8-2021

# Comparative Analysis of Genuine Acetaminophen by Raman Spectroscopy and Chemometrics

Elisabeth M. Good
*State University of New York College at Buffalo - Buffalo State College*, goodem01@mail.buffalostate.edu

**Advisor**
Jinseok Heo, PhD.
**First Reader**
Jinseok Heo, PhD.
**Second Reader**
Kimberly Bagley, PhD.
**Third Reader**
Joonyeong Kim, PhD.

**Comparative Analysis of Genuine Acetaminophen by Raman Spectroscopy and Chemometrics**

by

Elisabeth M. Good

An Abstract of a Project
in
Forensic Science

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2021

Department of Chemistry
Buffalo State College
State University of New York

ABSTRACT

**Comparative Analysis of Genuine Acetaminophen by Raman Spectroscopy and Chemometrics**

The use of Raman spectroscopy with Chemometric analysis has been well documented as an effective method for the identification of counterfeit pharmaceutical products. Due to an ongoing concern about the lack of transparency surrounding the manufacturing of generic pharmaceuticals in the US, this project aimed to examine the utility of the method for the differentiation of generic pharmaceuticals. A testing pool of six generic acetaminophen brands along with four unique samples of name brand Tylenol acetaminophen was obtained and representative tablets were characterized by Raman spectroscopy.

The results showed certain observable variance between brands and manufacturing lots. Two brands were determined to contain a titanium dioxide ($TiO_2$) coating, allowing for simple differentiation based on examination of the Raman spectra. The Raman spectra of brands in which $TiO_2$ was not observed were analyzed using Principal Component Analysis (PCA) to examine the unique variance present in each brand. This revealed discrete groupings of lots and brands due to variance observable within the PCA projections. The project provides proof of concept for the differentiation of genuine articles by Raman spectroscopy and chemometrics and potential for the method to be used for classification and prediction of products.

Department of Chemistry
Buffalo State College
State University of New York

**Comparative Analysis of Genuine Acetaminophen by Raman Spectroscopy and Chemometrics**

A Project in
Forensic Science

by

Elisabeth M. Good

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

Approved by

Jinseok Heo, Ph.D.
Associate Professor of Chemistry
Project Advisor, First Reader

Kimberly Bagley, Ph.D.
Chair and Professor of the Chemistry Department
Second Reader

Joonyeong Kim, Ph.D.
Associate Professor of Chemistry
Third Reader

## Acknowledgments

I would like to give my utmost thanks to Dr. Jinseok Heo for acting as my advisor on this project. His guidance, patience, and support has been of so much value throughout my work and I am extremely grateful. Also a large thank you to Dr. Jamie Kim and Dr. Kimberly Bagley for acting on my Project Committee and all the opportunities and guidance they offered throughout my time at Buffalo State. I am also grateful to my fellow graduate students, family, and friends who acted as sounding boards, proofreaders, and endless supporters throughout all my endeavors.

***Table of Contents***

**List of Abbreviations**

FDA = Food and Drug Administration

NDC = National Drug Code

TLC = Thin Layer Chromatography

NMR = Nuclear Magnetic Resonance

API = Active Pharmaceutical Ingredient

HCA = Hierarchical Cluster Analysis

PCA = Principal Component Analysis

CCD = Charge Coupled Device

Ft-IR = Fourier-transformed Infrared Spectroscopy

DoE = Design of Experiments

TOS = Theory of Sampling

MLR = Multiple Linear Regression

PC = Principal Component

S/N = Signal to Noise Ratio

$cm^{-1}$ = Wavenumber

DoE = Design of Experiments

TOS = Theory of Sampling

*List of Figures*

***List of Tables***

## I. Introduction

### 1.1 Research Motivations and Goals

Within the United States of America, the US Food and Drug Administration (US FDA) is responsible for the monitoring and oversight of all pharmaceutical research, development, and production. Every drug product in the US is registered with a National Drug Code (NDC) and the FDA regulates, approves, and documents all pharmaceuticals on the market. This is done in the hopes of ensuring all drug products have undergone significant testing to confirm their safety and efficacy. Additionally, by enforcing stringent manufacturing regulations for pharmaceuticals consistency and safety between manufacturing locations and times is maintained. However, despite the extensive reach of the FDA, certain unregulated products persist in reaching the market. One of the most prevalent sources of unregulated pharmaceuticals is the counterfeiting of genuine products.[1]

