grating, α is the incident light angle with respect to the grating normal, and β is the angle of the diffracted light. Adapted from Vandenabeele

2.4.5 Detection of Raman Scattered Light

Due to the inherent weakness of Raman scattered light, detectors for Raman systems need to be highly sensitive. Detectors work by converting the incoming signal, carried as photons, to an electronic signal. The measure of the number of electrons generated per photon striking the detector is referred to as quantum efficiency. The larger the quantum efficiency, the more sensitive the detector. This is often displayed by companies in a response curve, called a quantum efficiency curve, which is often depicted as the quantum efficiency versus wavelength (nm).

PMTs

The most sensitive detectors available for the first Raman systems (developed in the 1960s) were photon multiplier tubes (PMTs). These detectors consist of a photocathode followed by a series of dynodes, which act as electron multipliers. The incident light is directed to the photocathode, which produces a small number of electrons for each photon. The electrons are then directed to the first dynode, which releases a few more electrons for every incident electron; this processes continues down the series of dynodes, each iteration of the process strengthening the electronic signal.[1]

While PMTs are able to increase the signal, they can simultaneously multiply the noise and background signal. They are also single channeled, only able to detect one wavelength at a time.

CCDs

Charge-Coupled Devices (CCDs) have mostly replaced PMTs and any other detector in modern Raman spectroscopy instruments. CCDs are less susceptible to damage from high intensity exposure. Not only are they more sensitive and robust, they are also multi-channeled, meaning they are able to measure several wavelengths simultaneously.

CCDs are comprised of a 2-dimensional array of cells, laid out on a circuit. Each cell contains a light-sensitive component, which builds an electron potential proportional to the number of photons that strike it.

2.4.6 Sharma Lab Raman Instrument

The data collected for the work desribed in this dissertation was conducted using a homebuilt micro-Raman instrument, as shown in figure 2.5. In this system, there are four excitation lasers available: 488, 532, 633, and 785 nm. The 785 nm laser is a continuous wave diode laser (Innovative Photonic Solution, Monmouth, NJ), the 633 nm laser is a helium-neon (He-Ne) laser (Newport Corporation, Irvine, CA), the 532 and 488 nm lasers are both optically pumped semiconductor lasers (OPSLs) (Coherent, Inc., Santa Clara, CA).

The excitation light is directed to into the microscope from the laser source through a series of optics and a periscope. The optics and periscope are used to align the beam to be parallel to the table and to direct the beam of light towards the sample stage. The sample is placed on the stage of a confocal Nikon Ti-U inverted microscope (Nikon, Melville, NY) and viewed with either a 20X, 40X, or 100X objective. The light is then directed through the objective and focused on the sample.

The backscattered Raman light from the sample is collected through the objective and directed out ofthe microscope through a side port. Light that exits the side port is sent through a collection lens that focuses the scattered light through a long-pass filter on the 100 µm slit of an ISOPlane SCT 320 spectrometer (Princeton Instruments, Trenton, NJ). The Raman scattered light is dispersed within the spectrometer via a grating, and is then directed to a Pixis 400 charge-coupled device (CCD) detector (Princeton Instruments, Trenton, NJ), which converts the scattered light signal into a Raman spectrum.

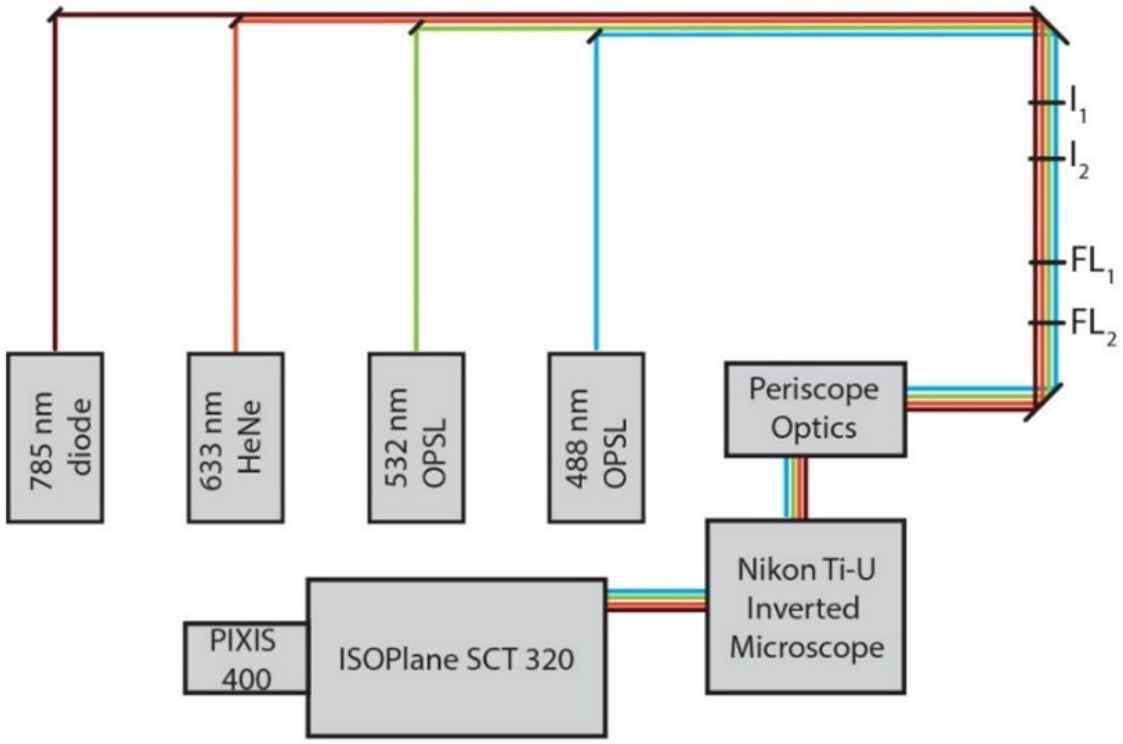


Figure 2.5: The Micro-Raman System in the Sharma Lab. The Raman system consists of four lasers, a Nikon inverted microscope, a spectrometer, and a CCD camera detector.

References

[1] Derek J Gardiner. *Practical Raman Spectroscopy*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989. ISBN 9783642740404. 9, 23

[2] Peter Vandenabeele. *Practical raman spectroscopy: An introduction*. John Wiley amp; Sons, Incorporated, 1 edition, 2013. https://ebookcentral.proquest.com/lib/utk/reader.action?docID=1272701. xii, 9, 12, 19, 21, 22

[3] John R. Ferraro, Kazuo Nakamoto, and Chris W. Brown. *Introductory Raman Spectroscopy*. Academic Press, 2 edition, 2003. https://ebookcentral.proquest.com/lib/utk/reader.action?docID=299529. xii, 9, 13

[4] L. A. Woodward. *Raman Spectroscopy: Theory and Practice*, pages 1–43. Springer US, Boston, MA, 1967. ISBN 978-1-4684-3024-0. doi: 10.1007/978-1-4684-3024-0_1. URL https://doi.org/10.1007/978-1-4684-3024-0_1. 12, 14

[5] M. Fleischmann, P.J. Hendra, and A.J. McQuillan. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters*, 26(2):163–166, 1974. ISSN 0009-2614. doi: https://doi.org/10.1016/0009-2614(74)85388-1. URL https://www.sciencedirect.com/science/article/pii/0009261474853881. 15

[6] David L. Jeanmaire and Richard P. Van Duyne. Surface raman spectroelectrochemistry: Part i. heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 84(1):1–20, 1977. ISSN 0022-0728. doi: https://doi.org/10.1016/S0022-0728(77)80224-6. URL https://www.sciencedirect.com/science/article/pii/S0022072877802246. 15

[7] Grant M. Albrecht and Alan J. Creighton. Anomalously intense raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society*, 99(15):5215–5217, Jun 1977. doi: 10.1021/ja00457a071. 15

[8] Yashar E. Monfared, Travis M. Shaffer, Sanjiv S. Gambhir, and Kevin C. Hewitt. Continuous-wave coherent raman spectroscopy via plasmonic enhancement. *Scientific Reports*, 9(1), 2019. doi: 10.1038/s41598-019-48573-8. 18

[9] S.M. Angel, M. Carrabba, and T.F. Cooney. The utilization of diode lasers for raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 51(11):1779–1799, 1995. ISSN 1386-1425. doi: https://doi.org/10.1016/0584-8539(95)01443-X. URL https://www.sciencedirect.com/science/article/pii/058485399501443X. 18

Chapter 3

Computational Data Analysis

3.1 Introduction

An important aspect of interpreting Raman spectra is data processing and the application of multivariate analysis. This chapter presents a Python-based package that was developed to manage large Raman spectral datasets, perform in-depth analysis, and visualize results. The significance of effective data management in handling large Raman spectral datasets has become increasingly evident during the course of the research presented in this dissertation, particularly since large datasets with multiple replicates are needed for statistical analysis and model development, with data for both testing and validating the model.

3.2 Considerations for Raman Spectroscopic Data

Generally, Raman spectra are collected in replicates, with more than one spectrum (often in triplicate) recorded per sample. When visualized on a plot, the x-axis constitutes the Raman shift, which represents the vibrational energy differences in the molecule upon excitation in wavenumbers (cm^{-1}), and the y-axis reflects the intensity of the Raman scattering detected at the corresponding energies.[1] During data collection, the two series of values are typically saved by the operating software of the spectrometer, such as LightField by Princeton Instruments. Each spectrum is then exported as a two-column CSV file for further processing, analysis, and visualization. In this state, the spectra are considered 'raw', as no pre-processing steps have been applied (see section 3.3).

3.2.1 Definition of 'spectra'

In the context of Raman spectroscopic data, the term 'spectra' can be understood in two ways, defined in this dissertation as 'discrete spectra' and 'composite spectra'. *Discrete* refers to the spectra described above, where each spectrum is an individual and unique observation, which is then recorded as a two-column file. When multiple discrete spectra are aggregated, the result is a *composite* spectra. They are generated by averaging

spectra based on a 'grouping factor', resulting in a representative spectrum that reflects the typical characteristics of a group of observations. In comparison, discrete spectra capture the variability and noise inherent in individual measurements.

3.2.2 Spectral Frames

During data acquisition, multiple frames may be collected for a single spectrum; frames are multiple spectral recordings of a given sample in a given spot. Each replicate frame is saved under the same name, differentiated by a frame number suffix. For heterogeneous samples, multiple spots for a single sample may be targeted, represented as technical replicates. These sets of frames are pre-processed separately, then averaged together; the resulting composite spectrum helps to reduce noise.

3.2.3 Metadata

Metadata is comprised of additional information related to the experiment, such as instrumental and experimental parameters (Figure 3.1). All information regarding a spectrum outside of the recorded Raman shifts and intensities is considered metadata. Common instrumental parameters include the excitation wavelength, excitation power, spectral acquisition time, and the type of Raman spectra collected (e.g., SERS or normal Raman). Meanwhile, experimental information helps encompass details, such as the timestamp, identity and concentration of an analyte within a solution, the source of a tissue sample, or the specific location on the sample where the excitation laser spot was focused. Both types of metadata can be used to determine the optimal collection method, identify sample outliers, and troubleshoot experimental issues.[1]

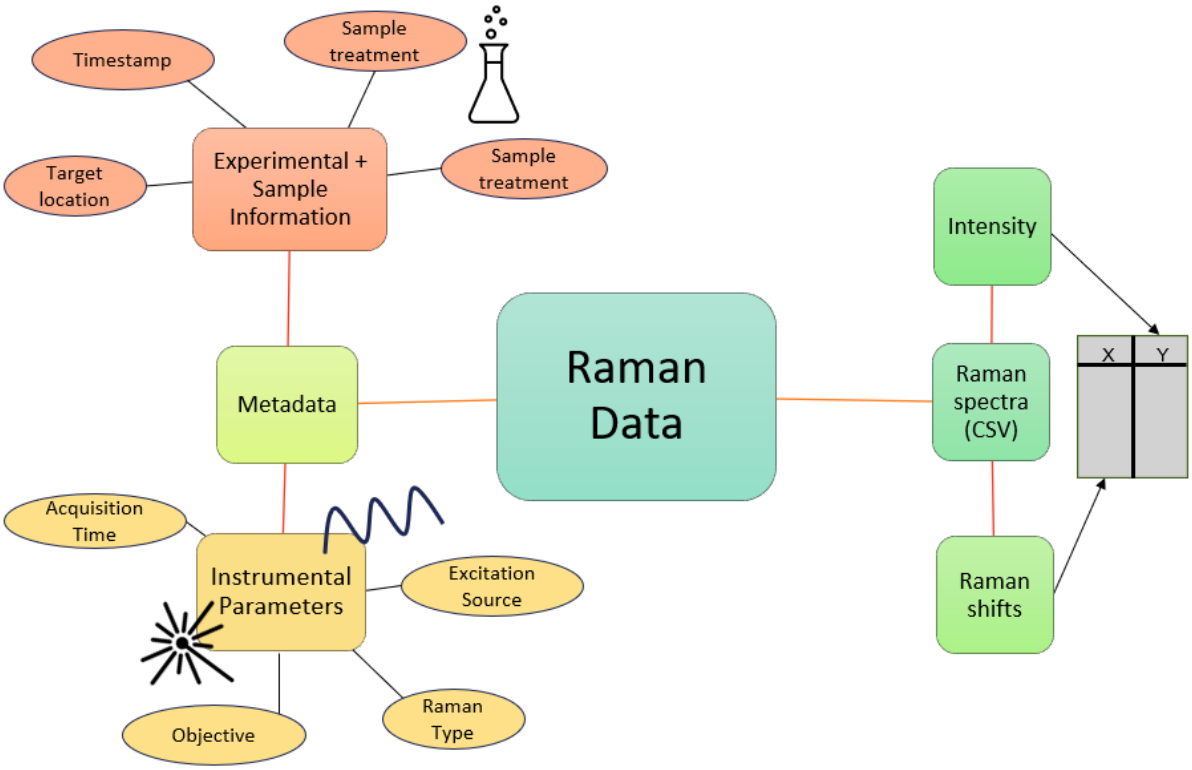


Figure 3.1: Raman data consists of information that falls into three categories: spectra, experimental, or instrumental. Spectral data, the Raman spectrum itself, consists of an array of Raman shifts and the Raman intensity recorded at that shift. Instrumental data represents parameters specific to the spectrometer. All information regarding the sample itself falls into the broad category of experimental data. Data collected for any singular spectra falls into three categories: instrumental parameters, sample information, and the Raman spectra itself

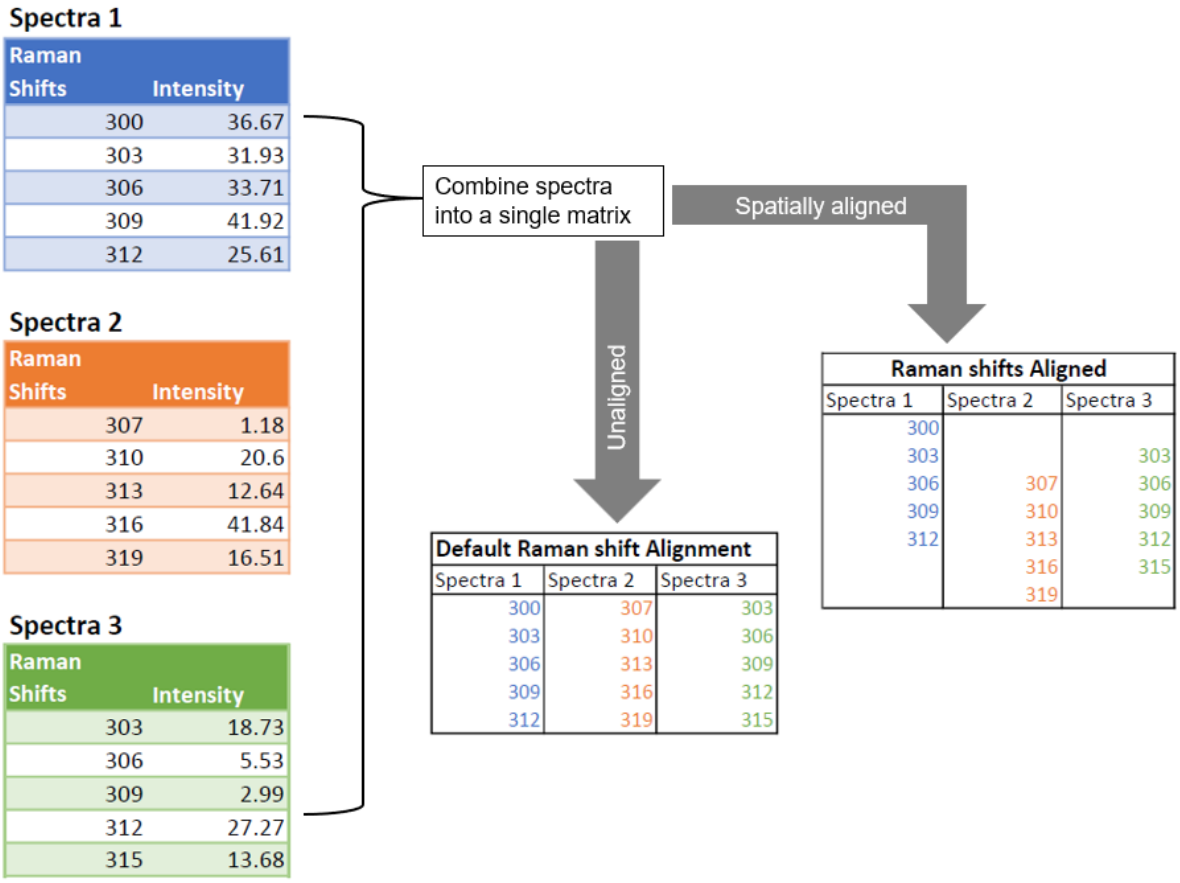


Figure 3.2: Illustration of the day-to-day variability in starting wavenumber position when compiling Raman spectra from multiple sources.

3.2.4 Challenges in Converting Raman Spectra to 2D Matrices

When collecting Raman spectra, one issue that arises is the day-to-day variability in data collection parameters. Since the grating in our spectrometer is tunable, on a given day there can be a slight position change in the grating. This results in collected Raman spectra from having different starting wavenumber positions or non-uniformly spaced 'x' values, or Raman shift values, and is exacerbated by differences arising from rounding decimal places of the recorded Raman shifts. This day-to-day variability does not affect the Raman peak positions, as those are reflective of the vibrational energy change in the molecule. We confirm this by measuring the Raman spectrum of a known standard at the beginning of each experiment and also routinely calibrating the instrument using a calibration lamp. However, these variations must be taken into account when comparing Raman spectra. If the Raman shift axes of the spectra do not start at the same wavenumber position, the data analysis becomes challenging because most multivariate and statistical analysis packages in Python accept input data in the form of a 2D matrix, in terms of Raman spectroscopy, this requires the spectra to be compressed along the Raman shift axis (Figure 3.2). The result, in its most basic form, is a 2D matrix where the rows represent individual discrete or composite spectra, and the columns represent Raman shifts. There cannot be any 'not a number' or 'NaN' values in the matrix.

3.3 Pre-Processing

Pre-processing steps are necessary before Raman spectra can be visualized and analyzed. These steps remove or minimize noise and random abnormalities from the spectra, which will otherwise mask the spectral features of the samples.[1] The main steps commonly performed during pre-processing are broadly described here: cosmic ray removal, baseline drift correction, and smoothing (Figure 3.3). These functions and algorithms transform the raw spectra into a usable form and have been developed in Python; they can be found in the *SpecPrePro* class of the *Data_Wrangle* tier within

RamanRoot, which will be discussed in detail later in this chapter. SpecPrePro is available as part of a stand alone Raman spectra batch preprocessing tool, which is hosted as a Google CoLaboratory within the Sharma Group Google Drive. The Juptyer Notebook environment of the CoLab is set up to guide a user to upload CSV files of Raman spectra, set parameters of the pre-processing steps, display before/after figures of the baselining process, and final export the results as new CSV files.

3.3.1 Cosmic Ray Spikes

High energy particles generated by cosmic rays cannot be entirely blocked from the Raman detector, resulting in distinctive, random sharp peaks on the spectrum. Some of the particles from the cosmic ray will interact with the detector, striking a pixel* at random.[2] Cosmic rays can be algorithmically detected and removed by assuming the pixel impacted is an outlier. This is the underlying principle used to correct cosmic ray spikes in the 'zap' function.[3]

First, outliers in a spectrum are determined using a modified Z-score, calculated for each intensity in the spectrum. A threshold variable is used to flag values that are outliers, a Z-score higher than the threshold. Next, for each outlier detected in the spectrum, the function selects a range of points (defined by a window variable) with the outlier in the middle. The outlier will be replaced with the average of the inlier values within the range. If all values in the range are outliers, the central one will be replaced by the average of all of the values in the range (central outlier included). Finally, the corrected series of intensity values is returned by the zap function.

*a pixel will reflect a single Raman shift value recorded as part of the final spectrum

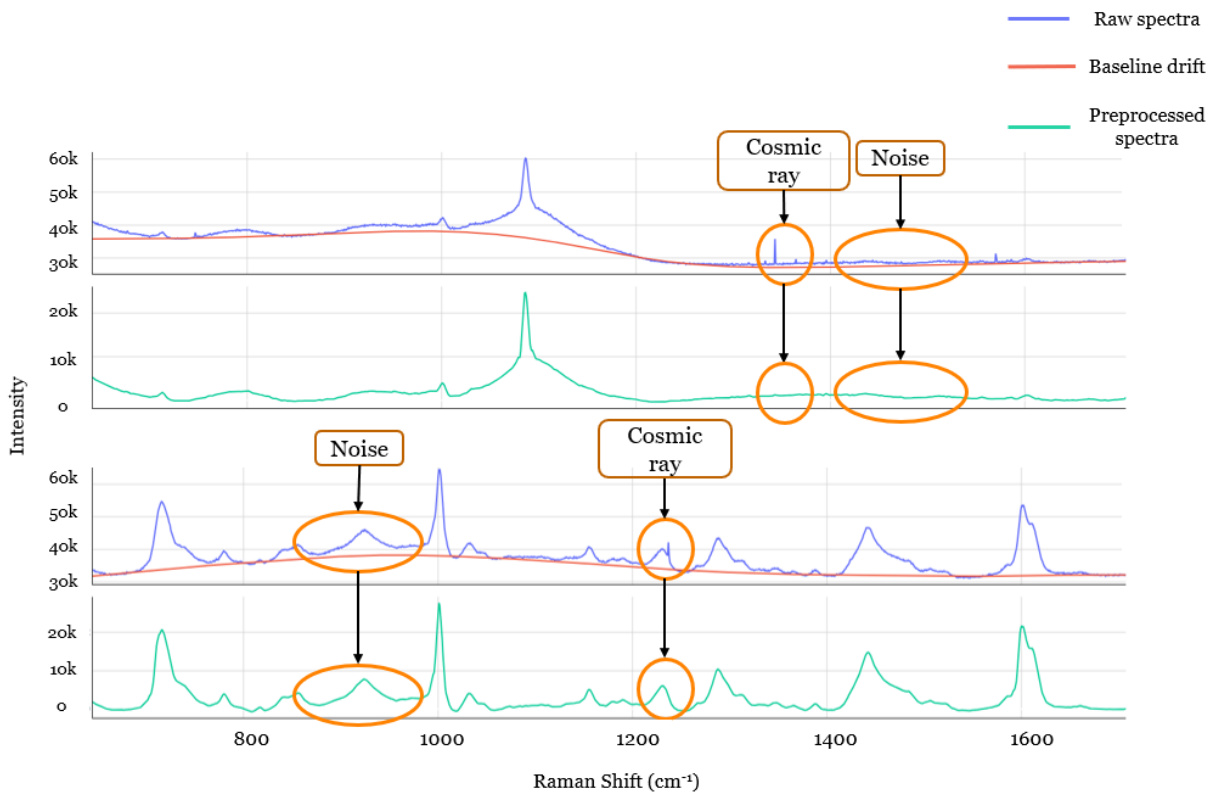


Figure 3.3: Example spectra from microalgal cells. The raw spectra are shown in blue, the approximated baseline of the raw spectra is calculated with an ASLS Spline function and shown in red. The final spectra are shown in green after cosmic ray removal, baseline removal, followed by smoothing via a Savitzky-Golay filter.

Z-Score

A traditional Z-Score is biased by extreme outliers, as it is dependent on the standard deviation and mean of the entire series. The modified Z-scores calculated in the 'zap' function avoids this bias by utilizing a more robust indicator, the Mean Absolute Deviation (MAD).[4] MAD evaluates how far a data point varies from the dataset, in relation to the variability within the data overall, using an algorithm based on the median.[5]

3.3.2 Baseline Correction

The baseline in Raman spectra arises from several factors, including fluorescence and dark noise (caused by random electrons created by the inherent nature of electronic devices and thermal energy). The baseline is seen in the spectra as a slow, steady rise in the intensity of the spectra.

Multiple algorithms exist that can be used for mathmatical baseline correction, many of them variations of the widely excepted Partial Least Squares (PLS) algorithm. Polynomial fitting, first derivitive, and wavelet transform-based are also commonly used to approximate the baseline of the raw spectra.[6] Regardless of the method used, the standard practice is that once the baseline has been approximated, it is subtracted from the original Raman spectrum. The 'baseline' function accepts a series of intensity values and approximates the baseline using the Asymmetric Least Squares (AsLS) algorithm,[7] which is imported from the 'PyBaselines' package.[8]

3.3.3 Noise

In terms of Raman data, noise is an inherent component of the observed Raman signal, the result of unavoidable imperfections within the spectrometer, which is layered on top of the true spectrum (eq 3.1).[9] Where $\bar{s}$ is the true spectra of the sample and e is the noise. The observed spectrum, s, is the spectrum reported to the operating computer

from the CCD detector.

$$s = \bar{s} + e \tag{3.1}$$

Noise poses a significant challenge because of the intrinsic weak intensity of Raman scattering, and can be reduced by applying a 'smoothing' function, which can be either a spectral and/or spatial filter.[1] The goal of applying smoothing is to minimize noise without significantly modifying the original spectrum. The smoothing function used in the pre-processing class applies a Savitzky-Golay filter, which is the standard filter used in Raman Spectroscopy.[10]

3.4 Motivation for *RamanRoot*

When applying Raman spectroscopy to detect and monitor biological samples, large numbers of replicate spectra are required for accurate statistical analysis due to the inherent variability and heterogeneity of biological matrices. A well-organized approach is essential to enable the analysis and interpretation of such extensive data, but is often overlooked. Traditionally, data are treated as a static pool of information. Data processing and visualization have separate and time-consuming steps, which limits the ability to dynamically adjust and compare datasets in real-time.

The motivation behind the *RamanRoot* package comes from the need to streamline and expedite the analysis of Raman spectroscopic data. Initial efforts focused on optimizing batch pre-processing and visualization, which significantly accelerated data analysis compared to conventional methods using graphical user interfaces, such as GRAMS AI (ThermoFisher Scientific). As our research projects progressed and larger datasets were collected, the primary Raman spectral analysis bottleneck shifted from data pre-processing to data management. It became increasingly clear that in addition to pre-processing data, it was necessary to develop methods that allowed for comparison of datasets collected on different days and under different conditions. Tools that utilize the Python framework, such as Dash and Plotly, have made interactive data visualization

easy to design; however, there are not tools that couple interactive visualization with computational analysis and data management for the specific needs of Raman spectroscopists. *RamanRoot* aims to fill this gap.

3.5 Program Framework Architecture

RamanRoot was developed of three independent tiers: *Data_Wrangle*, *Spectracles*, and *Analyst*; each tier is implemented separately, within an Object-Orientated Programming (OOP) structure. The modular design not only simplifies specific functionalities but also promotes the versatility of the code, which is crucial for a growing project. Individual tiers can be updated or replaced without impacting the rest of the package. Currently, the user interacts with *RamanRoot* using a Jupyter Notebook as a controller. A module within *RamanRoot*, 'utils', holds programs to facilitate user commands to the three tiers (Figure 3.4).

A central aspect of the *RamanRoot* design is the use of a MySQL database combined with Python programming to create a pipeline for data processing. A MySQL database is utilized to store Raman spectroscopic data in a structured format, allowing efficient organization and retrieval of the raw spectra and the associated metadata. MySQL is a type of SQL relational database where predefined relationships within table(s) are used to query subsets of data within the database. After the Raman spectra are exported from the operational software of the spectrometer and CCD detector as CSV files, the data within these sample-level CSV files can be imported into the database. Metadata for each CSV file can be extracted from the name of the file or can be found within lab notebooks. The database not only helps maintain organization, but also enables a quick and easy way for the user to edit and add details to each spectra as needed. Communication between the database and the controller is processed within the *Data_Wrangle* tier, along with dataset management and pre-processing of spectra. The products of this tier, standardized datasets and labels, can then be sent to the spectral visualization or data analysis tiers.

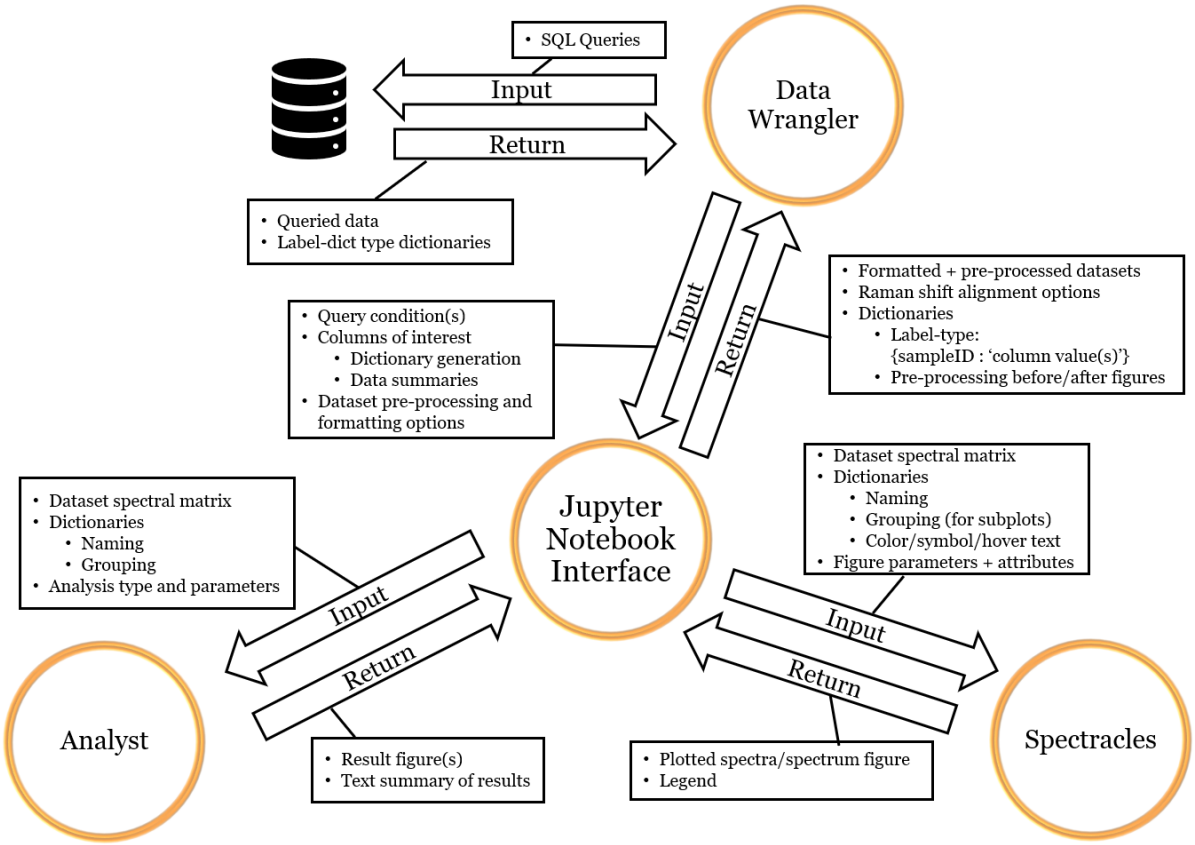


Figure 3.4: Broad overview of the *RamanRoot* architecture. The three tiers *Data_Wrangle*, *Spectracles*, and *Analyst*, are designed to be independent. There is a 'utils' object included in the package, which offers Python shortcuts for tasks frequency used when using the package in Python. The tiers use and produce standardized objects, providing functionality of the overall package to grow without needing to rewrite large sections of code. The user controls the analysis through a Jupyter Notebook. A summary of input and output objects is provided for each tier to interface interaction.

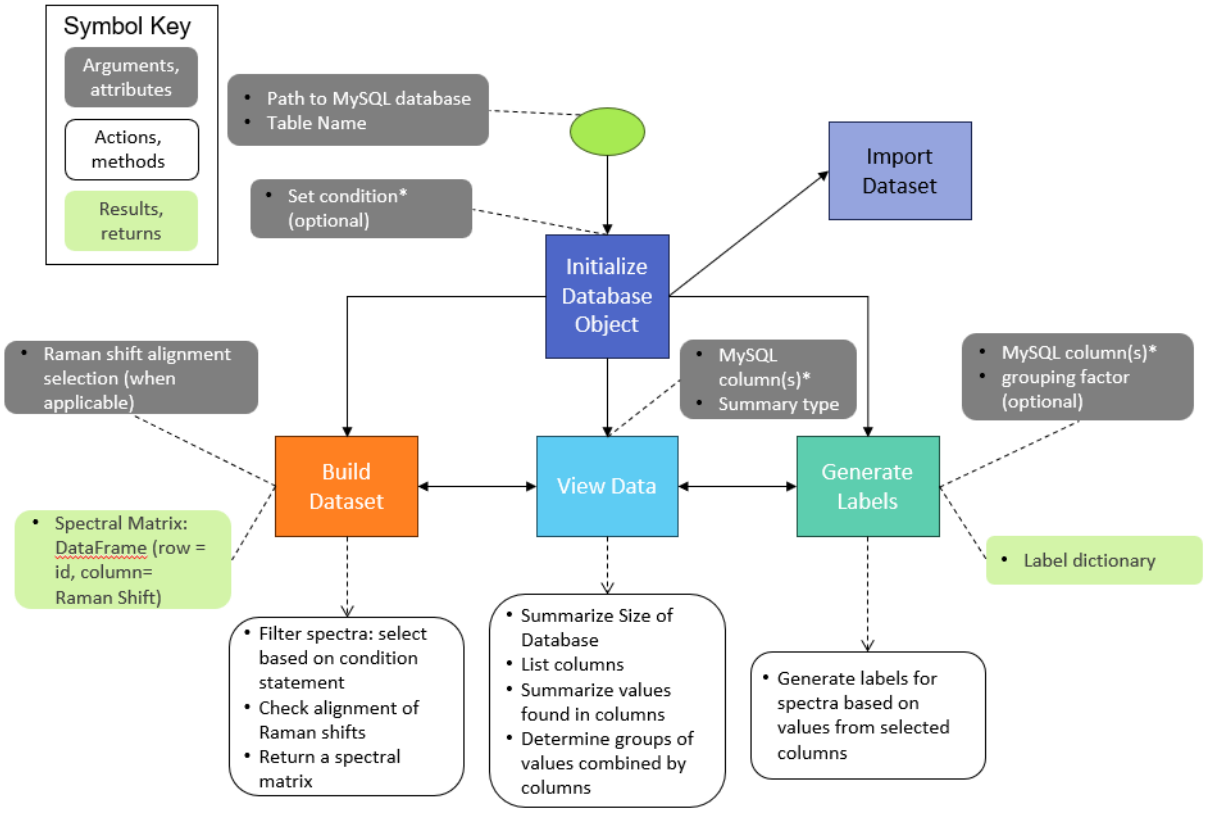


Figure 3.5: Overview of the *Data_Wrangle* tier. A database object must be initialized with a path to the database and a table name. Once initialized, the user can interact with the database to summaries of the data, compile a dataset for further analysis, or generate labels for spectra in the database using values from the MySQL columns. Functions within these modes produce standardized datasets for analysis, filtered and labeled based on the values contained with their metadata.

3.6 Tier: *Data_Wrangle*

The *Data_Wrangle* tier provides standardized data management for Raman spectra and metadata, as outlined in section 3.2. It comprises several components, which simplifies the storage, retrieval, pre-processing, and compilation of Raman spectra (Figure 3.5).

3.6.1 Motivation

Efficient data management is critical when working with biological samples, which often exhibit increased complexity and inherent variations. Meaningful results and robust statistical analysis require large datasets that capture the biological variability through numerous technical and biological replicates. As the size of the datasets, and the combinations of variables, grows, a database becomes increasingly important. Figure 3.6 depicts the overall order of processes executed to efficiently sort and import data to examine, apply and verify pre-processing steps, and generate organized objects that can be passed to *Spectracles* and *Analyst*.

3.6.2 Storage and Retrieval

The 'DataTable' class within *Data_Wrangle* interacts with the tables within a MySQL database. Data such as Raman shifts, corresponding intensity values, filenames, and additional metadata are stored within these tables (Figure 3.7). The metadata contained within this SQL database are used to execute queries and generate labels based on attributes stored in the database. When a new spectrum is added to the database, it is stored as a row in a data table, with a unique *sampleID* assigned to it.

To interact with the tables stored in the database, users can access and modify the data table through the Python *sqlite3* package or a desktop SQL application, such as MySQL Workbench or DB for SQLite. Along with publicly available tools, *Data_Wrangle* can query tables within a database and return data following a standardized format.

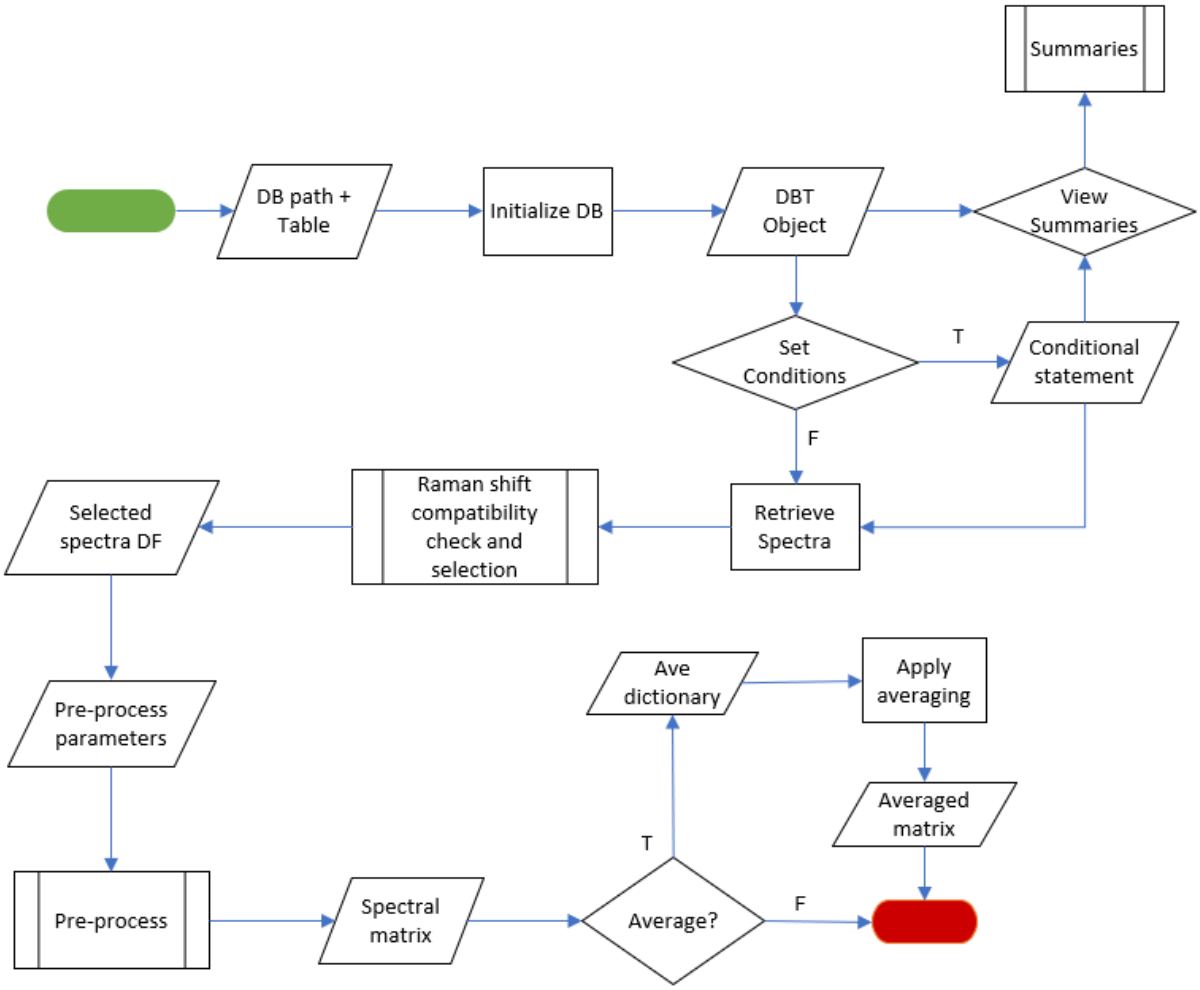


Figure 3.6: A flowchart diagram depicting the 'Build Dataset' module in *Data_Wrangle*. This pipeline grabs Raman spectra and metadata from a MySQL database can gets them to a standardized dataset that can be used as an input into *Spectracles* and *Analyst*. A database object is created when a user provdies the path and name of a SQL database table.

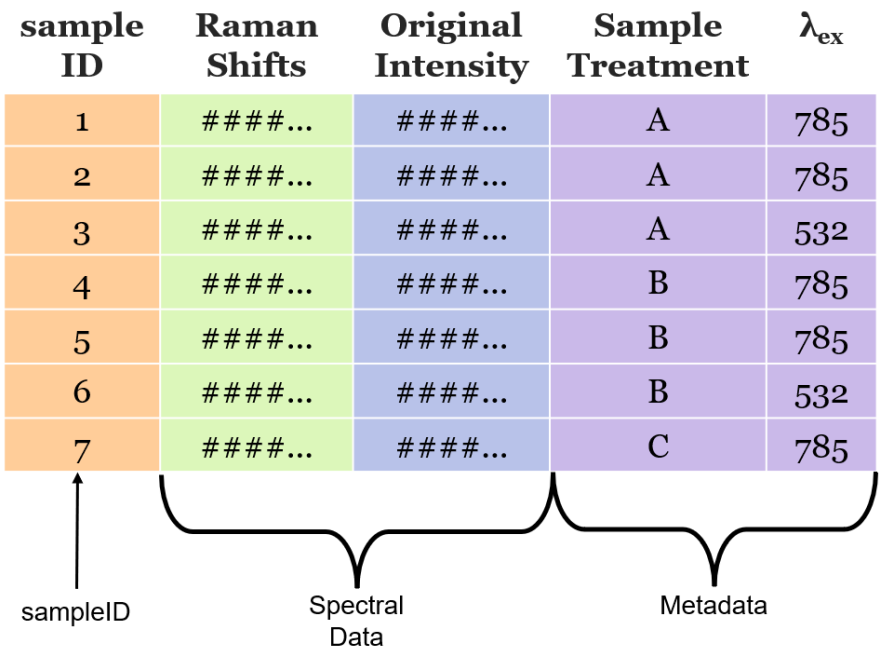


Figure 3.7: An example of a MySQL database table used to store RS data. Each csv file imported into the database is treated as an unique spectrum and stored as a row, with an unique sampleID assigned to it. The first and second columns within the csv files are expected to be the 'Raman shifts' and the 'original intensity' values, respectively. Both columns are converted into arrays and then stored within the database as blob-type data. Metadata can be stored within the database when the csv files are imported. After initial import, metadata can be viewed, added to, and edited using functions within *Data_Wrangle* or by directly accessing the database.