Counterfeiting of pharmaceutical drug products is a common offense within the US and globally. The presence of counterfeits within the market is primarily a safety concern, as any counterfeit product may contain illicit and unsafe compounds and result in injury or death for the individual who takes it. This issue alone is significant enough to make counterfeiting a large regulatory concern, and a key focus of the FDA and similar administrations internationally. These governing organizations are responsible for the identification of counterfeit products and ensuring that they are accurately attributed to the source for legal action and accountability.[1, 2]

However, the presence of counterfeit pharmaceuticals has additional ramifications on the pharmaceutical market, including causing deleterious impacts on the availability and cost of the drug product. The presence of a counterfeit product in the market may disrupt the supply and demand balance of the product, leading to manufacturing surplus or shortage, and corresponding artificial inflation or deflation of the drug cost.[2] Economic and marketing analysts specializing in

the pharmaceutical field attribute these issues to confusion within the supply chain, as the outside source of product (albeit counterfeit material) is undocumented and therefore obscures the true state of supply.

The issue of supply chain tracing within the pharmaceutical industry has become of increased importance over the past year due to the ongoing COVID-19 pandemic, which has caused shortages and manufacturing interruptions for many products. In February of 2020, with the pandemic looming, several economic and marketing analysts in the pharmaceutical industry highlighted the issue of transparency surrounding the manufacturing of generic pharmaceuticals in the US, mainly focusing on the fact that the manufacturing location and company of these products is not public information.[3] This article was a follow-up to one published in 2015, which had emphasized the same issue and requested that the FDA be authorized to publicize manufacturing information about generic products; however, no changes to policy have been made.[3, 4] This creates an additional issue for analysts attempting to predict the supply and demand of pharmaceuticals as there can be no communication between manufacturers and analysts regarding the source of pharmaceuticals.

When analyzing pharmaceuticals, identification of counterfeits and differentiation of counterfeits from genuine articles can be accomplished by a number of analytical methods. Most commonly, liquid or thin layer chromatography (TLC) and nuclear magnetic resonance (NMR) are used for this purpose.[1, 2, 5] These methods are effective, and are especially useful for identifying deficiencies in active pharmaceutical ingredients (API) or the presence of impurities, which can be inert but unregulated excipients or unsafe compounds, but they also have some impactful limitations. One primary issue is that these methods are destructive, resulting in a total loss or reduction of sample quantity, which can limit retesting possibilities. Additionally, these

methods require substantial sample preparation, separation, purification, and analysis time, resulting in higher cost of analysis and delay in obtaining results.[2]

To overcome these challenges, vibrational spectroscopy, such as near-infrared spectroscopy (NIR) and Raman spectroscopy, has been used as an alternative for identifying counterfeit pharmaceuticals. Vibrational spectroscopy alone has limited usefulness for assessing mixtures. However, when coupled with chemometrics, a multivariate data processing framework, its applications increase substantially. Several authors spanning the previous decade have reported that chemometric analysis of spectroscopic data can distinguish genuine and counterfeit pharmaceutical products utilizing several different projection methods and attribute counterfeit tablets to a source when a reference is available.[1, 2, 6, 7] The research has investigated a range of sample tablet sources and active ingredients, including acetaminophen and tadalafil, and successfully identified tablets from unlicensed sources and tablets beyond their expiration.[2, 7, 8] Tablets may display a reduction in API quantity, an increase in API heterogeneity, amorphous or polymorphic content, or impurities, all of which serve to create distinguishable variation in shifts or intensity of peaks in the Raman spectra.

Vibrational spectroscopy combined with chemometric analysis has been well documented as a viable method for the analysis of both genuine and counterfeit pharmaceuticals.[6, 7, 9, 10] Two main studies that served as support for the current work demonstrate this utility on samples of acetaminophen tablets. The first, a 2010 study, presented a successful application of NIR spectroscopy with principal component analysis (PCA) and hierarchical cluster analysis (HCA) for the distinction of genuine acetaminophen tablets from different countries of origin.[9] Small variations in spectra were modeled as discrete groupings of different brands in the PCA model. The authors outline a procedure for the classification of

acetaminophen tablets of different sources based on polymorph composition using NIR and chemometrics. Samples were gathered from international name brands, including Tylenol from the USA and Panadol from the United Kingdom. Though this article successfully demonstrated proof of concept for distinction and classification of genuine acetaminophen tablets, additional investigation into similar methods was not observed elsewhere in the literature.[9]

The general conjunction of vibrational spectroscopy and chemometrics appears often in investigations of counterfeit pharmaceuticals. The second study, a key study to the motivation of the current work, utilized Raman spectroscopy with HCA in order to accurately distinguish between genuine and counterfeit acetaminophen tablets.[6] This study also investigated different sample preparations, such as intact tablets versus ground tablets, and different strengths of genuine tablets from the same brand. While the most significant results were those of the genuine and counterfeit differentiation, promising results of genuine article differentiation were also observed.[6]

The ongoing issue of manufacturing transparency in generic pharmaceuticals, along with the increased need for the monitoring and tracing of pharmaceutical supply chains in the light of the COVID-19 pandemic, served as motivation for the investigation of the application of Raman scattering with chemometric analysis for the classification of acetaminophen tablets and differentiation of tablet brands and lots from generic and name-brand sources.[3]

The goal of the work was to provide proof of concept for the use of Raman spectroscopy and chemometrics for the assessment of inter- and intra-brand variability of pharmaceutical products. To achieve this, replicate samples of acetaminophen from varying brands and manufacturing lots were analyzed via Raman spectroscopy and the resultant spectra were projected by PCA to determine if the method was able to distinguish between tablets of different

brands, between tablets of the same brand but different manufacturing lots, and between tablets of the same brand by different API strengths.

### 1.2 Acetaminophen

Acetaminophen, otherwise known as paracetamol, is an analgesic, or pain reliever, and cited as the most common fever reducer and pain reliever worldwide, used both alone and in tandem with other APIs.[6, 11] Acetaminophen tablets are widely available, both domestically and internationally, from a range of name-brand and generic sources. Acetaminophen is most often obtained over-the-counter, but there are also over 25 million prescriptions written yearly in the US. The majority of prescriptions will be filled from generic sources rather than with name-brand product. Name-brand acetaminophen within the United States is available as Tylenol, manufactured by a subsidiary of Johnson and Johnson, which nets over 300 million dollars yearly.

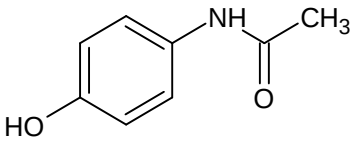



**Figure 1.** Acetaminophen structure.

Acetaminophen, or N-acetyl-para-aminophenol (Figure 1), is a *p*-aminophenol derivative first synthesized in 1877 by Harmon Northrop Morse at Johns Hopkins University by the reduction of *p*-nitrophenol in glacial acetic acid.[12] The compound exists in several forms of crystal structures. One main crystal structure is a monoclinic polymorph (form I); however, form II (orthorhombic) can also be isolated at standard conditions and three additional forms exist at high pressure conditions.[13] The monoclinic form is used for the production of acetaminophen tablets, however, the crystals alone exhibit poor compression behavior and therefore the pills are

produced by the tableting of monoclinic acetaminophen with various excipients in order to ensure proper molecular packing and tablet hardness.[11]

Although industrial procedures are highly similar between manufacturers, there are several possible sources of variations in genuine acetaminophen tablets, which we may be able to detect using Raman spectroscopy. The variations are most likely to result from the differences in the amount and the kind of inactive ingredients between samples. All brands are likely to have some variations in inactive ingredients, or excipients, used in the manufacturing process. However, even within the same brand we can expect possible differences between manufacturing lots because many excipients are interchangeable. As mentioned above, the excipients mainly serve to produce a drug tablet with greater structural integrity and hardness while not affecting the physical and chemical properties of the API.[11] A variety of excipients may be used interchangeably in manufacturing based on cost and availability.[14] In fact, we can often see the drug label stating, "one or more of these ingredients may be present" among the inactive ingredients listed on the label. Thus, the inactive ingredients and their relative abundance within a specific manufacturing lot may vary. Considering the excipients within a standard acetaminophen tablet ranges from 10-20 wt. %, these variations in ingredients could contribute to slight variations in the Raman spectra. Another possibility is a degradation of ingredients in the tablet depending on the tablet age or storage conditions like temperature and humidity.[15] This degradation may contribute to the variations in the composition of a drug tablet.

### 1.3 Raman Spectroscopy

Raman spectroscopy is best known for its usefulness as a chemical identification method that is applied to inorganic, organic, and biological materials. Samples can be analyzed nondestructively with minimal sample preparation by Raman spectroscopy, which makes it an

ideal method for the rapid identification of compounds.[16] Applications of Raman have significantly broadened over two decades in thanks to the lowered cost of instrumentation and the integration of imaging and spectroscopy systems to expand the method's utility.[17, 18] Raman spectroscopy now holds a place not only in research laboratories, but also in forensics, manufacturing, agriculture, art history, medical applications and beyond.[16]