During the retrieval process, the Raman spectra are pre-processed as a batch through the pre-processing class, which allows users to modify the parameters of the pre-processing algorithms.

When a user queries Raman data using *Data_Wrangle*, a series of data management steps are performed. First, the data are filtered to match specific groups of Raman spectra based on user input through the *condition* attribute. Summaries of the content within the database can be accessed using the *summary* method, which helps the user to set the condition statement. Next, the Raman frequencies of the spectra are inspected for starting position, (see 3.2.4). The user is shown a summary of each unique Raman shift array found in the selected data using *x_summary*, including its frequency in the selected data. The user uses the information on the Raman shift profiles to determine which spectra cam be compiled together for data analysis through the *get_datadf* function. All spectra that have not been filtered out are then converted from individual two-dimensional arrays into a single two-dimensional *spectral matrix*, where each row is a unique spectra aligned under the Raman shifts (matrix headers). Notably, the index value of each row in the spectral matrix is the unique sample ID assigned to each spectrum upon its initial import into the database.

3.6.3 Dataset and Dictionaries

Dictionaries are a fundamental data structure in Python, organizing data through key-value pairs. These standardized structures enable the development of functions that work efficiently with consistent data formats, and are easy to view and edit by users.

The dictionaries used in *RamanRoot* fall into three broad categories (Figure 3.8). Trace dictionaries are used in *Spectracles*; the structure of trace dictionaries will not be described here, as trace dictionaries are already well established in Plotly packages and are only used to set preferences for interactive spectral figures.

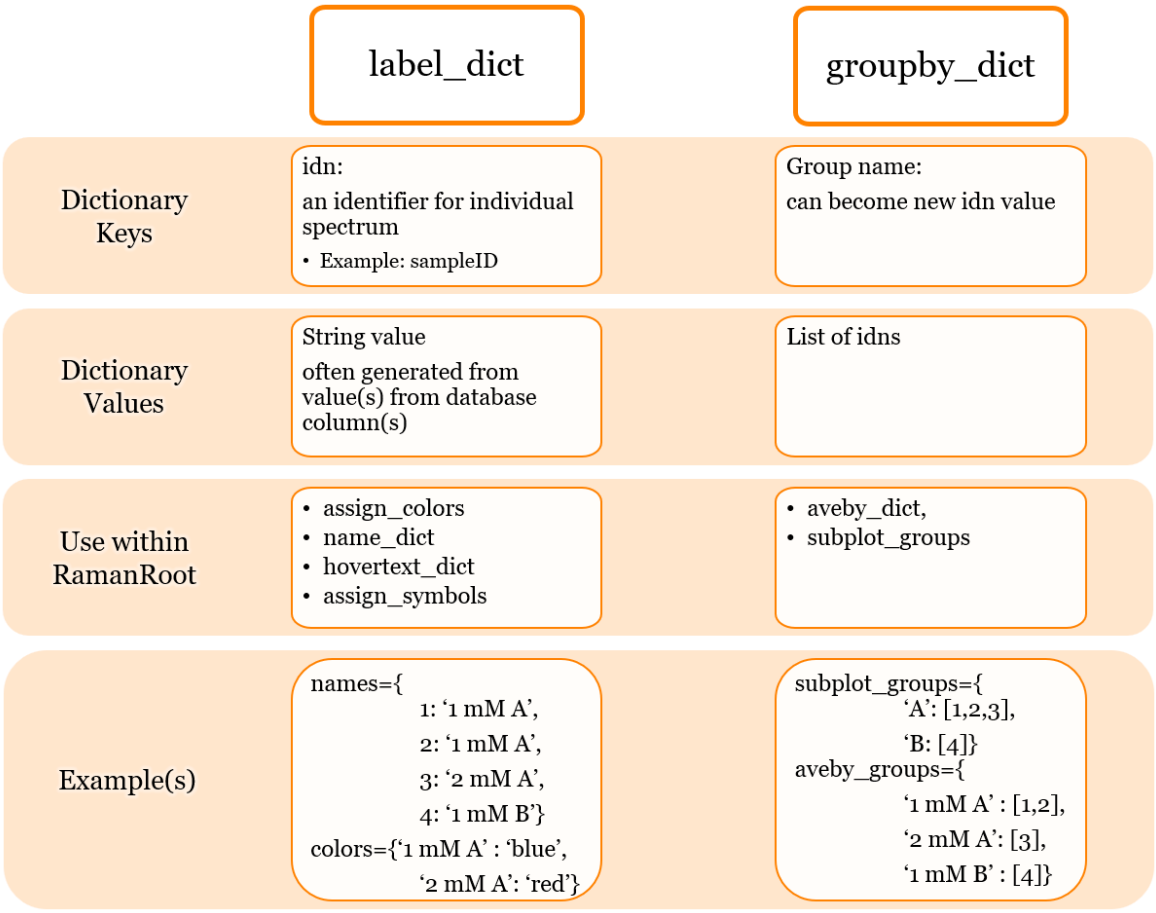


Figure 3.8: Summary of two types of dictionaries, utilized within *RamanRoot*

Labeling Dictionaries

When a Raman dataset is generated, the index is usually the sampleID. In order to assign proper names and labels for each spectrum, a labeling dictionary is used. The dictionary keys related to the identifier on the *spectral matrix*, such as the sampleID, while the value is a label. This dictionary format can be used to assign names to spectra on a figure, and is also used to set color, symbols, and other identifers when the spectra are plotted.

Group-by Dictionaries

Unlike a labeling dictionary that associates a single value with each key, a group-by dictionary contains a list of existing dictionary keys that can be used to call subsets of data. By assigning a new identifier to each group, subsets of data can be called quickly. This dictionary format is used to average spectra together and plot/compare spectra of the same group.

3.7 Tier: *Spectracles*

The *Spectracles* tier provides visualization functionalities for the *RamanRoot* package. Figures are generated using the 'plotly' package to produce interactive spectral figures. *spectral matrix*s and dictionaries generated in *Data_Wrangle* can be used as direct inputs for *Spectracles*.

3.7.1 Motivation

When plotting Raman data, each line on the figure is considered a *trace* and has a unique identifer(spectrum key). Both discrete and composite spectra can be traces. During visual analysis of Raman spectra, a user may want to view discrete spectra to identify potential outliers or anomalies. Most often, spectra are viewed as composite spectra, averaged together based on groupby-type dictionaries, which reflect categories

logged in the SQL database. Dictionaries are used to transfer the grouping information between the database and *Spectracles*.

3.7.2 Requirements

Technically, the only requirement to generate a basic figure in *Spectracles* is the dataset. However, each row of the dataset will be indiviually plotted separately and is rarely useful. Useful arguments for *Spectracles* are as follows:

- **Dataset**: comprised of a Raman shift array, and individual spectra identified by their index
- **Averageby dictionary**: a groupby-type dictionary
- **Plotby dictionary**: a groupby-type dictionary, each key indicating a subplot name, with values being a list of spectra keys of spectra to display on the subplot.
- **Namespec dictionary**: a label-type dictionary, where each key is an spectrum key, which can be directly or indirectly mapped into the *spectral matrix*. The values are names or labels that can be used to identify groups.
- **Order list**: a list containing spectra keys or group names, specifying the order in which the traces or subplots should be placed.

3.7.3 Initial Figure Object

When a figure object is initialized with *Spectracles*, the tier will use the input information to determine which figures are possible and display the information to the user. Figures can be single plots (having only one set of x and y axes) or multiplots which have multiple x/y axes in the figure.

3.7.4 Averaging

The *ave_spec* class is used to average spectra and show the results along with standard deviation. It requires a groupby-type dictionary where values are lists of spectra keys to be averaged together, the resulting spectra are identified by the dictionary key for that particular group. Multiple averageby dictionaries may be used to incrementally group data together.

The output of the ave_spec class is a dictionary, the keys set by the keys of the input groupby dictionary and the values containing the averaged intensity array and the corresponding standard deviation. When a figure is created using *ave_spec*, the user can specify whether or not to show the standard deviation on the plot.

3.7.5 Generating Figures

Single plot figures are generated by calling the *single_plot* function. This function can take various arguments, such as the dataset, namespec dictionary, and order list, to customize the plot accordingly. Similarly, multiplot figures can be generated using the *multiplot* function, providing more flexibility in visualizing multiple spectra at once.

3.8 Tier: *Analyst*

Analyst offers tools for multivariate and statistical analysis of Raman spectroscopic data. It enables users to efficiently implement analytical methods, adjust parameters, and explore the results.

It should be noted that smoothing filters (section 3.3.3) may introduce new patterns into spectra, leading to errors when smoothed spectra are used in multivariate analysis.[1] Spectra pre-processed without a smoothing filter may be used as the input dataset for multivariate analysis to avoid reduced performance of the analysis.

3.8.1 Object Initialization

Using *Analyst*, an object is generated for a given data set and analysis, from which the parameters for the analysis are displayed and can be adjusted. The tier is designed to easily integrate with Sklearn and SciPy[11] packages, essentially acting as an interpreter to apply the complex algorithms of these packages to the Raman datasets. The results of the analysis will be displayed in the Python interpreter as text or as an interactive figure. A save function can be called at any time to save the figures of the analysis when applicable.

3.8.2 PCA

Principal component analysis (PCA) is an unsupervised multivariate analysis technique commonly applied to Raman spectroscopic data to elucidate patterns and differences in complex spectral data. PCA reduces the dimensionality of the Raman spectra, which typically involves over a thousand data points per spectrum, based on the variance between samples. The array of Raman shifts is transformed into a set of variables called principal components (PCs). The number of PCs is determined by a PC number parameter. The PCs represent the most significant patterns of spectral variation, with PC 1 capturing the most variance, followed by subsequent PCs with decreasing variance.[9]

PCA is particularly useful for identifying spectral features and detecting patterns related to characteristics of the sample or differences between sample groups. These features are studied by viewing the loadings of the PCs. Loadings reflect the spectral patterns and contributions of the individual Raman shifts to each PC. The loadings of a PC consists of the Raman shift values and the weight each Raman shift had for the given PC. The direction and magnitude of these scores represent the influence each Raman shift had on the data. The Raman dataset PCA object can be set to return two- and three-dimensional plots of specified PCs, along with subplots of the corresponding PC loadings.

3.8.3 Leave One Out Cross-Validation

Leave One Out Cross-Validation (LOOCV) is used to measure the performance and accuracy of a model. Data from multivariate analysis, such as PCA, is used to create a predictive model via linear regression. To test the model, LOOCV divides the dataset into subsets, with each subset containing all but one data point. These data are used to build a linear regression model. The data point which was left out is then used as a test for this model. A mean squared error (MSE) is calculated based on the results and is a measure of success for the model. This process is repeated through the entire dataset, until all PCs have been used as a test. The MSE values for each iteration of the process are averaged and the value is used to determine the overall accuracy of the model, with low scores showing high accuracy.

Since the datasets used within *Analyst* are standardized, the LOOCV function can be applied to any data set, including the original input data set. However, due to the high level of computing power needed to perform the LOOCV algorithm, it is recommended to only be used on dimensionally reduced datasets, such as PCA.

3.9 Future Development

As code is constantly evolving, it is important to keep track of improvements that can be made to the code in the future. The *RamanRoot* tiers were designed with the goal of eventually developing into a web application. This application would access the data contained in an SQL database and generate interactive figures in response to user selections and prompts. A future improvement to the *Data_Wrangle* tier would be to enable comparisons between spectral 'frames' to better detect and remove cosmic rays. The challenge of varying starting points for day-to-day measurements may partially be resolved by integrating peak-fitting algorthims to better compile Raman spectra into the simplified matrix needed to perform statistical analysis.

References

[1] Shuxia Guo, Jürgen Popp, and Thomas Bocklitz. Chemometric analysis in raman spectroscopy from experimental design to machine learning–based modeling. *Nature Protocols*, 16:5426–5459, 2021. URL https://doi.org/10.1038/s41596-021-00620-3. 29, 30, 33, 37, 48

[2] Darren A. Whitaker and Kevin Hayes. A simple algorithm for despiking raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 179:82–84, 2018. ISSN 0169-7439. URL https://doi.org/10.1016/j.chemolab.2018.06.009. 34

[3] Uckert K, Bhartia R, and Michel J. A semi-autonomous method to detect cosmic rays in raman hyperspectral data sets. *Applied Spectroscopy*, 73(9):1019–1027, 2019. doi: doi:10.1177/0003702819850584. URL https://doi.org/10.1177/0003702819850584. 34

[4] Hause Lin. Data science: Use median absolute deviation instead of z-score to detect outliers, September 2019. URL https://hausetutorials.netlify.app/posts/2019-10-07-outlier-detection-with-median-absolute-deviation/#robust-solution-median-absolute-deviation. 36

[5] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median - sciencedirect, July 2013. URL https://www.sciencedirect.com/science/article/pii/S0022103113000668. 36

[6] Hao Chen, Weiliang Xu, and Neil G.R. Broderick. An adaptive and fully automated baseline correction method for raman spectroscopy based on morphological operations and mollification. *Applied Spectroscopy*, 73(3):284–93, 2019. URL https://doi.org/10.1177/0003702818811688. 36

[7] et al. Eilers, P. Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653, 2010. 36

[8] Donald Erb. pybaselines: A Python library of algorithms for the baseline correction of experimental data. URL https://github.com/derb12/pybaselines. 36

[9] Richard Brereton. *Chemometrics: Data Driven Extration for Science*. John Wiley Sons Ltd, 2018. 36, 49

[10] JAVIER FOCHESATTO and JAMES SLOAN. Signal processing of multicomponent raman spectra of particulate matter. *International Journal of High Speed Electronics*, 18(2):277–294, 2008. doi: 10.1142/S0129156408005345. URL https://doi.org/10.1142/S0129156408005345. 37

[11] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. 49

Chapter 4

Surfaced-Enhanced Raman Spectroscopy Combined with Multivariate Analysis for the Detection and Differentiation of Dopamine and DOPAC in Cerebrospinal Fluid

Surfaced-Enhanced Raman Spectroscopy Combined with Multivariate Analysis for the Detection and Differentiation of Dopamine and DOPAC in Cerebrospinal Fluid

A version of this chapter was originally published in *Journal of Raman Spectroscopy* by Paula Pimiento, Natalie Dunn, and Bhavya Sharma (https://doi.org/10.1002/jrs.6571). Paula Pimiento and Natalie Dunn are joint 1st authors.

This article describes the detection and differentiation of dopamine and DOPAC in artificial cerebrospinal fluid with Surface Enhanced Raman Spectroscopy. Multivariate analysis (principal component analysis) was applied to the Raman spectra of various mixtures in solution. The article was edited and subjected to peer review.

4.1 Abstract

Dopamine (DA) is a vital neurotransmitter for several brain functions, including pleasure, memory, motivation, and movement control. Dopaminergic neuron degradation is correlated to neurological disorders such as Parkinson's disease, and changes in DA concentration are related to mental health disorders. Currently, the gold standard for monitoring DA concentrations in the brain is electrochemistry. However, while electrochemistry has high sensitivity, it has low specificity for molecules with similar molecular structures, such as DA and its metabolites. Here, we demonstrate that surface-enhanced Raman spectroscopy (SERS), a vibrational spectroscopy technique that provides molecule-specific information with excellent sensitivity, requiring small sample sizes/volumes and is non-destructive, making it optimal for the detection and differentiation of DA and its metabolites. We show that SERS combined with multivariate analysis allows for the detection and differentiation of DA and its metabolites, including 3-O-methyl dopamine (3-OMD), homovanillic acid (HVA), 3,4-dihydroxyphenylacetic acid (DOPAC), and 3-methoxytyramine (3-MT) at physiologically relevant concentrations in artificial cerebrospinal fluid (aCSF).

4.2 Introduction

Dopamine (DA) is a vital neurotransmitter (NT) that plays a significant role in general brain function, especially memory, motor skills, pleasure, and reward.[1] Changes in DA concentrations within the brain are linked to various major neurological issues ranging from neurodegenerative disorders[2] (NDs) to mental health disorders, including depression, addiction, schizophrenia, bipolar disorder, and attention deficit disorder (ADHD). Direct detection of dopamine is challenging, partly due to the presence of dopamine metabolites, which can confuse the measurement of dopamine.[3–5]

Imaging techniques used to monitor dopamine involve indirect detection through secondary molecules to trace dopamine through fluorescence, magnetic resonance imaging,

and positron emission tomography.[6–10] Drawbacks of these techniques include complexity, overlap with other molecular signals, and labeling of traces with either fluorescent or radioactive probes.[7,11] High-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography (UPLC) combined with mass spectrometry (MS) are both commonly utilized in dopamine detection,[12–14] to assess complex biological matter without sacrificing the sensitivity and specificity that is required to analyze neurotransmitters and their metabolites; however, these are time-consuming and require expensive and bulky instrumentation.

Electrochemistry, specifically fast-scan cyclic voltammetry, is the gold standard for the sensitive and selective detection of electroactive neurochemicals.[15–17] Recent efforts have been aimed at improving the sensitivity, selectivity, and biocompatibility of electrochemical sensors.[18–21] Further, electrochemical detection methods have been paired with HPLC by modifying electrodes with molecules that have a specific affinity for dopamine to improve sensitivity and selectivity.[22,23] The combined electrochemical-HPLC method is very sensitive and selective; however, it requires considerable sample preparation. The primary challenge for electrochemical measurements is that dopamine and its metabolites[24] (Figure 4.1), 3,4-dihydroxyphenylacetic acid (DOPAC), 3-O-methyldopamine (3-O-MD), homovanillic acid (HVA), and 3-methoxytyramine (3-MT), have very similar chemical structures, differing by one substituent group, which leads to similar redox potentials, resulting in overlapping CVs. While HVA, 3-O-MD, and 3-MT are at lower concentrations (10s of nanomolar) than dopamine (1-3 µM), DOPAC is 3-6 times more concentrated (7-18 µM) than dopamine (Table **??**);[24–26]; thus, when DA and the precursors and metabolites are combined in a mixture, deconvoluting the individual contributions to the oxidation peak is challenging. Creative solutions such as the use of polymers, pH changes, complex electrodes, and other methods have been developed to isolate dopamine from its metabolites,[24,27,28] however, some of these solutions are not practical for direct measurement and quantification of dopamine. Thus, there is a pressing

need for sensing technologies that are rapid, label-free, sensitive, and selective for the detection of dopamine.

Our approach to addressing this need for sensing technologies to differentiate between dopamine and its metabolites is the use of surface-enhanced Raman spectroscopy (SERS) combined with multivariate analysis. SERS is a highly sensitive vibrational spectroscopy technique that yields fingerprint-like spectra based on the vibrational signatures of a sample. It can be applied to targeting molecules in complex biological media with little to no sample preparation and small sample sizes/volumes.[29–32] SERS harnesses the electric field generated through the localized surface plasmons of noble metal nanoparticles to enhance the Raman scattering signal by up to $\sim 10^8$,[33,34] allowing for ultra-low concentrations of analytes to be detected. The low detection limits enabled by SERS allowed for progress in the application of SERS for biosensing and quantification.[35–44] Our group previously applied SERS towards characterizing and differentiating between seven neurotransmitters: dopamine, epinephrine, norepinephrine, melatonin, serotonin, γ-aminobutyric acid (GABA), and glutamate at physiologically relevant concentrations (down to 100 nM concentrations)[45] and detection of the hormone cortisol to a limit of detection of 125 nM in complex media.[46] Here, we will push the limits of SERS further. While differentiating between the various metabolites of dopamine is important, the key challenge arises in the detection and identification of DA and DOPAC in a mixture because the concentration of DOPAC in the brain is orders of magnitude larger than the other metabolites (7-18 µM vs 10-100 nM; Table **??**). DOPAC is also found at higher concentrations than DA, thus further confounding the separation of these molecules with electrochemistry. In terms of chemical structure, DOPAC (3,4-dihydroxyphenylacetic acid) differs from DA (3,4-dihydroxyphenethylamine) by a single carboxylic acid functional group. We demonstrate here that SERS allows us to detect and differentiate between DA and DOPAC at physiologically relevant concentrations in artificial cerebrospinal fluid.

4.3 Methods

4.3.1 Materials

Phosphate buffered saline (PBS) was purchased from VWR. Citric acid (trisodium salt dehydrate, 99%), 3-O-methyldopamine hydrochloride (98%), dopamine (hydroxytyramine hydrochloride, 99%), 3-methoxytyramine hydrochloride (99%), 3,4-dihydroxyphenylacetic acid (98%), homovanillic acid (98+%), hydrogen tetrachloroauric (III) acid (HAuCl4), and perchloric acid ($HClO_4$) were purchased through Fisher Scientific. Artificial cerebrospinal fluid (aCSF) was purchased from Amazon. All glassware was cleaned in aqua regia and with Millipore water (18 MΩ·cm) prior to use. Millipore water was also used in all the experiments.