The basis of Raman spectroscopy is to collect and separate inelastic light scattering from molecules at wavelengths different from that of the incident light. This shift of wavenumber (or frequency) in Raman scattered light is called Raman shift and occurs due to interaction between the light and the vibrational modes of a molecule. Thus, the Raman shifts represent the vibrational frequencies of a molecule. A vibrational mode is said to be Raman active if an energy exchange with the incident light can result in a change of the polarizability of the molecule. The incident light will either be scattered at a lower energy, which will result in Stokes Raman scattering in which the scattered light has a longer wavelength, or scattered at a higher energy, resulting in anti-Stokes Raman scattering in which the scattered light has a shorter wavelength.[16]

The Raman instrument is designed in order to allow for the collection of inelastically scattered light from molecules. The coupling of a Raman instrument with an optical microscope allows for precise location of a sample surface to collect Raman scattering signals, and construction of a depth profile of a sample if a confocal pinhole is adopted in the optical system. Thus, Raman microscopy belongs to chemical microscopic methods.

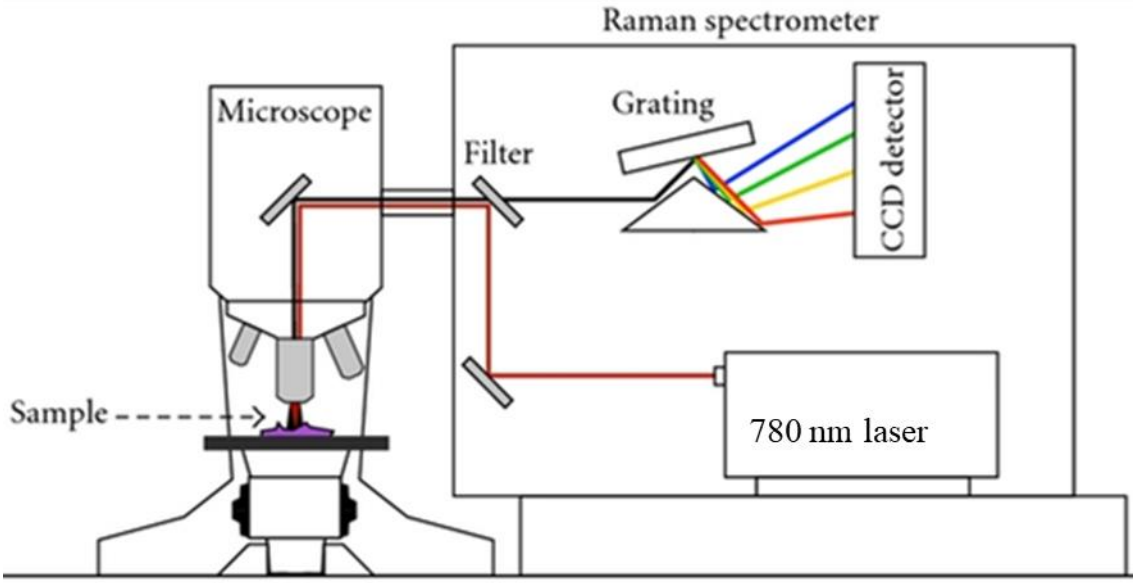



**Figure 2**. Schematic diagram of a Raman microscope.

Figure 2 outlines a schematic diagram of a Raman microscope. A sample surface is first focused by controlling the distance between the sample and the objective lens, while the surface image is observed through an eyepiece or a charge coupled device (CCD) camera. Since Raman scattering intensity is intrinsically weak, a light source with strong beam intensity is required. Thus, the incident light source within a Raman instrument is typically a monochromatic laser in the visible or near-IR wavelength range. The laser beam is focused on the sample surface through the objective lens and the scattered light from the sample is collected through the same objective lens. This scattered light is then passed through a notch filter which will remove Rayleigh scattering (elastic scattering) and transmit Raman scattering. Since Stokes Raman scattering is stronger than anti-Stokes Raman scattering, most commercial instruments are designed to detect Stokes Raman scattering. The inelastically scattered light that passes through the filter is sent to a grating which separates it into individual wavelengths of photons. These are then detected by a CCD detector and transcribed into a spectrum based on their intensity.

Since the positions and the relative intensities of the Raman peaks of a molecule is unique depending on the molecular structure and chemical bonds, Raman spectroscopy can be used to identify a molecule. Raman spectroscopy and Fourier-transform Infrared spectroscopy (Ft-IR) are complementary techniques, both capable of obtaining structural information based on the vibrational modes of a molecule. While Raman activity requires a change in polarizability of the molecule, IR activity depends on a change in molecular dipole moment, so certain modes may be only IR-active or only Raman-active depending on the symmetry properties of vibrational modes. This makes each method suited for different applications.

Raman spectroscopy has several key advantages over IR spectroscopy. First, the Raman bands are much sharper than the IR absorption bands, resulting in a better resolution of peaks. Second, the Raman spectroscopy is more useful for aqueous samples as water produces very weak Raman signals, while it shows strong absorption bands in an IR spectrum, which may interfere with analyte bands. The primary advantage for the current project is that the Raman scattering allows for the acetaminophen tablets to be tested in their intact form, while Ft-IR depends on light transmittance and therefore has thickness constraints that must be considered in sample preparation.[16] Thus, Raman spectroscopy does not require special sample preparation steps.

On the other hand, Raman analysis traditionally has limited discriminatory power for a sample that consisted of various mixtures, and is generally not useful for differentiating between similar samples when relying on direct comparison or library search. The ease of use and short analysis times still make Raman an ideal technique for many applications, therefore methods of expanding its practical utility are sought after.[6, 16] The combination of Raman spectroscopy with

chemometric analytical methodology has been the subject of much research over the last several decades.[6, 19]

### 1.4 Chemometric Analyses

Chemometrics can be simply described as the integration of chemical analysis with multivariate statistical methods to expand and optimize experimental methodology. This system of statistical analyses is the result of multivariate theories developed alongside applied mathematics to functionalize data and strengthen the inferences researchers are able to make from it.[4] A main catalyst of the development of chemometrics was the rapid growth of quantitative analytical chemistry throughout the late 19th and 20th centuries. This growth was driven by the development of instrumental analytical methods which generated an explosion of raw multi- or mega-variate data. This data was not reconcilable using traditional data processing methods and chemometrics played a key role in allowing researchers to meaningfully employ it.[20]

Chemometrics was initially known as "chemical pattern recognition" prior to its formal designation and the methods are designed and adapted to investigate features of multivariate chemical data sets. Though quantitative analytical chemistry catalyzed its development, chemometric methods are broadly applied and useful in providing additional depth to studies in many fields.[5] The successful application of chemometrics to a research study requires the study be designed with an awareness of the intrinsic link between the chemical data collection and the statistical framework that will be applied to present meaningful results. Chemometric studies generally proceed through five phases of analysis: design of experiments, sampling, data collection, data preprocessing, and data projection.[20, 21]

Design of experiments (DOE) and sampling are the preparation and planning stages of the experiment, and both serve to ensure that the resulting data is representative and complete.

DOE optimizes the analytical process while maintaining budgetary and time constraints on a study. In any experiment, specifically instrumental analyses, multiple parameters influence experimental results. When optimizing, testing across a range of each parameter is necessary, and ideally testing would be done as a parameter interaction matrix in order to avoid false maxima. However, when there are many parameters to test, testing each combination while varying one factor at a time is seldom feasible. DOE calculation creates a formalized design in order to minimize trials required to create a map of the experimental range to model how parameters interact. The design is done by coding each parameter range into 'high' and 'low' conditions and using software to create a range of experiments fitting an established framework to gather descriptive data.[20, 22]

When DOE has established the experimental setup, the theory of sampling (TOS) provides a theoretical framework establishing appropriate sampling methods. Proper sampling is required to accurately characterize a population or heterogeneous material, and TOS ensures that samples are sufficient for representation. TOS encompasses two realms of sampling, physical and statistical. Physical sampling gathers samples from a heterogeneous source in order to establish mixing, composite, and heterogeneity. Statistical sampling involves sampling and resampling from multiple sources, lots, etc. to provide cross-validation and promote the development of classifications.[15, 20]

The next step, data collection, is the chemical analysis portion of the process. After collection, raw data must be preprocessed prior to analyses to ensure accuracy and minimize the introduction of error in the projection process. Preprocessing methods are of two types" row-wise and column-wise. Row-wise preprocessing methods are applied to all variables of a sample and include filtering, peak alignment, and baseline corrections. Column-wise preprocessing

involves application of the calculation to a specified variable across all samples. Examples of column-wise preprocessing include weighting, scaling, and centering.[20]

The final phase is the data analysis utilizing chemometric projection methods in order to process and visualize the data. Analytical projection methods are designed for a variety of applications and can be used to demonstrate correlations, recognize patterns, classify results based on characteristic features, demonstrate consistencies, and identify distinguishing features of data that could not be shown by other methods. The type of projection method used depends on several features of the data including the number of variables, dimensionality, data type, and set size. The most common methods are multiple least squares regression, component analyses, predictive modeling, and multivariate classifications.[20, 21] The coupling of Raman spectroscopy with chemometrics now has a wide range of established applications from cancer diagnosis to the discrimination of foreign fats and oils in dairy products.[23, 24]