4.3.2 Cyclic voltammograms

Cyclic voltammograms (CVs) were collected using ceramic screen-printed electrodes (Pine Research Instrumentation, Durham, NC) with gold working (WE, 2 mm outer diameter) and counter (CE) electrodes and Ag/AgCl reference electrodes (RE). The ceramic gold electrodes (CGE) were used in conjunction with the Pine Research compact electrochemical cell and connected to the potentiostat (WaveNow Potentiostat with AfterMath Software, Pine Research Instrumentation Durham, NC). Analyte solutions were prepared in 0.1 M HClO4 in phosphate buffered saline (PBS). The potential scan rate was 100 mV/s.

4.3.3 Nanoparticle Synthesis

Gold nanoparticles (AuNPs) were synthesized following the Frens[47] method. Briefly, 50 mL of 10^{-2}% by weight tetrachloroauroric (III) acid trihydrate ($HAuCl_4 \bullet 3H_2O$) is heated to boiling. Then 0.3 mL of 1% by weight sodium citrate (Na_3-citrate) is added to the boiling solution. Within 40 seconds, the solution turns blue. After 140 seconds,

the solution turns red (with a yellow Tyndall Effect), indicating the formation of colloidal AuNPs, which are approximately 65 nm in diameter.

4.3.4 Sample Preparation

Stock solutions of DA, HVA, DOPAC, 3-O-MD, and 3-MT at 1 mM concentration were prepared in either 0.1 M $HClO_4$ in PBS for comparison with electrochemical data or in artificial CSF (aCSF). The 1 mM stock solutions were agitated on an orbital shaker (MaxQ 2000 Orbital Shaker, Thermo Scientific) at 300 RPMs for 10 minutes to fully dissolve solid materials. Serial dilutions were carried out to obtain the desired concentrations. 2 mL aliquots of AuNPs were centrifuged at 4200 RPMs for 15 minutes, and the supernatant was removed (Sorvall ST8R microcentrifuge, Thermo Scientific). The AuNPs were resuspended in 200 µL of each neurochemical solution at the desired concentration ranges. SERS spectra were collected immediately.

We collected SERS spectra of DA, HVA, DOPAC, 3-O-MD, and 3-MT at concentrations similar to those used for the electrochemical measurements for comparison (Figure A.1) and to establish the SERS limits of detection (Figures A.2 and A.3). For the detailed comparison of dopamine and DOPAC at physiologically relevant concentrations for the brain, twenty sample groups were prepared in aCSF, with 5 samples prepared at each concentration/mixture: 4 pure DA samples (1, 3, 7, 10 µM), 5 pure DOPAC samples (3, 7, 10, 12, 18 µM), 10 mixtures, and blanks (AuNPs in aCSF). Two mixtures were included that fall outside of the physiologically relevant range: 7 µM DA + 3 µM DOPAC and 10 µM DA + 10 µM DOPAC. The 7 µM DA + 3 µM DOPAC mixture was prepared to determine the effects on the SERS spectra when DA is present in a higher concentration than DOPAC. A summary of the samples can be found in the SI A.

4.3.5 Raman Instrumentation

SERS spectra were collected on a homebuilt confocal micro-Raman system. Raman scattering was excited using a 785 nm diode laser (IPS Lasers, New Jersey). The laser light

was guided through a series of optics before entering the back of an inverted Nikon Ti-U microscope. The Raman scattered light was collected in a 180° backscattering geometry. An ultrasteep long-pass edge filter (RazorEdge, Semrock) was used to reject the Rayleigh light, and the Raman scattered light was directed onto the 100 µM slit of an IsoPlane SCT-320 spectrometer (Princeton Instruments) equipped with a PIXIS 400 CCD camera (Princeton Instruments). A laser power of 10 mW was used to illuminate each sample for 30 seconds, and five replicate spectra were collected for each sample.

4.3.6 Raman Spectra Preprocessing

Raw Raman spectra were preprocessed with a home-written, Python-based batch preprocessing program. The program applies the same steps and parameters to each raw spectra and exports the result. Not only does this increase preprocessing speed, but it also removes the variability of preprocessing by hand. The program removes cosmic rays and the baseline drift inherent to all Raman spectra. The spectral baseline is generated from the pspline_asls algorithm in the Python package, PyBaselines. Finally, a 3rd order Savitzky-Golay filter with a window of 9 was applied for smoothing and generated by the popular SciPy package.

4.3.7 Multivariate Analysis

In this work, multivariate analyses, including principal component analysis (PCA), linear regression, and cross-validation leave-one-out analysis, were performed on 500 spectra from the DA/DOPAC dataset in Python using the Python sklearn package. Prior to the multivariate analyses, the data were also standardized using sklearn. In PCA, spectral data points (1340 data points for each Raman spectrum in this case) are deconvoluted by being assigned to a series of principal components (PCs), which each contain a spectral pattern (or loading) that is comprised of the points in the data set that indicate the largest variance, and a score that indicates the relationship between the sample and PC. PCs are plotted against each other in either two-or three-dimensional

(2D or 3D) plots. In these plots, the similar Raman spectral data will cluster together while dissimilar spectral data will separate in the plot. For the analyzed Raman dataset, 10 PCs were calculated.

Cross-validation was used to assess the prediction accuracy of a linear regression model built on the PCA data, consisting of the PC score for each sample for the 10 PCs calculated, using the leave one out method. From the PCA scores data set, a single sample was removed as the 'test' sample. Then, the remaining data was used to build a linear regression model. Cross validation was then applied, where the test sample was compared to the model. From this, a mean squared error (MSE) was calculated to determine the accuracy of the model. This process was repeated 500 times, one time for each spectrum in the data set.

4.4 Results and Discussion

Electrochemistry is the gold standard for in vivo electroactive neurochemical detection. While electrochemical methods offer sub-second time resolution, high spatial resolution, and excellent sensitivity, their selectivity is moderate. Electrochemical methods have been used for the detection of dopamine with limits of detection (LODs) in the low nanomolar ranges.[48] However, differentiation of dopamine from its metabolites, due to lack of selectivity in CV methods has proven challenging, as overlap of oxidation potentials commonly occurs with chemically similar molecules.[49] We collected individual cyclic voltammograms (CVs) of dopamine (DA) and its metabolites, 3-methoxytyramine (3-MT), 3-O-methyl dopamine (3-OMD), homovanillic acid (HVA), and 3,4- dihydroxyphenylacetic acid (DOPAC) at 100 µM concentrations (figure 4.1). The CV data show the considerable overlap between the oxidation peaks for all five molecules.

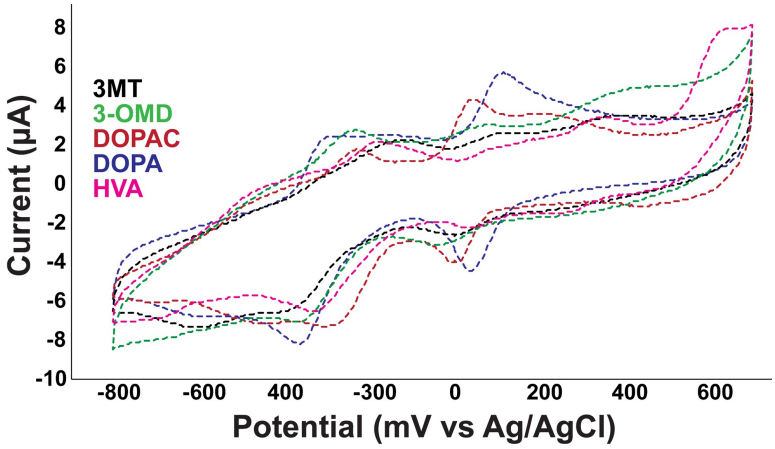


Figure 4.1: Cyclic voltammogram of 100 µM 3-MT (black), 3-OMD (green), DOPAC (red), DOPA (blue), HVA (pink) in 0.1 M $HClO_4$ and PBS. The potential was scanned from -0.80 V to 0.75 V at a 100 mV/s scan rate.

Table 4.1: SERS peak assignments for DA (orange), DOPAC (green), DA/DOPAC mixtures (purple), and blanks (AuNPs in aCSF, gray).

DA	DOPAC	Mixture	Blank
475	476	476	486
505	499	505	
535			
588	577	591	563
			648
670	670	670	
755	755	755	
			838
			1021
1117			
1143			
			1219
	1259	1259	1250
1282			
1404			
1424	1430	1430	
1479	1477	1477	

The results demonstrate that the limit of detection (LOD) of SERS is similar to that of electrochemical methods and, in combination with multivariate analysis, SERS allows for differentiation between DA and DOPAC. We also established SERS detection of dopamine and its metabolites at a concentration of 10 nM in 0.1 M $HClO_4$ and PBS (Figure A.1). At this concentration, which is equivalent to the LODs established by electrochemistry, we show distinct SERS vibrational modes for each analyte. Note that the SERS spectra of the metabolites in 0.1 M $HClO_4$ and PBS are shifted in position when compared to the SERS spectra in aCSF, due to solvent interactions with the molecules of interest and the AuNPs.

SERS spectra of twenty samples of DA, DOPAC, and mixtures in artificial CSF were collected, with twenty-five spectra per sample. After preprocessing, each group of 25 spectra were averaged and plotted together for comparison (Figure 4.2). The Raman peaks present in the SERS spectra of the blanks (648, 1021, and 1219 cm^{-1}) arise primarily from citrate anions on the surface of the gold nanoparticles.[50] The Raman peaks from the blanks are negligible in the DA and DOPAC solution spectra, allowing for characterization of the SERS spectra of the analytes without interferents. For the DA spectra (orange), the vibrational modes observed are: 588, 670, 1117, 1143, 1282, 1424, and 1479 cm^{-1}.46, 52-54 Similar Raman peaks are found in the SERS spectra of DOPAC (green), with slight shifts in peak position: 591, 670, 1282, 1430, and 1477 cm^{-1}. The minor changes in peak positions for the DA versus DOPAC spectra are expected based on the little difference in chemical structure between the two molecules. The DA/DOPAC mixture SERS spectra (blue) are comprised of peaks from both DA and DOPAC, as well as slight shifts in the peak positions: 591, 670, 1283, 1430, and 1477 cm^{-1}. Through visual inspection, the SERS spectra of pure DA, pure DOPAC, and their mixtures are similar and almost identical; there are, however, minor changes to the shape of the broad Raman peaks evident in the regions centered around 500, 580, 1270, and 1430 cm^{-1} (Table 4.1). We see that the DA spectra show a clear concentration dependence, while the DOPAC spectra do not show a similar dependence. We attribute the distinct DA concentration dependence to stronger

interactions between the DA and the AuNP, while DOPAC has a weaker interaction with the AuNP surface. To further elucidate the differentiation and concentration dependence of DA and DOPAC, we applied multivariate analysis to aid in decomposition of the complex SERS spectra.

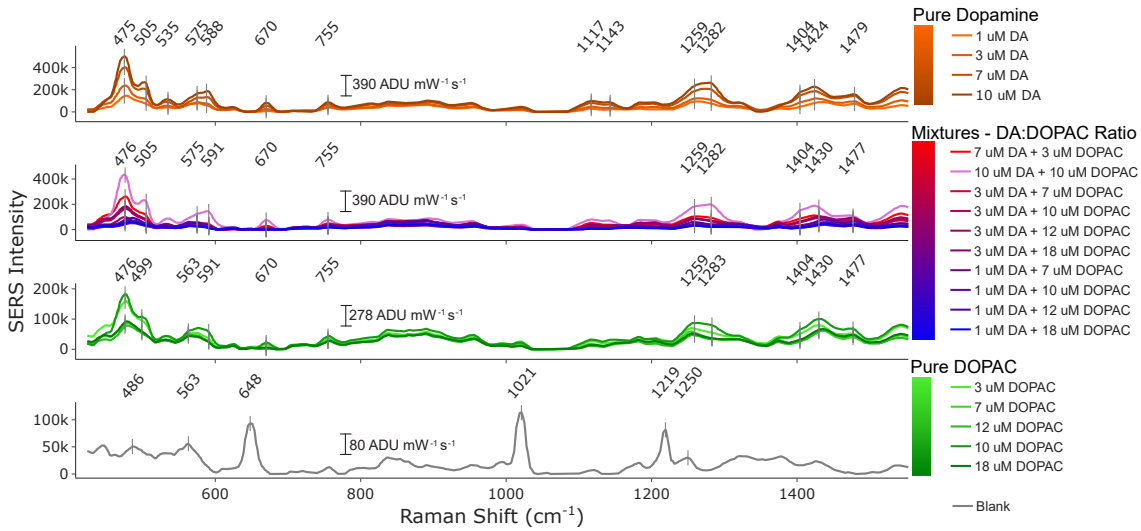


Figure 4.2: Averaged SERS spectra of DA, DOPAC, and DA/DOPAC mixtures and blanks. Pure DA SERS spectra are shaded orange and pure DOPAC SERS spectra are shaded green, both with the darkest color representing the greatest analyte concentration. Mixture SERS spectra are colored on a red-blue gradient based on the ratio of DA to DOPAC in the mixture: greater ratio being red and lowest ratio being blue. λ_{ex} = 785 nm, P= 10 mW, t= 30 sec.

We applied principal component analysis (PCA), an unsupervised multivariate technique, towards the classification of the SERS spectra from DA, DOPAC, and DA/DOPAC mixtures. The PCA analysis resulted in ten principal components (PCs) from the data; however, PCs 1 to 3 accounted for more than 90% of the variance shown in the data and were thus chosen for further analysis. Cross-validation using the leave-one-out (LOOCV) approach was performed on the first 3 PCs of the PCA dataset as described in the Methods section. From this, a mean squared error (MSE) value of 0.2525 and a root mean squared error (RMSE) value of 0.5025 were calculated for the entire dataset. The MSE value reflects the magnitude of error in the dataset through comparison of the observed (measured) and predicted values. If a model had no error, the MSE value would be equal to 0. The larger the error, the MSE values increase. Similarly, the RMSE value is based on the difference between the observed and predicted values and is used to assess the performance of the overall model for the entire dataset. Both the MSE and the RMSE demonstrate that there is good agreement between the linear regression model and the Raman dataset.

PCs 1 through 3 were plotted on a three-dimension plot to aid with visualization of any patterns that arise (Figure 4.3). The 3D PCA plot shows the SERS spectra cluster in four broad groups: (1) blank samples, which are comprised of the AuNPs plus CSF (gray), (2) pure DOPAC at concentrations of 3-18 µM (greens), (3) mixtures of low concentration DA (1-3 µM) and high concentration DOPAC (12-18 µM, blue to dark purple); (4) pure DA at concentrations of 1- 10 µM (orange), along with mixtures of high concentration DA (3-10 µM) and lower concentration DOPAC (3-10 µM, purple to red). Also included in the fourth cluster is an equimolar mixture of DA and DOPAC at 10 µM (pink). The 3D PCA shows that for SERS spectra of mixtures of DA and DOPAC, where the ratio of DA to DOPAC is higher, cluster towards the pure DA data. When the ratio of DA to DOPAC is lower, the SERS spectra of the mixtures is in-between the pure DA and pure DOPAC data, even at ratios of 1 µM DA to 18 uM DOPAC.

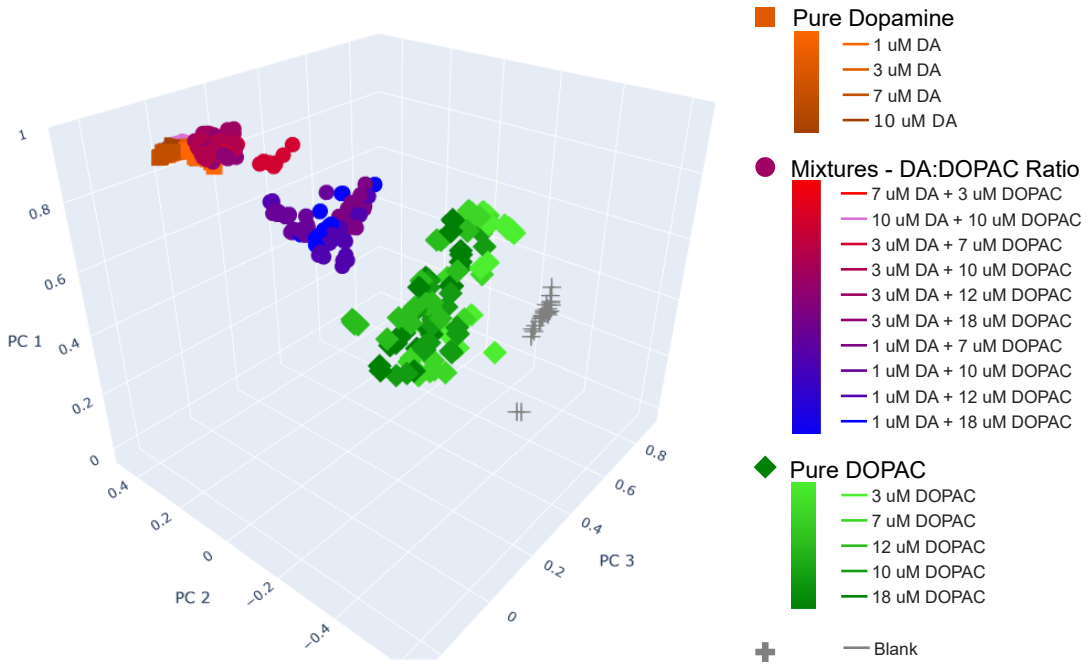


Figure 4.3: 3-dimensional PCA plot of PC 1 vs. PC 2 vs. PC3. Pure DA SERS spectra are represented by the orange shades and pure DOPAC SERS spectra are represented by the green shades, both with the darkest color representing the greatest analyte concentration. DA/DOPAC mixture SERS spectra are colored on a red-blue gradient based on the ratio of DA to DOPAC in the mixture; greater ratio being red and lowest ratio being blue. The SERS spectra for the blanks are shown as gray crosses. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec.

For an in-depth analysis of the contributions of the PCs, as well as to further extract information about the differentiation of DA and DOPAC through clustering and the spread pattern, we plotted 2D PCA plots (Figures 4.4, S5, S6). While the comparison of PC 1 vs. PC 3 and PC 2 vs. PC 3 both provide additional information, the 2D PCA plot that most clearly demonstrates the clustering and spread of the DA and DOPAC data is the comparison of PC 1 vs. PC 2. PC 1 reflects 73% of the variance between the samples, arising from contributions of the blanks (AuNPs + aCSF), as well as Raman peaks from DOPAC and the DA/DOPAC mixture (see PC 1 loading; Figure 4.4). Along PC 1, the variance primarily distinguishes the blank from the DA/DOPAC solutions. We also see separation between pure DOPAC samples and those which contain DA along PC 1; however, the spread is limited and does not resolve pure DA from the mixtures. The greatest resolution between sample groups is seen in PC 2 (Figure 4.4), which reflects 13% of sample variance. When examining the DA/DOPAC samples along PC 2, four distinct groups are seen (moving from negative to positive values): pure DOPAC, low DA:DOPAC ratio mixtures, high DA:DOPAC mixtures, and pure DA. Solutions of pure DA score near 0.4 along PC 2 and pure DOPAC score between -0.4 and -0.6. Mixtures of DA and DOPAC have PC scores between the pure solutions, with those containing 3 µM of DA scoring closer to pure DA and mixtures containing 1 µM of DA (a lower ratio of DA:DOPAC) scoring below zero. Peaks at 471, 673, 1278, and 1483 cm^{-1}) present in the loadings of PC 2 and are also found in both DA and DOPAC SERS spectra. Raman peaks shifts distinct to the DA Raman spectra at 595, 1113, and 1406 cm^{-1} are also found in the PC 2 loadings. We find that through the 2D PCA plot of PC 1 vs. PC 2, the samples are separated along PC 2 by concentration, from pure DA (high to low concentrations), through the mixtures (also separated by concentration), through to pure DOPAC (low to high concentrations). The SERS peaks contributing to each loading are outlined in Table 4.2.

Table 4.2: SERS peak positions from the loadings of PCs 1-3 with colors indicating the SERS spectra from which they arise: DA (orange), DOPAC (green), DA/DOPAC mixtures (purple), blanks (AuNPs in aCSF, gray).

PC 1	PC 2	PC 3
478	471	
577	595	
		649
	673	
838		
890		
955		
		1021
	1113	
		1219
1258		
	1278	
		1382
	1406	
1438		
	1483	

The two mixtures,10 µM DA + 10 µM DOPAC (light pink) and the 7 µM DA + 3 µM DOPAC (raspberry pink, Figure 4.3), with concentrations that would not be found in the healthy brain vary slightly in this trend, where the pattern of concentration dependence is not followed along PC 2. The 10 µM equimolar mixture is the only mixture which scores above pure DA on PC 2, indicating that with equivalent molar concentrations, the mixture more highly resembles DA than DOPAC. The 7 µM DA + 3 µM DOPAC mixture also clusters with the pure DA solutions. One explanation for this difference is that the amine group on the DA has a higher affinity for the negatively charged (citrate reduced) AuNP surface than the charged carboxylic acid, resulting in increased DA SERS signal. It has been previously established in the literature that nitrogen groups, particularly amines, interact strongly with gold *d* and *s* electrons, resulting in a high affinity of the nitrogen for the gold surface.

For further analysis of the trends revealed in the clustering of the SERS data along PC 2, we zoomed into the 2D PCA plot of PC 1 (region of 0.8-1) vs. PC 2 (region of 0.15 – 0.5) and hid the mixture data to examine the concentration dependence of DA (Figure 4.5). The 2D PCA shows that SERS spectra for high concentration solutions of DA (7 or 10 µM DA) scored above 0.4 on PC 2. SERS spectra for samples with 3 µM DA scored between 0.3 and 0; all solutions with 1 µM DA or no DAs scored below zero. If we isolate the SERS spectra data for the mixtures (hide the pure DA data, zoom into PC 1 region 0.86 to 0.96 and PC 2 region -0.4 to 0.5), the 10 µM DA + 10 µM DOPAC mixture shows that Raman spectroscopy is highly sensitive to dopamine in aCSF, despite having equal concentrations of DA and DOPAC, the sample clusters entirely in the space where the pure DA spectra would be, followed closely by the 7 µM DA + 3 µM DOPAC mixture data. Viewing the pure samples and mixtures separately allows for the clustering and spread of the pure DA and DA/DOPAC mixture data to show a concentration gradient from high to low along PC 2 (Figure 4.5).