**1.5 Principal Component Analysis (PCA)**

There are a number of chemometric projections with potential for the comparative analysis of pharmaceuticals including multiple linear regression (MLR), hierarchical cluster analysis (HCA), and principal component analysis (PCA).[6, 9, 25] While each method has advantages, PCA was chosen for this work as it showed consistent success in the differentiation of similar samples with the characteristic peaks of a single component overwhelming the Raman spectra. PCA is generally appropriate for spectral data due to its reliance on the covariance of variables, making it capable of detecting small sources of variance, and its high power for dimension reduction creates a far easier interpretation than the raw, multi-variate data set.[20, 26, 27]

In fact, the main focus of PCA is dimensionality reduction, which is accomplished by modeling sets of dependent variables in the sample data along single dimensions.[26] This allows

for the reduction of the number of dimensions needed while still reflecting the majority of total variance within the raw data set. The dimensions calculated to represent each set of dependent data are termed principal components, or PCs. Each PC refers to a single dimension of the projection, which reflects one grouping of dependent variables. Each of these PCs is calculated as the linear combination of variables within the data set. The specific PCA projection depends on the number of PCs chosen for modeling, which will correspond to the total cumulative variance that is modeled by the PCA projection as a whole.[27]

Each principal component is orthogonal, therefore, uncorrelated with the others, creating a basis set of dimensions in which to model the variance. The orthogonal components are calculated by performing an orthogonal decomposition of the data covariance matrix. The decomposition formula is shown in eq. 1.[27]

$$X = t_1 {p_1}^T + t_2 {p_2}^T + \cdots + t_k {p_k}^T + E \qquad \text{(eq. 1)}$$

The data matrix, X, is decomposed into vector components of loadings and scores and has dimensions of *m* samples and *n* variables. Loading vectors, $p_i$, and scores vectors, $t_i$, are calculated by determining the orthogonal basis of the matrix. The decomposition begins by determining the eigenvectors of the covariance matrix of X (eq. 2), which correspond to the loading vectors, $p_i$. The covariance matrix is the representation of how each of the variables within the original data set, which contain linear dependence on each other, show dependent variance.

$$cov(X) = \frac{X^T X}{m-1} \qquad \text{(eq. 2)}$$

The eigenvectors of the covariance matrix are described as the 'characteristic roots' of the matrix, such that the product of a scalar (the eigenvalue, $\lambda$) and the eigenvector ($p_i$) is equivalent to the product of the covariance matrix and the eigenvector (eq. 3).[26, 27]

$$cov(X)p_i = \lambda_i p_i \quad \text{(eq. 3)}$$

The eigenvalue is proportional to the amount of variance being captured by the dimension. The loading vector, a modeling single independent variable within the covariance matrix, then allows us to calculate the score vector via eq. 4, which is the projection of the data matrix onto the loading vector to yield the score vector for the dimension.

$$t_i = Xp_i \quad \text{(eq. 4)}$$

Each pair of score and loading vectors represents a single principal component within the decomposition of the data matrix. The number of scores and loadings vectors in the decomposition depends on how many principal components are deemed necessary to model the meaningful variance within the data set. Any remaining variance not modeled by the PCs is captured in the error matrix, *E*.

As suggested by their derivation, the loading vectors, $p_i$, and score vectors, $t_i$, describe the relationships between variables and samples, respectively. The loading vectors reflect how highly weighted a given variable is, or how much of the variance being modeled is observed at a given variable. The scores vectors indicate how strongly a sample is correlated with a given dimension. A sample which exhibits strong signal at variables with a high positive load will

score strongly positively along that PC, while samples which display low signals at variables with a high positive load may score negatively along the PC.[20, 27]

When performing PCA on spectral data, there is one phenomenon, generally only observed along PC 1, which results in the loading plot that resembles the sample spectra. This is because PC 1 captures the majority of the variance within the data set, which arises from the characteristic peaks of the compound(s) within the sample. The variables with the strongest signals will have the largest absolute variance, and therefore be more strongly loaded than those with lower signals. This resemblance of loading plots to the spectral data should not persist past the first PC, unless there are additional sources of known variance in the data, such as multiple chemical components each contributing characteristic peaks to the spectra.[27]

## II. Materials and Methods

### 2.1 Acetaminophen Sampling

A total of ten acetaminophen samples were purchased locally in order to test the intra- and inter-brand variability of the tablets. To test inter-brand variability, six different extra strength tablet samples were purchased from generic brands, and to test intra-brand variation three replicates of extra strength Tylenol brand were purchased, each from a distinct lot, and one regular strength Tylenol brand sample was purchased for comparison. The extra strength tablets contain 500 mg of acetaminophen per tablet, and regular strength contains 325 mg of acetaminophen per tablet.