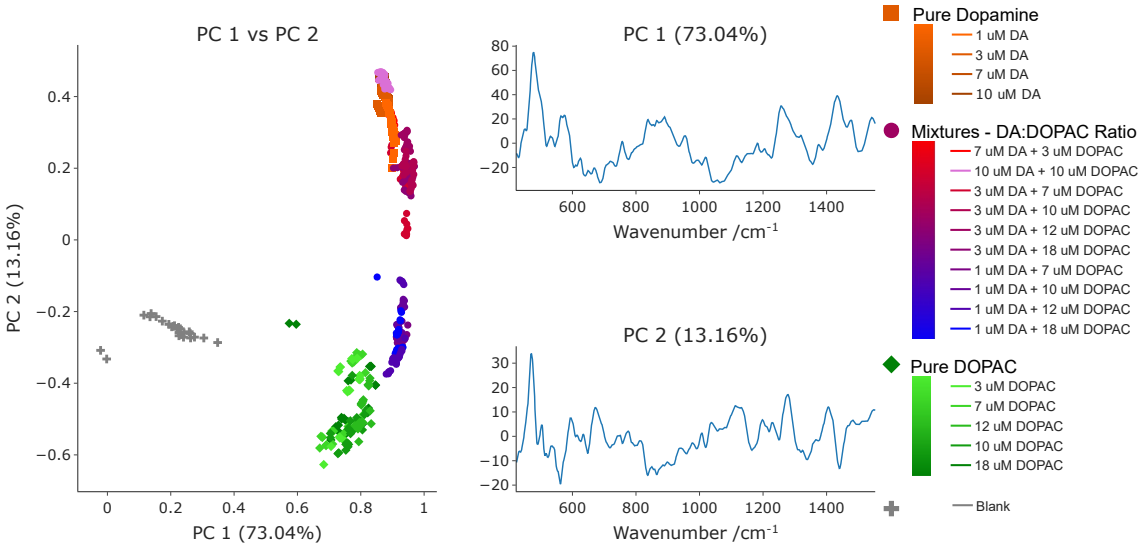


Figure 4.4: 2D PCA plot of PC 1 vs. PC 2 of the SERS spectra of DA, DOPAC, DA/DOPAC and blank solutions. Pure dopamine solutions are marked with an orange square (■), pure DOPAC solutions are marked with a green diamond (♦), mixtures are marked with a circle (●) and the blanks are marked with grey crosses (+). A red to blue gradient was applied to the mixtures depending on the ratio of DA to DOPAC in the solution, with red having the highest ratio and blue the lowest. Pink represents two mixtures that are not naturally found in the brain 10 µM DA + 10 µM DOPAC (light pink) and 7 µM DA + 3 µM DOPAC (raspberry pink). The loadings for PC 1 and PC 2 are included. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec.

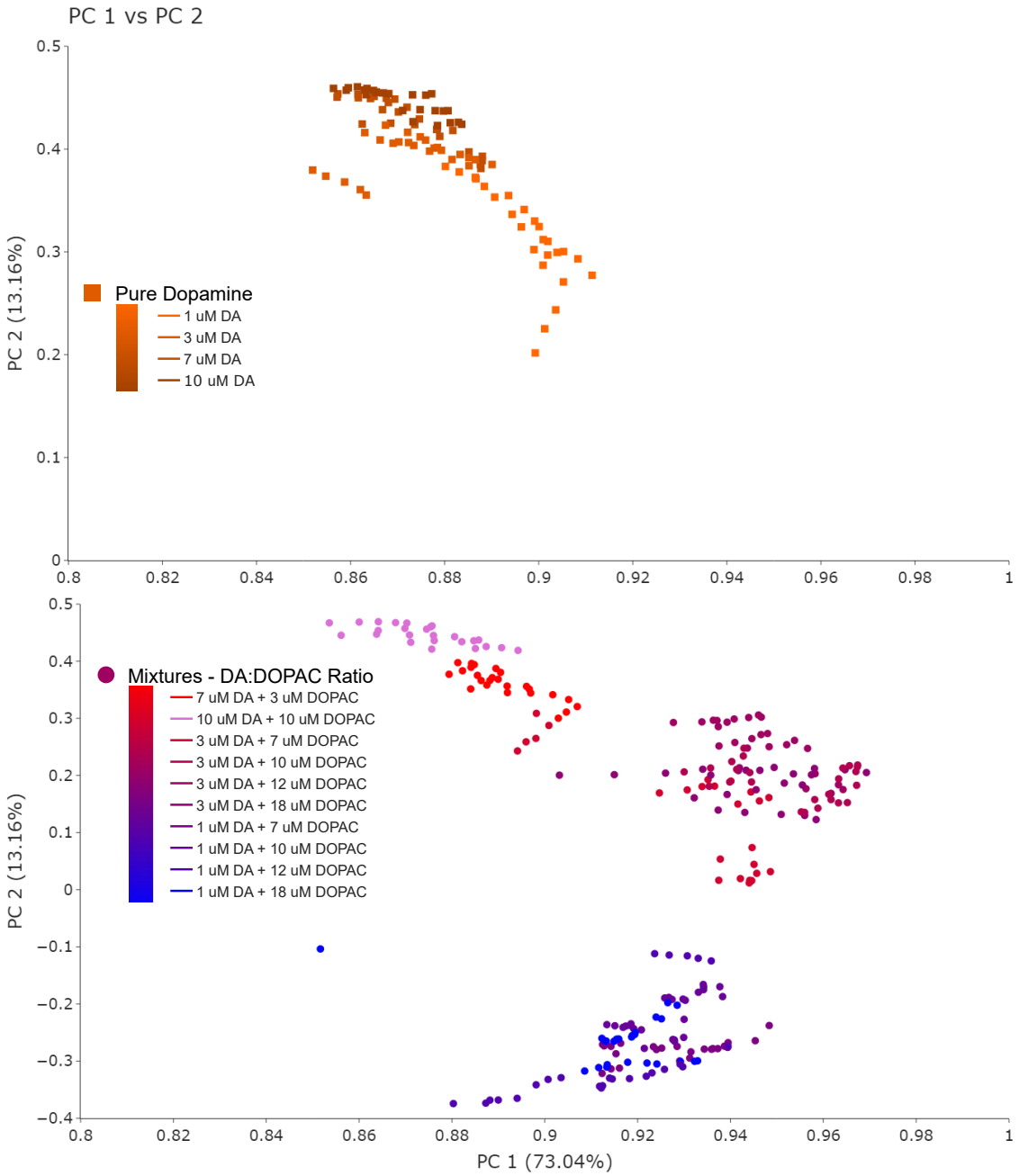


Figure 4.5: 2D PCA plot of PC 1 vs PC 2 zoomed in to highlight the concentration dependence of the pure DA samples (top, orange) and DA/DOPAC mixtures (bottom, pink to blue). In both cases, the samples separate along a gradient from highest DA concentration to lowest DA concentration. λ_{ex} = 785 nm, P = 10 mW, t = 30 sec.

Pure DOPAC solution SERS spectra are clearly separated from the mixtures (Figure 4.4, green); this distinction is not as noticeable between pure DA and mixtures. We attribute these differences in clustering behavior to the lower concentration range for the DA in the mixtures (1-3 µM) versus the higher concentration range (7-18 M) for DOPAC. 2D PCA plots allow for DA concentrations to be distinguished. The DOPAC SERS spectra do not separate along the concentration range as well as the DA, which indicates that Raman spectroscopy is more sensitive to DA than DOPAC. We attribute this to the amine group on the DA molecule interacting with the gold surface of the nanoparticles with higher affinity than the carboxyl group of DOPAC, resulting in better sensitivity to DA in solution. This higher affinity of DA to the AuNPs is a distinct advantage of SERS over electrochemistry. Even at lower physiological concentrations than DOPAC and in mixtures, the DA SERS spectra have higher intensity. Combining SERS with multivariate analysis allows for the differentiation of the SERS spectra by concentration.

4.5 Conclusion

We have established the detection and differentiation of DA and its metabolites at a limit of detection of 10 nM in 0.1 M $HClO_4$ and PBS solutions, which served as the supporting electrolyte for CV experiments. While the electrochemistry experiments provided the appropriate reduction and oxidation potentials for these neurochemicals, the CVs were overlapped, and deconvoluting the individual contributions to the oxidation peak is challenging. SERS was applied towards detecting and differentiating between DA and its metabolites in 0.1 M $HClO_4$ and PBS solutions. Focusing specifically on discriminating between DA and the high concentration metabolite DOPAC, we identified vibrational modes for DA, DOPAC, and DA/DOPAC mixtures in the SERS spectra of these molecules at physiologically relevant concentrations in aCSF. Multivariate analysis, specifically PCA, further allowed for the classification of the DA, DOPAC, and DA/DOPAC mixtures by concentration in 2D PCA plots. We show that SERS

AuNPs have a higher affinity for the amine group of DA over the anionic DOPAC, providing advantages and increased sensitivity in molecular identification compared to other methods. We highlight here the potential of SERS as a stand-alone technique for the detection of neurochemicals and their metabolites in biological media.

Acknowledgements

Work on "Non-Invasive Spectroscopic Quantification of Depression Biomarkers" is supported by Wellcome Leap as part of the Multi-Channel Psych Program. We would like to also thank the University of Tennessee for funding. We acknowledge undergraduate students Lillian Mooney, Tarek Atyia, Jason Cornelio, and Christopher Quizhpe for their assistance in these experiments.

References

[1] Ja-Hyun Baik. Stress and the dopaminergic reward system. *Experimental Molecular Medicine*, 52(12):1879–1890, 2020. doi: 10.1038/s12276-020-00532-4. 55

[2] Valery L Feigin and et al. Global, regional, and national burden of neurological disorders during 1990–2015: A systematic analysis for the global burden of disease study 2015. *The Lancet Neurology*, 16(11):877–897, 2017. doi: 10.1016/s1474-4422(17)30299-5. 55

[3] Allison Basse-Tomusk and George V. Rebec. Regional distribution of ascorbate and 3,4-dihydroxyphenylacetic acid (dopac) in rat striatum. *Brain Research*, 538 (1):29–35, 1991. doi: 10.1016/0006-8993(91)90372-3. 55

[4] Mark R. Deakin, Paul M. Kovach, K. J. Stutts, and R. Mark. Wightman. Heterogeneous mechanisms of the oxidation of catechols and ascorbic acid at carbon electrodes. *Analytical Chemistry*, 58(7):1474–1480, 1986. doi: 10.1021/ac00298a046.

[5] Zijun Shao and B. Jill Venton. Different electrochemical behavior of cationic dopamine from anionic ascorbic acid and dopac at cnt yarn microelectrodes. *Journal of The Electrochemical Society*, 169(2), 2022. doi: 10.1149/1945-7111/ac4d67. 55

[6] Michael R. Post and David Sulzer. The chemical tools for imaging dopamine release. *Cell Chemical Biology*, 28(6):748–764, 2021. doi: 10.1016/j.chembiol.2021.04.005. 56

[7] Rajendra D. Badgaiyan. Detection of dopamine neurotransmission in "real time". *Frontiers in Neuroscience*, 7, 2013. doi: 10.3389/fnins.2013.00125. 56

[8] Sofia Elizarova, Abed Alrahman Chouaib, Ali Shaib, Björn Hill, Florian Mann, Nils Brose, Sebastian Kruss, and James A. Daniel. A fluorescent nanosensor paint detects dopamine release at axonal varicosities with high spatiotemporal resolution. *Proceedings of the National Academy of Sciences*, 119(22), 2022. doi: 10.1073/pnas.2202842119.

[9] Mareike Fauser, Manuel Ricken, Franz Markert, Nikolai Weis, Oliver Schmitt, Jan Gimsa, Christine Winter, Kathrin Badstübner-Meeske, and Alexander Storch. Subthalamic nucleus deep brain stimulation induces sustained neurorestoration in the mesolimbic dopaminergic system in a parkinson's disease model. *Neurobiology of Disease*, 156:105404, 2021. doi: 10.1016/j.nbd.2021.105404.

[10] Juan M. Rojas Cabrera, J. Blair Price, Aaron E. Rusheen, Abhinav Goyal, Danielle Jondal, Abhijeet S. Barath, Hojin Shin, Su-Youne Chang, Kevin E. Bennet, Charles D. Blaha, and et al. Advances in neurochemical measurements: A review of biomarkers and devices for the development of closed-loop deep brain stimulation systems. *Reviews in Analytical Chemistry*, 39(1):188–199, 2020. doi: 10.1515/revac-2020-0117. 56

[11] Tyler J. Bruinsma, Vidur V. Sarma, Yoonbae Oh, Dong Pyo Jang, Su-Youne Chang, Greg A. Worrell, Val J. Lowe, Hang Joon Jo, and Hoon-Ki Min. The relationship between dopamine neurotransmitter dynamics and the blood-oxygen-level-dependent (bold) signal: A review of pharmacological functional magnetic resonance imaging. *Frontiers in Neuroscience*, 12, 2018. doi: 10.3389/fnins.2018.00238. 56

[12] P. A. LeWitt, M. P. Galloway, W. Matson, P. Milbury, M. McDermott, D. K. Srivastava, and D. Oakes. Markers of dopamine metabolism in parkinson's disease. *Neurology*, 42(11):2111–2111, 1992. doi: 10.1212/wnl.42.11.2111. 56

[13] Chiara Muzzi, Erica Bertocci, Lucia Terzuoli, Brunetta Porcelli, Ilaria Ciari, Roberto Pagani, and Roberto Guerranti. Simultaneous determination of serum concentrations of levodopa, dopamine, 3-o-methyldopa and α-methyldopa by hplc. *Biomedicine amp;amp; Pharmacotherapy*, 62(4):253–258, 2008. doi: 10.1016/j.biopha.2007.10.018.

[14] Andrej Kovac, Zuzana Somikova, Norbert Zilka, and Michal Novak. Liquid chromatography–tandem mass spectrometry method for determination of panel of

neurotransmitters in cerebrospinal fluid from the rat model for tauopathy. *Talanta*, 119:284–290, 2014. doi: 10.1016/j.talanta.2013.10.027. 56

[15] Michael Heien, Amina Khan, Jennifer Ariansen, Joseph Cheer, Paul Phillips, Kate Wassum, and R Mark Wightman. Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. *Proc Natl Acad Sci USA*, 2005. doi: DOI: 10.1073/pnas.0504657102. URL https://pubmed.ncbi.nlm.nih.gov/16006505/. 56

[16] E. J.; Chang S.-Y.; Lujan-J. L.; Jang-D. P.; Bennet K. E.; Lee-K. H. Paek, S. B.; Knight. Dopamine measurement during prolonged deep brain stimulation: a proof-of-principle study of paired pulse voltammetry. *Biomed Eng Lett*, March 2013. doi: 10.1007/s13534-013-0086-y.

[17] J. G. Roberts, L. Z. Lugo-Morales, P. L. Loziuk, and L. A. Sombers. Real-time chemical measurements of dopamine release in the brain. *Methods Mol Biol*, 2013. doi: 10.1007/978-1-62703-251-3_16. 56

[18] Diane Miller and James O'Callaghan. Biomarkers of parkinson's disease: present and future. *Metabolism*, October 2014. doi: 10.1016/j.metabol.2014.10.030. 56

[19] Elaheh Farjami and Elena Ferapontova. Effect of the dna end of tethering to electrodes on electron transfer in methylene blue-labeled dna duplexes. *Langmuir*, November 2012. doi: 10.1021/la3032336. URL https://pubmed.ncbi.nlm.nih.gov/23106377/.

[20] E. Farjami, R. Campos, J. S. Nielsen, K. V. Gothelf, J. Kjems, and E. E. Ferapontova. Rna aptamer-based electrochemical biosensor for selective and label-free analysis of dopamine. *Anal Chem*, January 2012. doi: 10.1021/ac302134s. URL https://pubmed.ncbi.nlm.nih.gov/23210972/.

[21] J. Huang, Y. Liu, H. Hou, and T. You. Simultaneous electrochemical determination of dopamine, uric acid and ascorbic acid using palladium nanoparticle-loaded carbon

nanofibers modified electrode. *Biosens Bioelectron*, December 2008. doi: 10.1016/j.bios.2008.06.011. URL https://pubmed.ncbi.nlm.nih.gov/18640024/. 56

[22] H. Gu, E. L. Varner, S. R. Groskreutz, A. C. Michael, and S. G. Weber. In vivo monitoring of dopamine by microdialysis with one-minute temporal resolution using online capillary liquid chromatography with electrochemical detection. *Anal Chem*, June 2015. doi: 10.1021/acs.analchem.5b00633. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4835028/. 56

[23] L. Lin, J. Yang, R. Lin, L. Yu, H. Gao, S. Yang, and X. Li. In vivo study on the monoamine neurotransmitters and their metabolites change in the striatum of parkinsonian rats by liquid chromatography with an acetylene black nanoparticles modified electrode. *J Pharm Biomed Anal*, January 2013. 56

[24] R. H. Dopamine Synthesis-Uptake Metabolism Elsworth, J. D.; Roth and Receptors. Dopamine synthesis, uptake, metabolism, and receptors: relevance to gene therapy of parkinson's disease. *Exp Neurol*, March 1997. doi: 10.1006/exnr.1996.6379. URL https://pubmed.ncbi.nlm.nih.gov/9126143/. 56

[25] R. C. Pierce and G. V. Rebec. Dopamine-, nmda- and sigma-receptor antagonists exert differential effects on basal and amphetamine-induced changes in neostriatal ascorbate and dopac in awake, behaving rats. *Brain Res.*, May 1992. doi: 10.1016/0006-8993(92)90741-q.

[26] M. K. Zachek, P. Takmakov, J. Park, R. M. Wightman, and G. S. McCarty. Simultaneous monitoring of dopamine concentration at spatially different brain locations in vivo. *Biosens Bioelectron*, January 2010. doi: 10.1016/j.bios.2009.10.008. 56

[27] Y. Oh, M. L. Heien, C. Park, Y. M. Kang, J. Kim, S. L. Boschen, H. Shin, H. U. Cho, C. D. Blaha, K. E. Bennet, H. K. Lee, S. J. Jung, I. Y. Kim, K. H. Lee, and D. P. Jang. Tracking tonic dopamine levels in vivo using multiple cyclic square wave

voltammetry. *Biosens Bioelectron*, December 2018. doi: 10.1016/j.bios.2018.08.034. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6775780/. 56

[28] D. Raju, A. Mendoza, P. Wonnenberg, S. Mohanaraj, M. Sarbanes, C. Truong, and A. G. Zestos. Polymer modified carbon fiber-microelectrodes and waveform modifications enhance neurotransmitter metabolite detection. *Anal Methods*, March 2019. doi: 10.1039/c8ay02737d. URL 10.1039/c8ay02737d. 56

[29] J. H. An, D.-K. Choi, K.-J. Lee, and J.-W. Choi. Surface-enhanced raman spectroscopy detection of dopamine by dna targeting amplification assay in parkisons's model. *Biosens Bioelectron*, May 2015. doi: 10.1016/j.bios.2014.10.049. URL https://pubmed.ncbi.nlm.nih.gov/25465795/. 57

[30] L. Qin, X. Li, S.-Z. Kang, and J. Mu. Gold nanoparticles conjugated dopamine as sensing platform for sers detection. *Colloids Surf B Biointerfaces*, Feb 2015. doi: 10.1016/j.colsurfb.2014.12.028.

[31] L. Tang, S. Li, F. Han, L. Liu, L. Xu, W. Ma, H. Kuang, A. Li, L. Wang, and C. Xu. Sers-active au@ag nanorod dimers for ultrasensitive dopamine detection. *Biosens Bioelectron*, September 2015. doi: 10.1016/j.bios.2015.04.013. URL https://pubmed.ncbi.nlm.nih.gov/25880832/.

[32] V. Ranc, Z. Markova, M. Hajduch, R. Prucek, L. Kvitek, J. Kaslik, K. Safarova, and R. Zboril. Magnetically assisted surface-enhanced raman scattering selective determination of dopamine in an artificial cerebrospinal fluid and a mouse striatum using fe(3)o(4)/ag nanocomposite. *Biosens Bioelectron*, March 2014. doi: 10.1021/ac500394g. URL https://pubmed.ncbi.nlm.nih.gov/24555681/. 57

[33] Bhavya Sharma, Renee R. Frontiera, Anne-Isabelle Henry, Emilie Ringe, and Richard Van Duyne. Sers: Materials, applications, and the future. *Materials Today*, 15(1):16–25, 2012. ISSN 1369-7021. doi: https://doi.org/10.

1016/S1369-7021(12)70017-2. URL https://www.sciencedirect.com/science/article/pii/S1369702112700172. 57

[34] B. Sharma, K. Ma, M. R. Glucksberg, and R. P. Van Duyne. Seeing through bone with surface-enhanced spatially offset raman spectroscopy. *J Am Chem Soc*, November 2013. doi: https://doi.org/10.1021/ja409378f. URL https://pubs.acs.org/doi/10.1021/ja409378f. 57

[35] Alvarez-Puebla RA. Rodriguez-Lorenzo L, Fabris L. Multiplex optical sensing with surface-enhanced raman scattering: a critical review. *Anal Chim Acta*, October 2012. doi: 10.1016/j.aca.2012.08.003. 57

[36] DanDan Li, Guangchao Zheng, HaoWei Jia, and Jin Wang. Direct readout sers multiplex sensing of pesticides via gold nanoplate-in-shell monolayer substrate. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 451:48–55, 2014. doi: 10.1016/j.colsurfa.2014.03.034.

[37] Balasundaram G. Chang-Y. T.- Olivo-M. Dinish, U. S. Actively targeted in vivo multiplex detection of intrinsic cancer biomarkers using biocompatible sers nanotags. *Scientific Reports*, 4(4075), 2014. URL https://doi.org/10.1038/srep04075.

[38] Liu B-Ni H-Sun L-Su E Chen H Gu Z Zhao X Zhang D, Huang L. Quantitative and ultrasensitive detection of multiplex cardiac biomarkers in lateral flow assay with core-shell sers nanotags. *Biosens Bioelectron*, 2018. doi: 10.1016/j.bios.2018.01.062. URL https://pubmed.ncbi.nlm.nih.gov/29428590/.

[39] I.; Moulia V.; Elka A.; Gaitanis G.; Bassukas I. Kourkoumelis, N.; Balatsoukas. Advances in the in vivo raman spectroscopy of malignant skin tumors using portable instrumentation. *International Journal of Molecular Sciences*, 16(12):14554–14570, 2015. URL https://pubmed.ncbi.nlm.nih.gov/29428590/.

[40] B.-Y.; Jie L.-J.; Tian X.-D.; Yang Z.-L.; Radjenovic P. M.; Luo S.-Y.; Tian Z.-Q.; Li J.-F. Zhu, W.; Wen. Rapid and low-cost quantitative detection of creatinine in

human urine with a portable raman spectrometer. *Biosensors and Bioelectronics*, 154, 2020. doi: 10.1016/j.bios.2020.112067. URL https://pubmed.ncbi.nlm.nih.gov/32056962/.

[41] B.; König M.; Salehi-M.; Schlücker S. Tran, V.; Walkenfort. Rapid, quantitative, and ultrasensitive point-of-care testing: A portable sers reader for lateral flow assays in clinical chemistry. *Angewandte Chemie International Edition*, 58(2):442–446, 2019. URL https://doi.org/10.1002/anie.201810917.

[42] D.; Faulds K. Goodacre, R.; Graham. Recent developments in quantitative sers: Moving towards absolute quantification. *TrAC Trends in Analytical Chemistry*, 102: 359–368, 2018. URL https://doi.org/10.1016/j.trac.2018.03.005.

[43] Jacob E. Olson Xin Gu, Michael J. Trujillo and Jon P. Camden. Sers sensors: Recent developments and a generalized classification scheme based on the signal origin. *Annual Review of Analytical Chemistry*, 11(1):147–169, 2018. URL https://pubmed.ncbi.nlm.nih.gov/29547340/.

[44] S Laing, K Gracie, and K Faulds. Correction: Multiplex in vitro detection using sers. *Chemical Society Reviews*, 46(19):1901–1918, September 2017. URL https://doi.org/10.1039/c7cs90089a. 57

[45] A Moody and Bhavya Sharma. Multi-metal, multi-wavelength surface-enhanced raman spectroscopy detection of neurotransmitters. *ACS Chem Neurosci*, June 2018. doi: 10.1021/acschemneuro.8b00020. URL https://pubmed.ncbi.nlm.nih.gov/29601719/. 57

[46] T. Moore and B Sharma. Multi-metal, multi-wavelength surface-enhanced raman spectroscopy detection of neurotransmitters. *ACS Chem Neurosci*, April 2018. doi: 10.1021/acschemneuro.8b00020. URL https://pubmed.ncbi.nlm.nih.gov/29601719/. 57

[47] G Frens. Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions. *Nature Physical Science*, 241(105):20–22, 1973. doi: 10.1038/physci241020a0. URL https://doi.org/10.1038/physci241020a0. 58

[48] Richard B. Keithley, Pavel Takmakov, Elizabeth S. Bucher, Anna M. Belle, Catarina A. Owesson-White, Jinwoo Park, and R. Mark Wightman. Higher sensitivity dopamine measurements with faster-scan cyclic voltammetry. *Analytical Chemistry*, 83(9):3563–3571, 2011. doi: 10.1021/ac200143v. 61

[49] Y.-R. Kim, S. Bong, Y.-J. Kang, Y. Yang, R. K. Mahajan, J. S. Kim, and H. et al Kim. Electrochemical detection of dopamine in the presence of ascorbic acid using graphene modified electrodes. *Biosensors and Bioelectronics*, 25(10):2366–2369, 2012. 61

[50] D.-B. Grys, B. de Nijs, and A. R. et al. Salmon. Citrate coordination and bridging of gold nanoparticles: The role of gold adatoms in aunp aging. *ACS Nano*, (7): 8689–8696, 2020. 64

Chapter 5

SERS and Multivariate Analysis for Discrimination Between Neurotoxic Arsenic Species in Cerebrospinal Fluid

5.1 Abstract

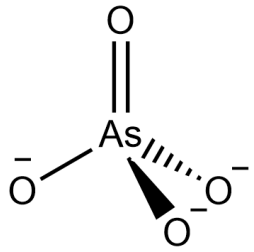
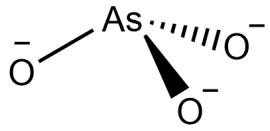
Arsenic is a heavy metal, found in the environment due to natural phenomena and human activity. It is considered one of the most toxic elements and is classified as a neurotoxin, capable of causing severe health effects upon exposure, including cancer and death. This study aimed to investigate the detection and analysis of arsenic species in artificial cerebral spinal fluid (aCSF) using surface-enhanced Raman spectroscopy (SERS). SERS spectra were collected from aCSF containing physiologically relevant concentrations of two inorganic arsenic species: arsenite (As(III)) and arsenate (As(V)). Arsenite is considered to be the more toxic of the two species. The results demonstrate that SERS is successful in distinguishing between these toxic inorganic species. The application of principal component analysis (PCA) further supported the detection of arsenic in aCSF, particularly in determining the presence of As(III) in the sample.

5.2 Introduction

Arsenic, known as one of the most toxic elements, exists in the environment due to natural phenomena and human activity.[1] It is classified as a neurotoxin, capable of causing severe health effects upon exposure, including cancer and death.[2,3] Groundwater contaminated with arsenic poses a significant health threat to over 140 million people worldwide.[3] The two most toxicology important forms of naturally occurring arsenic exist in inorganic states: trivalent arsenite (As(III)) and pentavalent arsenate (As(V)), with arsenite being the more toxic of the two (Table 5.1).[3]

Typically, arsenic enters the body, usually as As(V), after contaminated drinking water is consumed, passing through the small intestine, and accumulates in organs such as the liver, kidneys, lungs, spleen, and brain. Once inside the body, most arsenate is metabolized into the more toxic As(III).[2] During this process, As(V) is reduced through a redox reaction, leading to the formation of reactive oxygen species (ROS)[3]

Table 5.1: Structures of inorganic arsenic species and reported lethal amounts values.

	As III	**As V**
Structure		
BBB Permeability[4]	High	Low
CSF Abundance	High	Low
Water Abundance	Low	High
LD50[5]	50 µM	180 µM
LC50[6]	19.8 µM	1006 µM

The central nervous system benefits from some protection against arsenic due to the blood-brain barrier (BBB), which is effective in restricting the passage of arsenic in low concentrations. Unfortunately, the underdeveloped BBB in young children is more susceptible to arsenic crossing the BBB, thereby rendering children vulnerable to the neurological effects of arsenic.[1] The mechanism of arsenic entry into cerebrospinal fluid remains poorly understood, hindering our understanding of arsenic neurotoxilogical effects. This is critical, given the prevelance of arsenic poisoning, particularly as arsenic has medicinal uses. Although several studies have reported the detection of arsenic species in water using silver nanoparticles and silver-based substrates,[7–9] direct detection of arsenic in biofluids has not yet been explored. While arsenic is known to pass the BBB, the mechanism is still unknown. To address this knowledge gap, we propose a novel approach: utilizing surface-enhanced Raman spectroscopy (SERS) with gold nanoparticles (AuNPs) in artificial cerebral spinal fluid (aCSF).

This study aims to improve the detection of arsenic in a biological context and expand our understanding of its neurotoxic effects, given the limited understanding of its impact on the brain and associated health risks. By employing SERS with AuNPs, we investigate the presence and behavior of arsenic species in aCSF, providing insights into its implications for neurological health.

5.3 Background

5.3.1 Water Contamination: Bangladesh

The impact of arsenic contamination is particularly concerning for regions such as Bangladesh, where millions of people suffered the largest poisoning event in history due to the high levels of arsenic detected in public drinking water.[10] Most of the estimated 35 to 77 milllion Bangladeshi people have been exposed to arsenic through well water, where concentrations are often twenty times higher then the World Health Organization recommend limit.[10,11]

Table 5.2: A summary of neuropathy in APL patients treated with ATP in literature. Severe neurpathy is defined as severity classification grades of 3 or 4 (affecting daily living or incapacitating) , with reported side effects being dizziness, moodiness, musculoskeletal pain, and seizures. Adapted from Kühn et al. (2016).

Publication	No. of Cases	Cases with Neuropathy (# / %)	Cases with Severe Neuropathy (# / %)	Clinical Course of Neuropathy
Soignet et al.[13]	12	3 (25)	0	NA
Niu et al.[14]	58	6 (10.3)	0	NA
Soignet et al.[15]	40	17 (42.5)	1 (2.5)	Full Recovery
Lazo et al.[16]	12	2	1	Improved
Raffoux et al.[17]	20	2	0	NA
Shen et al.[18]	41	0	NA	NA
Shigeno et al.[19]	34	10 (29)	0	NA
Mathews et al.[20]	72	14 (19.4)	2 (2.7)	Improved
Estey et al.[21]	45	1 (2.2)	Not Reported	Not Reported
Ghavamzadeh et al.[22]	197	0	NA	NA
Iland et al.[23]	124	9 (7.2)	9 (7.2)	Not Reported

Between 1960 and 2000, Bangladesh's population grew from 48.01 million to 127.66 million.[24] This population boom was accompanied by a rise in the number of waterborne diseases. By the 1970s, a large outbreak of cholera swept through the region, garnering attention from agencies around the globe.[25] The use of water for drinking and cooking from surface sources, where bacteria thrive, was quickly determined to be the cause of the massive outbreak. As water penetrates the ground, it drains into aquifers. Soil, rocks, and other materials act as filters, effectively filtering out bacteria from the water.[26] World agencies, including UNICEF, The World Bank, and the British Geological Survey encouraged the use of tube-wells to provide the Bangladeshi people a clean water source from underground, spurring rapid installation of 10 million wells across the country.[25] However, it was not long after tube wells were installed that there was an increase in patients presenting with skin lesions at the School of Tropical Medicine. The lesions were soon diagnosed as arsenical keratosis, an obvious sign of arsenic posioning.[27] Early water tests soon revealed community wells contained high levels of arsenic, leading to an extensive public health study published in 1984 by Professor Kshitish Chandra Saha.[27] More than a decade later* the government of Bangladesh recognized the problem.[28] Nearly 57 million people were exposed to contaminated water since well installation began 40 years earlier.[10]

There are multiple factors that are believed to cause the levels of arsenic to be higher in the region in and around Bangladesh. Arsenic is a mobile element, traveling in solution with water or attached to other sediments. Arsenic has been transported to the Bengal basin during the natural weathering process of rocks in the Himalayan region, which washes minerals into the Bengal basin system.[11] While surface dwelling microbes are generally unable to survive in underground aquifers, there are multitudes of species that thrive in these dark environments with low oxygen. Without access to oxygen, organisms substitute other elements to use as an electron acceptor during cellular respiration or as an oxidizing agent during the metabolism of organic carbon matter. Iron oxides are suitable

*1996

substitutes and are commonly available in subsurface layers. Unfortunately, arsenic and iron tend to coexist in mineral deposits, and the reduction of iron by organisms facilitates the release of arsenic into groundwater.[11,29] While the arsenic poisoning in Bangladesh has been widely reported in recent years, the phenomenon is not uncommon, with naturally high levels of arsenic found around the world, including Taiwan, Chile, and the United States.[25]

5.3.2 Therapeutic Use and Detection in Biofluids

In contrast to the devastating effects of arsenic poisioning, arsenic can also serve as a life saving measure for people with acute promyelocytic leukemia (APL), a subtype of acute myeloid leukemia, usually the result of chromosomal aberration, t(15, 17).[30] Arsenic trioxide (ATO) was approved as a cancer treatment for APL by the FDA in 2003, after its effectiveness in inducing remission was confirmed across multiple countries.[31] However, concerns about its toxicity arise, including severe toxicity and neuropathy (Table 5.3.1).[12] Furthermore, due to the BBB, it is difficult to transport therapeutic levels of ATO into CSF. Currently, there is no direct clinical test to determine arsenic levels in cerebrospinal fluid (CSF). Further research in this area is needed due to its importance in understanding APL treatment and neurotoxic effects.[4,12,32]

5.4 Experimental Section

5.4.1 Materials

Trisodium citrate dihydrate, hydrochloric acid (HCl), nitric acid (HNO_3), sodium hydroxide (NaOH), sodium hydrogen arsenate heptahydrate (98.0 -102.0%), and sodium arsenite (98.0 - 101.0%) were purchased from Fisher Scientific. Tetrachloroauric(III) acid trihydrate was purchased from ACROS Organics. Artificial cerebrospinal fluid (aCSF) was purchased from EcoCyte Biosciences. All glassware was cleaned with aqua regia and rinsed with Millipore water (18 MΩ·cm) prior to use.

5.4.2 Synthesis of Nanoparticles

Gold nanoparticles (AuNPs) were synthesized using a modified version of the Frens method[33] by reducing $HAuCl_4$ with citrate. A 0.01% solution of tetrachlorauric(III) acid trihydrate ($HAuCl_4 \cdot 3H_2O$) was brought to a brief boil and a 1% trisodium citrate dihydrate ($C_6H_5Na_3O_7 \cdot 2H_2O$) solution was added. The solution was boiled for 5 minutes. Large volumes ($>$ 200 mL) of AuNPs were synthesized to maintain consistency in nanoparticles.

5.4.3 Sample Preparation

SERS spectra of both arsenic species dissolved in aCSF were collected. AuNPs in 4 mL aliquots were centrifuged and the supernatant was removed. The AuNPs were resuspended in 400µL of the various arsenic solutions. The concentration of arsenic in these solutions ranged from 1 µM to 100 mM. AuNPs were incubated for 15 minutes with each solution. After incubation, the pHs of the solutions were taken using pH paper strips (Fisher Scientific). The pH of each solution was adjusted to approximately 7 by adding 1-3µL of 3M HCl and NaOH dropwise. Spectra were obtained immediately.

5.4.4 Micro-Raman Spectroscopy

SERS spectra were collected using the home-built inverted confocal micro-Raman instrument, as previously described in Section 2.4.6. A 785 nm diode laser was selected as the excitation laser source (Innovative Photonic Solutions) and focused on the sample through a 40X objective (Nikon), with a power of 10 mW. For each sample, 5 replicate spectra were collected at a time, with each spectral acquisition time of 120 seconds. Each spectral set was collected from a sample 5 times.

5.4.5 Raman Data Pre-processing and Multivariate Analysis

The raw Raman spectra were pre-processed using a home-written, Python-based program. The program applies the same parameters and algorithms to each spectrum, then exports the result. This increases pre-processing speed, as well as removes human error.

Multivariate analysis techniques are often applied to Raman spectra, especially when spectral features and differences are difficult to detect with the human eye. These steps are discussed further in Chapter 3.

5.5 Results and Discussion

5.5.1 Arsenic Detection with SERS

In the literature, aqueous As(V) has been described as having two characteristic SERS peaks: a large, broad peak between 780-812 cm^{-1} and a smaller peak at 420 cm^{-1}.[7,8] The former is associated with a ν_1(A1) symmetric As-O stretch mode, while the latter results from a superposition of the ν_2(A1) and ν_5(E) stretching modes.[7,8] Here, we present the detection of arsenic species using SERS with AuNPs in aCSF. SERS spectra of eight samples containing As(III), As(V), and mixtures in aCSF were collected, along with a sample of the aCSF only. Twenty-five spectra were collected from each of the nine samples, resulting in a dataset of 225 spectra. Figure 5.1 shows these spectra averaged and plotted together for comparison.

Similar to previously described aqueous As(III), we observed a distinctive peak at 850 cm^{-1} for As(III) in aCSF.[7,8] There is not a peak at 420 cm^{-1}, however, a small peak at 413 cm^{-1} was present, which also appeared as a shoulder in the broad ~370 cm^{-1} peak of As(V).Through visual inspection, the SERS Raman spectra of As(III) differ from those of As(V) (Table 5.3). It is clear that As(V) is the less Raman active molecule when compared to As(III). This can be attributed to the structural differences between the

molecules, as the lone-pair electrons in As(III) increase its polarizability, thus increasing the capacity for Raman scattering. Due to the weaker Raman signal from As(V), the peaks from the blank are still discernible in the pure As(V) spectra. However, the molecular interactions between As(V) and the aCSF solution result in distinct changes in the shape and intensity of several peaks in the pure As(V) solution, particularly in the 950-1050 cm^{-1}, 1180-1400 cm^{-1}, and 1500-1600 cm^{-1} regions. The small, but distinct peak at 367 cm^{-1} is a notable characteristic of As(V). Furthermore, the 1200-1400 cm^{-1} region shows activity in As(V) solutions, with a peak observed at 1374 cm^{-1}. Additionally, a doublet is seen in the 1500-1600 cm^{-1} for As(V), consisting of a larger peak at 1537 cm^{-1} merging into a smaller peak at 1585 cm^{-1}. Upon closer inspection, a similar doublet appears in the As(III) samples, but with slightly shifted peaks at 1548 and 1580 cm^{-1}. As expected, the spectra from pure As(III) solutions shows limited contribution from the blank due to the strong SERS signal of As(III). A peak at 850 cm^{-1} is a clear indicator that toxic As(III) is present in the aCSF.

Comparison of these arsenic peaks with the blank solution, consisting of AuNPs and aCSF, reveals noticeable differences. Peaks in As(V) at 406 cm^{-1}, 563 cm^{-1}, $\sim$1000 cm^{-1}, $\sim$1200 - 1400 cm^{-1}, and $\sim$1600 cm^{-1} appear to arise in the blank. Citrate is used during the synthesis of AuNPs, citrate anions adhere to the surface of the nanoparticles and contribute to peaks around 1021 and 1219 cm^{-1}. However, the change in structure of these peaks indicates interaction between As(V) and the nanoparticles. In the blank solution, a peak is present at 367 cm^{-1}, but it is different in shape and size from the 367 cm^{-1} peak in As(V). Additionally, a doublet is observed in the 1500 cm^{-1} region, but the peaks have shifted slightly to 1541 and 1589 cm^{-1}, respectively, with 1589 cm^{-1} being the largest.

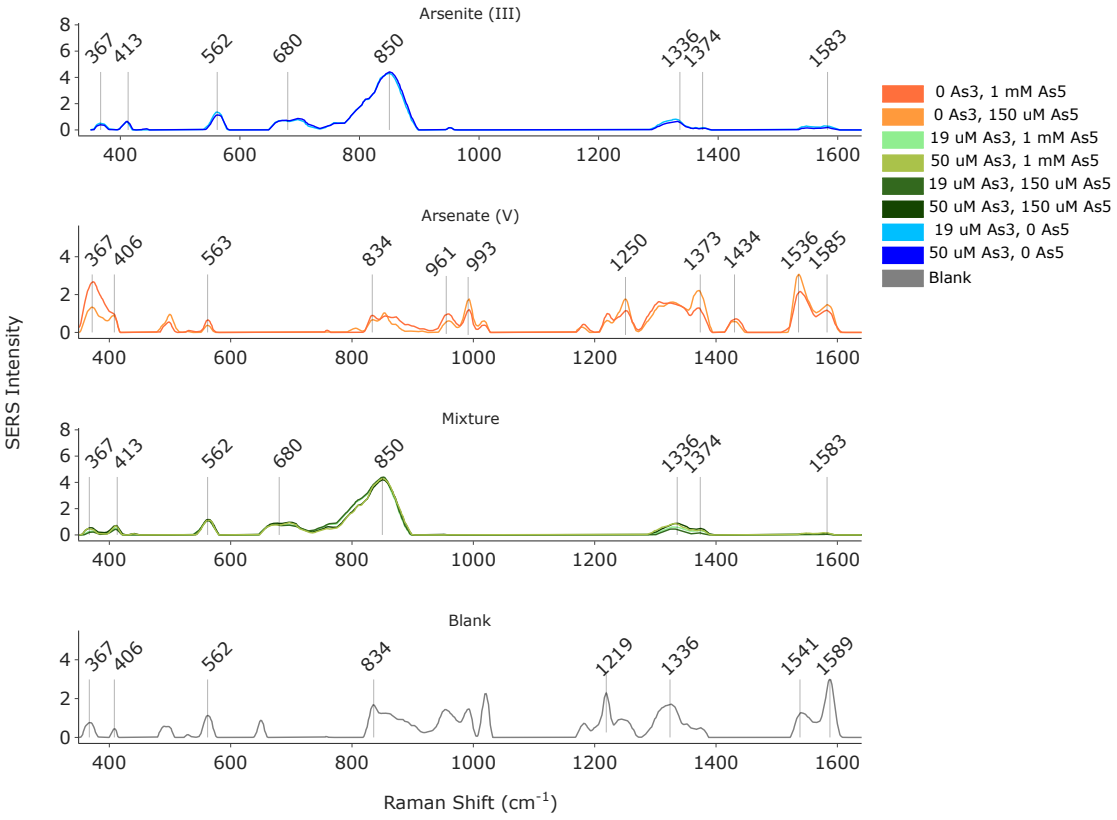


Figure 5.1: SERS spectra of arsenic samples in aCSF

Table 5.3: SERS peak assignments (cm^{-1}) for As(III), As(III)/As(V) mixtures, As(V) and blanks (AuNPs in aCSF)

As(III)	**Mixture**	**As(V)**	**Blank**
		367	367
413	413	406	
		498	
562	563	562	
663-709	663-709		
850	850	834	834
		961	
		993	995
			1021
			1219
		1250	
	1336		
	1374	1373	
		1434	
		1473	
		1537	1541
		1585	1589

5.5.2 Multivariate Analysis

PCA was applied to the arsenic data set and yielded 200 principal components. The first four PCs accounted for greater than 95% of the variance and were thus selected for further analysis. Cross-validation was performed using the Leave-One-Out method, yielding a mean squared error (MSE) of 0.0124 and a root mean squared error (RMSE) of 0.1115, indicating that the PCA model effectively captured the underlying patterns and variations in the data. The MSE score is a measurement between the predicted and actual PC values, indicating the magnitude of the error in the model where 0 equals no error. A lower MSE score indicates a better fit of the model to the data, and the value increases along with the error. Similarly to the MSE, the RMSE score is used to assess the ability of the model to predict values across the dataset, however, it also takes the scale of the dependent variable into account. MSE and RMSE scores are crucial in determining the performance of the model and determining its predictive power.