The information of the tablets sampled, including their brand, the vendor, their NDC, weight, and API wt. % is summarized in Table 1. The inactive ingredients listings from each brand are included in Appendix 1. Three tablets from each sample were scanned by Raman microscope in order to examine possible intra-lot variation of the tablets.

**Table 1**.Physical and brand information of purchased acetaminophen tablets.

| Brand | Abbreviation | NDC | Vendor | Lot # | Ave Weight, n=12 | API w/w%, n=12 |
|---|---|---|---|---|---|---|
| TopCare Health | TC | 36800-484-62 | Wegmans | ODE2413 | 0.553 | 90.11 |
| Ready In Case | RIC | 59726-494-24 | Walgreens | H09800 | 0.612 | 80.32 |
| Walgreens | Wal | 0363-0175-08 | Walgreens | P119534 | 0.554 | 90.66 |
| Equate | EQ | 49035-308-82 | Walmart | 0EV1292 | 0.553 | 89 |
| Family Wellness | FW | 55319-494-50 | Family Dollar | OFE2816 | 0.553 | 89.08 |
| Up&Up | UU | 11673-838-05 | Target | C030G | 0.556 | 89.63 |
| Tylenol | SEA041 | 50269-449-35 | Family Dollar | SEA041 | 0.61 | 81.1 |
| Tylenol | SEA035 | 50269-449-35 | Wegmans | SEA035 | 0.607 | 82.5 |
| Tylenol | SHA101 | 50269-449-35 | Dash's Markets | SHA101 | 0.601 | 82.21 |
| Tylenol | Reg | 50580-495-01 | Wegmans | PMA121 | 0.389 | 83.32 |

### 2.2 Raman Method

Raman spectra of all the samples were obtained with the confocal DXR Raman microscope (Thermo Fisher, Waltham, MA) equipped with a 780 nm diode laser, a CCD detector (2048 pixels), and an automated microscope stage. The set-up parameters for Raman microscopy were slightly modified from those used in others' previous reports.[9,28] Raman data were collected and processed using OMNIC software. We optimized the number of Raman scans (number of Raman spectra for average) and exposure time (detector integration time) considering both the signal to noise ratio (S/N) of Raman spectrum and total data acquisition time for a sample.

Samples for analysis were placed on microscope glass slides and focused using a 10× objective lens. The laser power was set to 2 mW. The spectrograph aperture was set to 50 μm slit width. The CCD exposure time was set to 4 s. The Raman spectrum of each sample was the result of averaging 32 exposures after the automatic background subtraction. For each tablet sample, three Raman spectra were collected at three different locations along the long axis of the tablet using the same instrumental parameters. Then, using an analysis tool in the OMNIC, these three Raman spectra were averaged to represent the composition of each tablet sample.

**2.3 Tablet Preparation**

Tablet samples were initially prepared in three different ways: pulverized powder, and tablet with and without a top coating layer. This was to establish the preferred method of sample preparation for further tests and to investigate if there were discernible differences in the Raman spectra of the samples that were prepared in different ways.

The powder sample was prepared by pulverizing using a mortar and pestle and a small amount of the power was placed on a microscope slide to acquire its Raman spectrum. To remove the top coating layer of a tablet sample, the top surface was first ground using a coarse sandpaper and polished using a finer grade to smooth the surface. A sample of six tablets from a single lot of Tylenol brand acetaminophen were prepared as described above and tested for homogeneity using a 2-D Raman mapping, as suggested by Kwok and Taylor.[7] Each point in the Raman map was gathered using a 4 s exposure time and 20 exposures in order to reduce acquisition time while maintaining an acceptable S/N. The 2-D mapping covered a 0.5 mm × 0.5 mm area, with scans at 0.1 mm intervals for a total of 25 points per Raman map.

The Raman mapping experiments showed no distinguishing features between tablets of the same lot, and no visible variation in spectra gathered from the tablet form versus the

powdered sample. Therefore, subsequent testing was performed using only tablet samples with and without the coating layer.