The contributions of the PCs were investigated by comparing the values of the different sample types, which were plotted in two-dimensional (2D) figures. Both PC 1 and PC 2 indicate the presence of As(III) in the sample (Figure 5.2). The pure As(III) samples and the As(III)/As(V) mixtures have a PC 1 value of roughly 1, while the blank and pure As(V) samples have PC 1 values between 0.08-0.4. Pure As(III) and As(III)/As(V) mixtures cluster tightly together along PC 2 as well, with values ranging between -0.2 and 0.1, however the blank is separated from the pure As(V) solutions. The blanks, with PC 2 values between 0.5 and 0.8, clustered between As(III) solutions and the pure As(V) samples. The pure As(V) samples score just above the blanks, with values between 0.8 and 0.9. The substantial overlap between the pure As(III) and As(III)/As(V) mixtures is likely due to the overpowering intensity of the As(III) Raman signal over As(V). When looking at the scatter plot of PC 1 vs PC 2, the separation between solutions containing As(III) from the blanks and pure As(V) samples is striking.

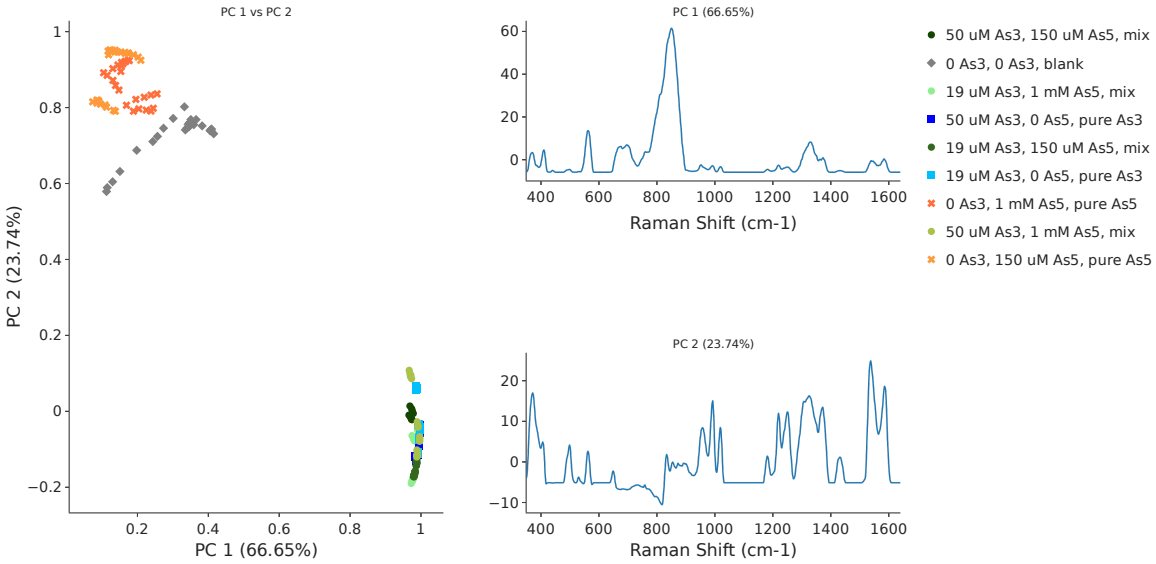


Figure 5.2: 2D PCA plot of PC1 vs. PC 2 for SERS spectra of arsenic species in aCSF. The loadings for PC 1 and PC 2 are included. λ_{ex} = 785 nm, P = 10 mW, t = 120s.

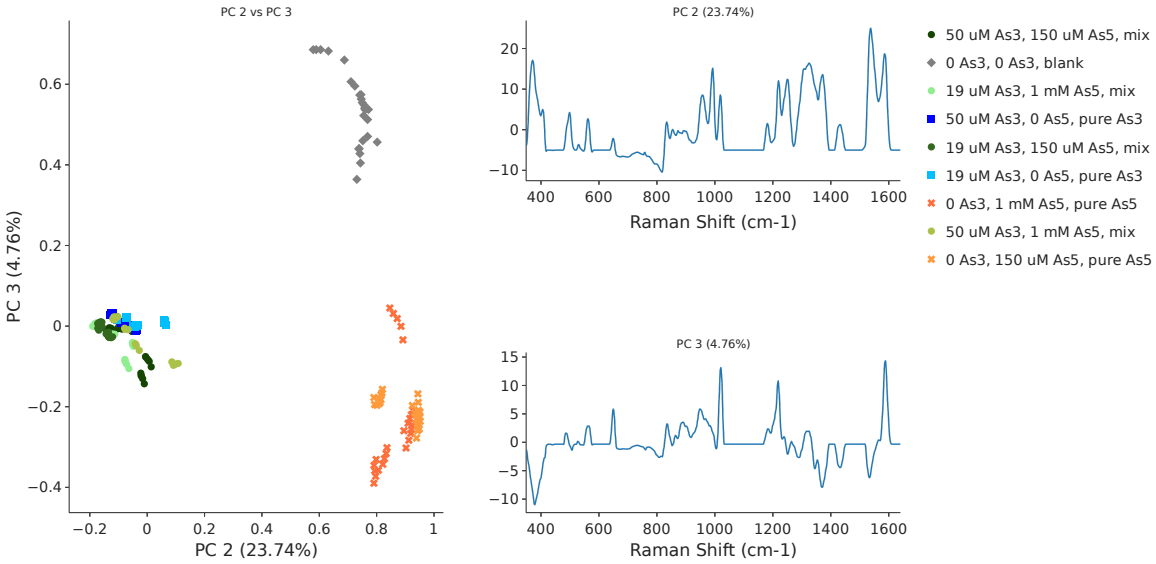


Figure 5.3: PCA plot of PC 2 vs PC 3, with loadings included. There is limited spread of the pure As(III) samples along the PC 3, especially compared to the spread seen in the mixtures. λ_{ex} = 785 nm, P = 10 mW, t = 120s

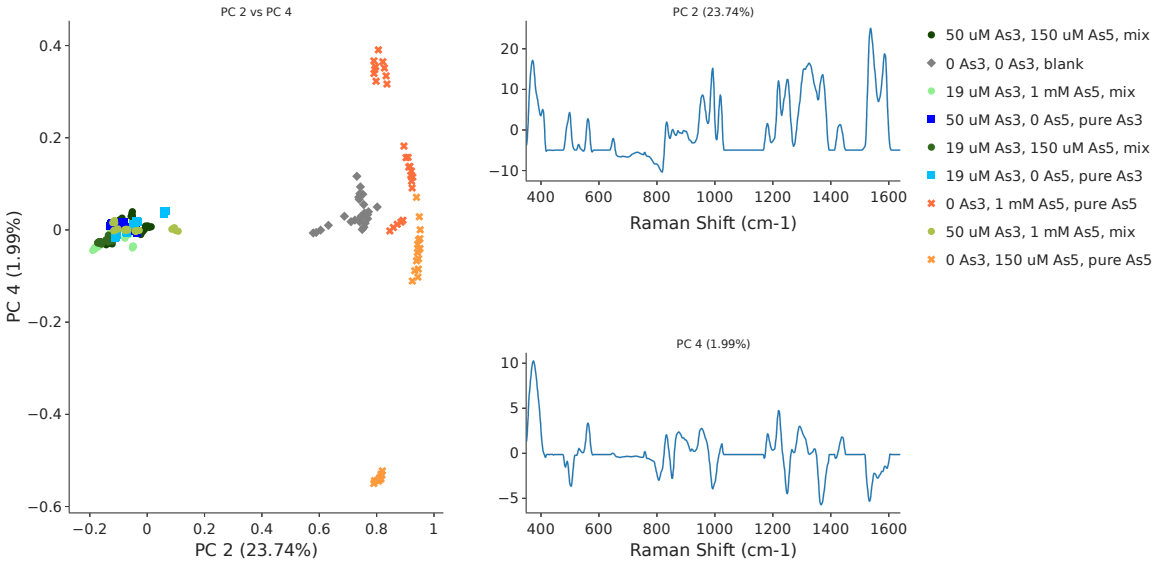


Figure 5.4: PCA plot of PC 2 vs PC 4, with loadings included. λ_{ex} = 785 nm, P = 10 mW, t = 120s.

Table 5.4: SERS peak positions from the loadings of PCs 1-4 with symbols indicating the SERS spectra from which they arise.

SERS peak (cm^{-1})	PC 1	PC 2	PC 3	PC 4
367	■●	■×		■×
498		■		
562	▲■●	▲■●	▲■●	▲■●
834		■×	■×	■×
850	▲●			
961		■		■
993		■×		■×
1021		×	×	
1219		×	×	×
1250		■		
∼1336	●			
∼1373	■●	■●		
1434		■		
∼1539	■	■		
1585	■	■		
1589			×	

As(III) ▲	As(V) ■	Mixture ●	Blank ×

Upon examination of PC 1 loadings, it is evident that it reflects the presence of the distinctive As(III) peak at 850 cm^{-1} in the aCSF solution. There are also a few small peaks in the 671-702 cm^{-1} region which reflect the presence of As(III). Unlike PC 1, the loadings of PC 2 are not dominated by a single large peak. The peak at 367 cm^{-1} is one of the higher valued in PC 2, reflecting its strong contribution and correlating with the presence of As(V). Also of interest is the doublet at 1538 and 1586 cm^{-1}, which reflects the contribution of As(V). A summary of the SERS peak contributions to each loading can be found in Table 5.4

PC 3 mainly indicates the absence of an arsenic species in the solution, effectively distinguishing the blanks from the other samples. In examining the PC 3 loadings, it is found that most of the peaks seen in PC 3 correlate with the blanks, which have values between 0.3 and 0.7. The pure As(III) samples have values between -0.02 and 0.04, while the mixtures lie between -0.1 and 0.03. The As(III) samples mainly have values between -0.2 and 0.03, with a few data points near zero. Negative peaks are also of note in PC 3, located at 380, 1371, and 1536 cm^{-1} which reflect the absence of As(V). Comparing PC 3 against PC 2 (Figure 5.3) allows for better distinction between the blanks and pure As(V) samples.

PC 4, accounting for only 1.7% of variance within the samples, shows some separation of the As(V) samples based on their concentration (Figure 5.4). The concentration dependent separation of the As(V) samples are seen in the PC 4 scores: 1 mM As(V) solutions are valued between 0.4 and 0.04, while the 150µM samples are between -0.5 and 0.06. The PC 4 values for blanks, mixtures, and pure As(III) solutions lay between -0.09 and 0.1.

A peak at 562 cm^{-1}, an indicator of the presence of an arsenic species in aCSF solution, was found in the four PCs. Despite the weaker Raman signal of As(V), PC loadings exhibit numerous As(V) peaks, allowing As(V) differentiation from the other samples. In solution, the Raman signal of As(III) is stronger than that of As(V), effectively obscuring

the As(V) Raman signature and making it challenging for As(III)/As(V) mixtures to be distinguished from samples containing As(III) only through PCA.

5.6 Conclusions

Our study successfully utilized surface-enhanced Raman spectroscopy (SERS) with gold nanoparticles (AuNPs) to detect and analyze arsenic species in artificial cerebrospinal fluid (aCSF). Understanding the cellular interactions between arsenic and the central nervous system is crucial, as it directly impacts the treatment and accurate diagnosis of arsenic related individuals. Gaining comprehensive insight into the short- and long-term health effects of arsenic exposure in the brain will not only aid in developing treatment strategies for the therapeutic use of arsenic, but also enable early detection and intervention to protect the millions of people at risk drinking arsenic laden water.

We found that despite weaker Raman scattering, As(V) showed more discernible peaks compared to As(III) in aCSF samples, allowing effective differentiation through PCA analysis (Figure 5.5). However, challenges arose in distinguishing samples containing both arsenic species from those with As(III) only, as its stronger SERS signal dominated the spectra. Further analysis of PCA loadings and SERS spectra confirmed that mixed solutions exhibited peaks from both As(III) and As(V). This is particularly relevant for understanding the mechanisms of arsenic entry and behavior in the brain, as the presence of arsenic can quickly be detected, along with the determination of whether or the more toxic species is present. Our research also suggests that additional research applying SERS to detect and quantify arsenic in CSF shows great promise.

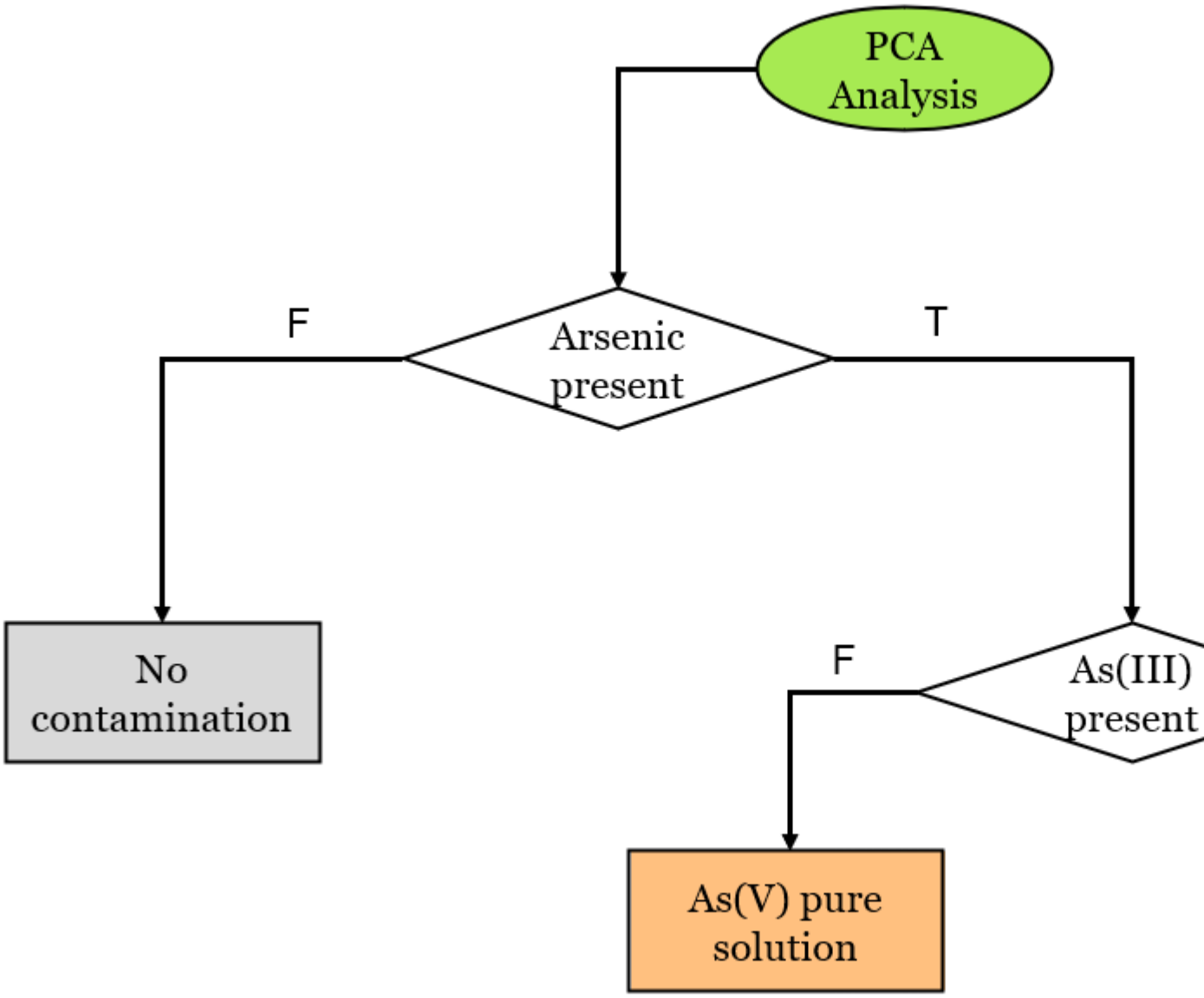


Figure 5.5: A flow chart summarizing the conclusions which can be drawn from the PCA analysis of Arsenic species in aCSF.

References

[1] Hitoshi Mochizuki. Arsenic neurotoxicity in humans. *International Journal of Molecular Sciences*, 20(14), 2019. ISSN 1422-0067. doi: 10.3390/ijms20143418. URL https://www.mdpi.com/1422-0067/20/14/3418. 85, 87

[2] Chandra Prakash, Manisha Soni, and Vijay Kumar. Mitochondrial oxidative stress and dysfunction in arsenic neurotoxicity: A review. *Journal of Applied Toxicology*, 36(2):179–188, Oct 2015. doi: 10.1002/jat.3256. URL https://doi.org/10.1002/jat.3256. 85

[3] Lou B Wu R Wang G Lu C Wang H Pi J Xu Y Hu Y, Li J. The role of reactive oxygen species in arsenic toxicity. Feb 2020. doi: 10.3390/biom10020240. URL https://pubmed.ncbi.nlm.nih.gov/32033297/. 85

[4] Shengjin Fan Zhiqiang Wu Liwang Lin Hongzhu Chen Yanhui Gao Xin Hai Meihua Guo, Qilei Zhao. Characteristics of arsenic species in cerebrospinal fluid (csf) of acute promyelocytic leukaemia (apl) patients treated with arsenic trioxide plus mannitol. *British Pharmacological Society*, 87(10):4020–4026, Feburary 2021. doi: https://doi.org/10.1111/bcp.14804. 86, 90

[5] Hitoshi Mochizuki. Arsenic neurotoxicity in humans. *International Journal of Molecular Sciences*, 20(14):3418, 2019. doi: 10.3390/ijms20143418. URL https://www.mdpi.com/1422-0067/20/14/3418. 86

[6] Jay S. Petrick, Felix Ayala-Fierro, William R. Cullen, Dean E. Carter, and H. Vasken Aposhian. Monomethylarsonous acid (mmaiii) is more toxic than arsenite in chang human hepatocytes. *Toxicology and Applied Pharmacology*, 163(2):203–207, 2000. ISSN 0041-008X. doi: https://doi.org/10.1006/taap.1999.8872. URL https://www.sciencedirect.com/science/article/pii/S0041008X99988725. 86

[7] Xiaochen Lv, Shu Li, Qing Yang, Shaoying Zhang, Jie Su, Shi-Bo Cheng, Yongchao Lai, Jing Chen, and Jinhua Zhan. Robust, reliable and quantitative sensing of aqueous arsenic species by surface-enhanced raman spectroscopy: The crucial role of surface silver ions for good analytical practice. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 281:121600, Nov 2022. doi: 10.1016/j.saa.2022.121600. URL https://www.sciencedirect.com/science/article/pii/S1386142522007491. 87, 92

[8] Mei-Juan Han, Jumin Hao, Zhonghou Xu, and Xiaoguang Meng. Surface-enhanced raman scattering for arsenate detection on multilayer silver nanofilms. *Analytica Chimica Acta*, 692(1):96–102, 2011. ISSN 0003-2670. doi: https://doi.org/10.1016/j.aca.2011.02.054. URL https://www.sciencedirect.com/science/article/pii/S0003267011002790. 92

[9] Jumin Hao, Mei-Juan Han, Jinwei Li, and Xiaoguang Meng. Surface modification of silver nanofilms for improved perchlorate detection by surface-enhanced raman scattering. *Journal of Co lloid and Interface Science*, 377(1):51–57, 2012. ISSN 0021-9797. doi: https://doi.org/10.1016/j.jcis.2012.03.036. URL https://www.sciencedirect.com/science/article/pii/S0021979712003050. 87

[10] A. H. Smith, E. O. Lingas, and M. Rahman. Contamination of drinking-water by arsenic in bangladesh: a public health emergency., 1999. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2560840/. 87, 89

[11] Stephane Guillot and Laurent Charlet. Bengal arsenic, an archive of himalaya orogeny and paleohydrology. *Journal of Environmental Science and Health*, 42(12):1785–1794, October 2007. doi: https://doi.org/10.1080/10934520701566702. 87, 89, 90

[12] M. Kühn, K. Sammartin, M. Nabergoj, and F. Vianello. Severe acute axonal neuropathy following treatment with arsenic trioxide for acute promyelocytic leukemia: a case report. *Mediterranean journal of hematology and infectious*

diseases, 8(1), 2016. doi: https://doi.org/10.4084/MJHID.2016.023. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4848016/. xi, 88, 90

[13] SL Soignet, SR Frankel, Douer D, and et al. United states multicenter study of arsenic trioxide in relapsed acute promyelocytic leukemia. *J Clin Oncol*, 19(18):3852–3860, 2001. URL https://pubmed.ncbi.nlm.nih.gov/11559723/. 88

[14] C. Niu, H. Yan, T. Yu, and et al. Studies on treatment of acute promyelocytic leukemia with arsenic trioxide: remission induction, follow-up, and molecular monitoring in 11 newly diagnosed and 47 relapsed acute promyelocytic leukemia patients. *Blood*, 94(10):3315–3324, 1999. URL https://pubmed.ncbi.nlm.nih.gov/10552940/. 88

[15] SL Soignet, P Maslak, ZG Wang, and et al. Complete remission after treatment of acute promyelocytic leukemia with arsenic trioxide. *The New England journal of medicine*, 339(19):1341–1348, 1998. URL https://pubmed.ncbi.nlm.nih.gov/9801394/. 88

[16] G Lazo, H Kantarjian, E Estey, D Thomas, S O'Brien, and J Cortes. Use of arsenic trioxide (as2o3) in the treatment of patients with acute promyelocytic leukemia: the m. d. anderson experience. *Cancer*, 97(9):2218–2224, 2003. URL https://pubmed.ncbi.nlm.nih.gov/12712474/. 88

[17] E Raffoux, P Rousselot, J Poupon, MT Daniel, and et al. Combined treatment with arsenic trioxide and all-trans-retinoic acid in patients with relapsed acute promyelocytic leukemia. *Journal of clinical oncology*, 21(12):2326–2334, 2003. URL https://doi.org/10.1200/JCO.2003.01.149. 88

[18] ZX Shen, ZZ Shi, J Fang, BW Gu, JM Li, and et al. All-trans retinoic acid/as2o3 combination yields a high quality remission and survival in newly diagnosed acute promyelocytic leukemia. *Proceedings of the National Academy of Sciences of the*

United States of America, 101(15):5328–5335, 2004. URL https://pubmed.ncbi.nlm.nih.gov/15044693/. 88

[19] K Shigeno, K Naito, N Sahara, M Kobayashi, and et al. Arsenic trioxide therapy in relapsed or refractory japanese patients with acute promyelocytic leukemia: updated outcomes of the phase ii study and postremission therapies. *International journal of hematology*, 82(3):224–229, 2005. URL https://pubmed.ncbi.nlm.nih.gov/16207595/. 88

[20] Vikram Mathews, Ezhilarasi Chendamarai, Biju George, Auro Viswabandya, and Alok Srivastava. Treatment of acute promyelocytic leukemia with single-agent arsenic trioxide. *Mediterranean journal of hematology and infectious diseases*, 3(1), 2011. URL https://doi.org/10.4084/MJHID.2011.056. 88

[21] E Estey, G Garcia-Manero, A Ferrajoli, S Faderl, and et al. Use of all-trans retinoic acid plus arsenic trioxide as an alternative to chemotherapy in untreated acute promyelocytic leukemia. *Blood*, page 3469–3473, 2006. 88

[22] A Ghavamzadeh, K Alimoghaddam, S Rostami, SH Ghaffari, M Jahani, M Iravani, SA Mousavi, B Bahar, and M Jalili. Phase ii study of single-agent arsenic trioxide for the front-line therapy of acute promyelocytic leukemia. *Journal of clinical oncology*, 29(20):2753–2757, 2011. URL https://pubmed.ncbi.nlm.nih.gov/21646615/. 88

[23] HJ Iland, K Bradstock, S G Supple, and et al. All-trans-retinoic acid, idarubicin, and iv arsenic trioxide as initial therapy in acute promyelocytic leukemia (apml4). *Blood*, 120(8):1570–1752, 2012. URL https://pubmed.ncbi.nlm.nih.gov/22715121/. 88

[24] World Bank. Population growth in bangladesh, 2023. URL https://www.worlddata.info/asia/bangladesh/index.php. 89

[25] S. Sambu and R. Wilson. Arsenic in food and water - a brief history. *Toxicology and industrial health*, 24(4):217–26, 05 2008. URL https://utk.

idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/arsenic-food-water-brief-history/docview/223271590/se-2. 89, 90

[26] United States Geological Survey (USGU). Quality of ground water. URL https://pubs.usgs.gov/gip/gw/quality.html. 89

[27] Arijit Coondoo. Professor kshitish chandra saha. *Indian J Dermatol*, 41(4), August 2016. doi: 10.4103/0019-5154.185741. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4966417/. 89

[28] Fakir M. Yunus, Safayet Khan, Chowdhury, Priyanka, Abul H. Milton, Sumaira Hussain, and Mahfuzar Rahman. A review of groundwater arsenic contamination in bangladesh: The millennium development goal era and beyond. *International Journal of Environmental Research and Public Health 13*, (2), 2016. doi: https://doi.org/10.3390/ijerph13020215. 89

[29] Zahid Hassan, Munawar Sultana, Hans V. Westerhoff, Sirajul I. Khan, and Wilfred F. M. Röling. Iron cycling potentials of arsenic contaminated groundwater in bangladesh as revealed by enrichment cultivation. *Geomicrobiology Journal*, 33 (9):779–792, 2016. doi: 10.1080/01490451.2015.1111471. URL https://www.tandfonline.com/doi/full/10.1080/01490451.2015.1111471. 90

[30] Kamran Alimoghaddam. A review of arsenic trioxide and acute promyelocytic leukemia. *International Journal of Hematology-Oncology and Stem Cell Research 8*, (3):44–54, 2014. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4305381/. 90

[31] Yoshinaga-Sakurai K Rosen BP Yoshinaga M. Paul NP, Galván AE. Arsenic in medicine: past, present and future. *Biometals*, 36(2):283–301, April 2023. doi: 10.1007/s10534-022-00371-y. URL https://pubmed.ncbi.nlm.nih.gov/35190937/. 90

[32] Ashraf V valappil and Abraham Mammen. Subacute arsenic neuropathy: Clinical and electrophysiological observations. *Journal of Neurosciences in Rural Practice 10*, (3):529–532, 2019. doi: 10.1055/s-0039-1695693. URL https://doi.org/10.1055/s-0039-1695693. 90

[33] G Frens. Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions. *Nature Physical Science*, 241(105):20–22, 1973. doi: 10.1038/physci241020a0. URL https://doi.org/10.1038/physci241020a0. 91

Chapter 6

Conclusions

Raman spectroscopy, particularly Surface-Enhanced Raman Spectroscopy (SERS), has shown great potential in the field of biology, offering rapid chemical information on biological samples, including living cells and tissues. Unlike traditional methods, Raman spectroscopy can monitor chemical interactions within living targets without the need for fluorescent probes, presenting potential for selective and rapid biomolecule detection.

The combination of multivariate analysis with Raman spectroscopy is particularly powerful, as demonstrated by the successful detection of arsenic species in artificial cerebral spinal fluid (aCSF) using SERS with gold nanoparticles. Inorganic arsenic, a highly toxic heavy metal with neurotoxic effects, shows characteristic SERS peaks in aCSF, enabling differentiation. Further information on the chemical behavior of arsenic species in aCSF was obtained using Principal Component Analysis (PCA) and further aids in distinguishing As(III) from As(V). SERS detection of DA and its metabolite, DOPAC, were able to be detected at physiologically relevant concentrations in aCSF, demonstrating its potential for detection and differentiation of neurochemicals in biological media.

Continued research in Raman spectroscopy holds promise for further advancements in selective and rapid biomolecule detection, particularly coupled with advances in Raman tags. Selective Raman tags the simultaneous monitoring of multiple biomolecules or proteins, expanding research possibilities and facilitating non-invasive real-time analysis

of cellular processes. Monitoring and understanding the chemical dynamics at this level is essential for early disease diagnosis and medical research.

Over the course of this research, a comprehensive Python package *RamanRoot*, was developed to efficiently manage Raman datasets, conduct in-depth analysis, and visualize results. The significance of effective data management for handling large Raman spectral datasets became increasingly evident during this research, especially when dealing with extensive, growing datasets that require robust statistical analysis and validation. *RamanRoot* offers tools specifically tailored for Raman spectroscopy, providing a dynamic system that integrates data and facilitates real-time understanding of patterns and trends. By adopting an organized, standardized approach, Raman datasets can be efficiently explored and interpreted.

As Raman spectroscopy continues to evolve, it is poised to play a pivotal role in the advancement of research in various disciplines, contributing to scientific breakthroughs and further advancing our understanding of complex biological systems. The knowledge gained from this work provides a foundation for future explorations in RS.

Appendix

Appendix A

Chapter 4 Supplemental Information

Table A.1: Healthy Brain Concentrations for DA, DOPAC, HVA, 3-OMD, and 3-MT

Neurotransmitter	**Concentration Level in a Healthy Brain (nM)**
DA	1,000 - 3,000[1]
DOPAC	7,000 - 18,000[2]
HVA	35-250[3]
3-OMD	14 - 20[4,5]
3-MT	10 - 100[6]

Table A.2: Summary of the samples prepared for DA/DOPAC in aCSF study. For each sample group, five samples were prepared at each concentration or mixture

[DA] (µM)	DOPAC (µM)	Sample Group	Mixture Ratio (DA:DOPAC)
1	0	DA	
3	0	DA	
7	0	DA	
10	0	DA	
0	3	DOPAC	
0	7	DOPAC	
0	10	DOPAC	
0	12	DOPAC	
0	18	DOPAC	
7	3	Mix	2.33
10	10	Mix	1
3	7	Mix	0.43
3	10	Mix	0.3
3	12	Mix	0.25
3	18	Mix	0.17
1	7	Mix	0.14
1	10	Mix	0.1
1	12	Mix	0.08
1	18	Mix	0.06
0	0	Blank	

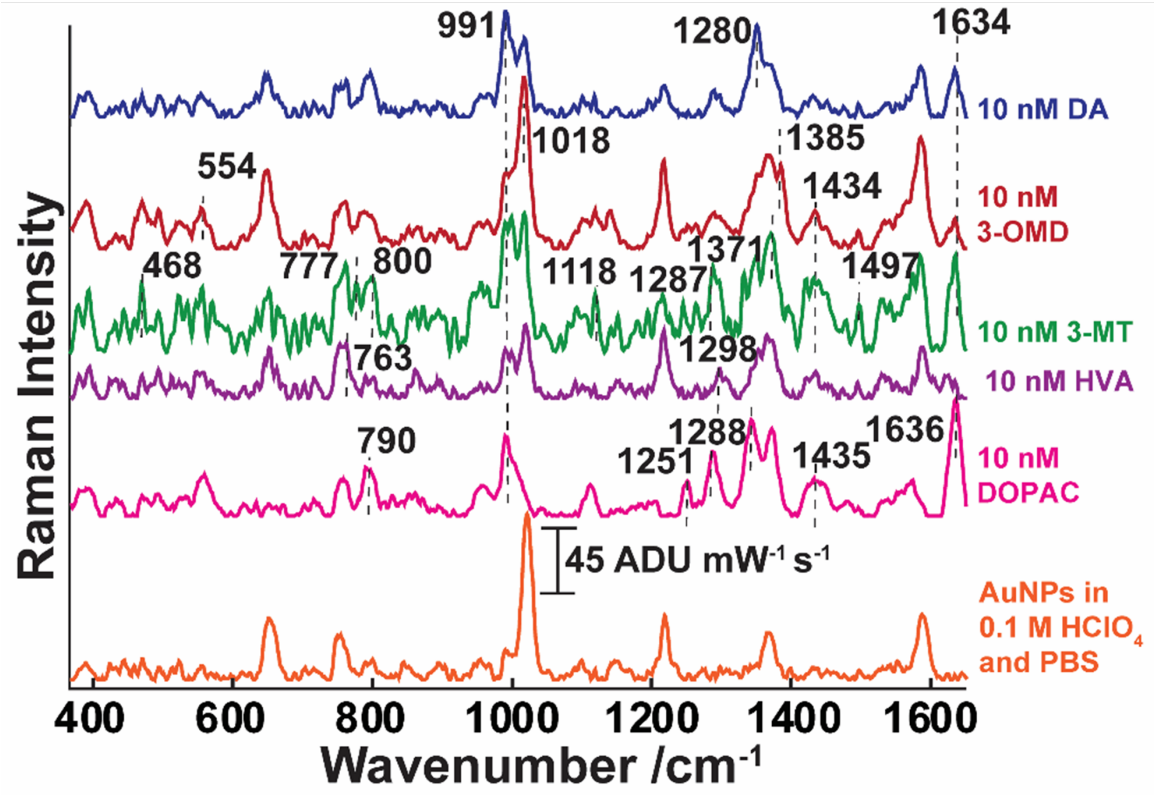


Figure A.1: SERS spectra of 10 nM DA (blue), 3-OMD (red), 3-MT (green), HVA (purple), DOPAC (pink) and a spectrum of the blank (AuNPs in 0.1 M HClO4 and PBS). λ_{ex} = 785 nm, P= 10 mW, t= 30 s. Figure produced by P. Pimento.

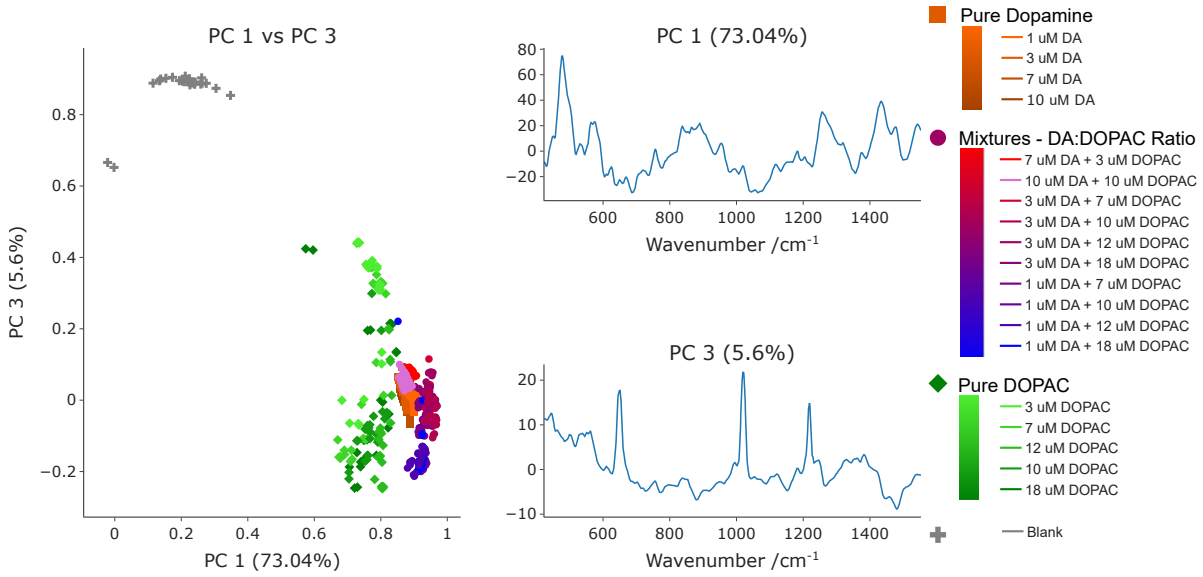


Figure A.2: PC 1 vs PC 3 of DA and DOPAC solutions. Both PC 1 and 3 strongly differentiate DA/DOPAC solutions from the blank. λ_{ex} = 785nm, P= 10 mW, t= 30s.

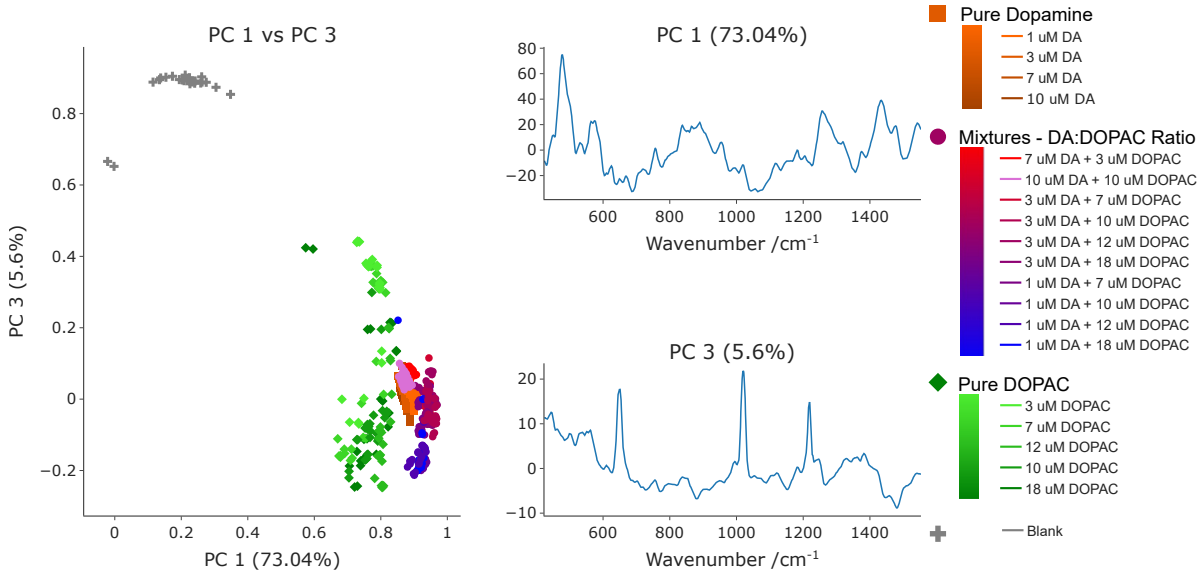


Figure A.3: PC 2 vs PC 3 of DA and DOPAC solutions. From this perspective, there is clear distinction between DOPAC and its P mixtures. Pure DA and high concentration DA mixtures cluster tightly compared to the pure DOPAC and low concentration DA mixtures. This difference in clustering is attributed to the smaller concentration range for the DA in mixtures (1-3 µM) vs the large concentration range (7-18 µM) for DOPAC. λ_{ex} = 785nm, P= 10 mW, t= 30s.

References

[1] M. K. Zachek, P. Takmakov, J. Park, R. M. Wightman, and G. S. McCarty. Simultaneous monitoring of dopamine concentration at spatially different brain locations in vivo. *Biosens Bioelectron*, January 2010. doi: 10.1016/j.bios.2009.10.008. 114

[2] R. C. Pierce and G. V. Rebec. Dopamine-, nmda- and sigma-receptor antagonists exert differential effects on basal and amphetamine-induced changes in neostriatal ascorbate and dopac in awake, behaving rats. *Brain Res.*, May 1992. doi: 10.1016/0006-8993(92)90741-q. 114

[3] Westenberg HG Kahn RS Scheepers FE, Gispen-de Wied CC. The effect of olanzapine treatment on monoamine metabolite concentrations in the cerebrospinal fluid of schizophrenic patients. *Neuropsychopharmacology*, 25(4):468–75, October 2001. doi: doi:10.1016/S0893-133X(01)00250-0. URL https://pubmed.ncbi.nlm.nih.gov/11557160/. 114

[4] Tuomainen P. Tuominen R. K. Männistö, P. T. Different in vivo properties of three new inhibitors of catechol o-methyltransferase in the rat. *British journal of pharmacology*, 105(3):569–574, 1992. URL https://doi.org/10.1111/j.1476-5381.1992.tb09020.x. 114

[5] Takahashi S Nozaki Y Kikuchi T Tohgi H, Abe T. Concentrations of tyrosine, l-dihydroxyphenylalanine, dopamine, and 3-o-methyldopa in the cerebrospinal fluid of parkinson's disease. *Neurosci Lett*, 127(2):212–4, June 1991. doi: 10.1016/0304-3940(91)90796-v. URL https://pubmed.ncbi.nlm.nih.gov/1908962/. 114

[6] P. T. Hyland, K.; Clayton. Aromatic l-amino acid decarboxylase deficiency: diagnostic methodology. *Clin Chem*, 38(12):2405–10, 1992. URL https://pubmed.ncbi.nlm.nih.gov/1281049/. 114

Vita

Natalie was born in Grand Rapids, MI, the only child to Bill Dunn and Ruth Athman. She grew up between the woods and lakes of Michigan and the farmlands of Central Minnesota, developing a strong love of nature and family (biological and found). She attended Northview Public Schools where she worked in technical theater and participated in the marching band.

After graduation, Natalie attended Aquinas College, a liberal arts college in downtown Grand Rapids, where she was inspired by the close-knit andsupportive environment in the science deparments. She worked towards a biology degree focusing on plant and environmental studies, while also obtaining her minor in chemistry. While initially having an aversion to chemistry due to negative high school experiences, she found general chemistry one of her more enjoyable classes. She quickly became a tutor for the chemistry department, resident of the chemistry student lounge / tutorial center, assistant to the chemistry stock room, and vice president of the chemistry society. She also became the vice president to Tri Beta biological society and worked in the biology stock room.

Natalie met Dr. Vogt when he was a guest speaker at Aquinas College in the 2014-2015 school year. Her friend and classmate invited her to the school sponsored lunch with Dr. Vogt after his lecture, and became passionate about his algal research projects, and spent a summer working as an undergraduate researcher in his lab.

After earning her B.S. in Biology in May of 2016, Natalie started as a chemistry graduate student at the University of Tennessee in August of 2016. She worked with Dr. Vogt until his passing in December of 2018. Since then she has worked in the

Raman spectroscopy lab of Dr. Sharma, developing methods of live cell and tissue data Raman data collection and analysis. During her time at UTK, she has been a member of the Association of Chemistry Graduate Students as the Philanthropy Chair. She has developed a passion for teaching and tutoring across all age groups and volunteering for science and environmental outreach projects. In 2019, she was awarded the Keenan Teaching Award, which is given to a graduate student who is regarded as an exceptional teacher.