### 2.4 Data Preprocessing

Prior to any following data analysis, the wavenumber range of Raman spectra of all tested samples was adjusted to 35 - 1850 $cm^{-1}$. Then, the baselines of the spectra were first corrected using the auto-baseline correction tool in the OMNIC software. Following correction, the spectra were overlaid to inspect any irregularities in the spectra caused by the correction. After the baseline correction, the Raman spectra obtained at three different locations on the tablet surface were averaged for further chemometrics analysis.

In order to remove any non-significant variance within the data and improve the results of chemometric projections of the data, additional preprocessing of mean centering was performed. Mean centering is a column-wise preprocessing technique which is performed by subtracting the mean of a variable across all observations from that variable in each observation. When considering a group of Raman spectra, this is performed by subtracting the average signal intensity at a specific wavelength from the signal at the same wavelength across all replicate spectra.[20] This forces a mean of zero for all variables and places all variables on the same scale, rather than having some being more highly weighted due to stronger signal intensities. Mean centering is appropriate for spectra data sets as it only considers equivalent variables, individual wavelengths, and does not depend on the data as a whole. Mean centering, and the following PCA projections, were performed on MATLAB using a PLS Toolbox chemometrics software package.

### 2.5 Principal Component Analysis (PCA)

The mean-centered Raman spectra were projected via a PCA method in order to more accurately and sensitively identify distinctions and groupings between tablets and brands.

Several data sets were created in order to test what combinations of brands and what ranges of wavenumbers within the Raman spectra were capable of producing the most robust classification of tablets and brands. Thirteen data sets prepared to build a PCA model are summarized in Table 2. Four projections (data set #1 - #4) yielded useful information, and the results of each will be presented in the Results and Discussion section. The other data sets (#5 - #13) did not generate useful information and the results will not be discussed.

**Table 2**. Data sets for the PCA projection.

| Data set | Tablet Preparation | Brands | Wavenumber Range |
|---|---|---|---|
| #1 | With top coating | All | 35 - 1850 $cm^{-1}$ |
| #2 | With top coating | RIC and UU excluded | 35 - 1850 $cm^{-1}$ |
| #3 | With top coating | All | 700 - 1850 $cm^{-1}$ |
| #4 | Without top coating | All | 35 – 1850 $cm^{-1}$ |
| #5 | With top coating | All | 1000-1850 $cm^{-1}$ |
| #6 | With top coating | RIC excluded | 35 - 1850 $cm^{-1}$ |
| #7 | With top coating | RIC excluded | 700 - 1850 $cm^{-1}$ |
| #8 | Without top coating | All | 700 - 1850 $cm^{-1}$ |
| #9 | Without top coating | All | 35 – 700 $cm^{-1}$ |
| #10 | Without top coating | All | 1000 - 1850 $cm^{-1}$ |
| #11 | Without top coating | All | 100 - 1850 $cm^{-1}$ |
| #12 | Without top coating | All | 150 - 1850 $cm^{-1}$ |
| #13 | Without top coating | EQ excluded | 35 - 1850 $cm^{-1}$ |

## III. Results and Discussion

### 3.1 Intact Tablet Spectral Results

The representative Raman spectra of 10 intact tablet samples, which were listed in Table 1, were compared with that of standard acetaminophen powder as shown in Figure 3. The complete Raman spectra of all the tablets (three tablets per brand) are included in Appendix 2. Overall, the Raman spectra of intact tablet samples were identical to that of the standard acetaminophen except for the generic brands, such as Ready-In-Case (RIC) and Up & Up (UU). Even for the RIC and UU brands, both of them showed main Raman peaks of acetaminophen standards. This suggests that the Raman scattering signals of all tested tablets mainly result from the acetaminophen and additional Raman peaks arise from the inactive ingredients. It is interesting to observe that there were slight differences in the absolute Raman intensities of the active ingredient in the samples, which may have been contributed from different amount of active ingredient present in the tablet.

There were noted discrepancies in the API wt. % among the tested tablet brands as shown in Table 1. There were two general groupings of brands. All Tylenol brands including both extra and regular strength had an API wt. % between 82-84%, while all generic brands except for RIC were close to 90 wt. % API. RIC was the only generic brand to have a similar API wt. % to that of the Tylenol brand samples. In general, the intensities of Raman bands that correspond to the acetaminophen appeared weaker in the tablets with 80 wt. % API than in those with 90 wt. % API as shown in Figure 3. On the other hand, there were no other distinctions on the Raman spectra as a result of the differences in API wt. %. Although the wt. % of excipients took up as high as 20 % of the weight of a tablet, the Raman peaks of these inactive ingredients did not appear because they consisted of small amounts of various compounds.

As mentioned above, the RIC and UU tablets showed more peaks than other tablets below 700 $cm^{-1}$. We suspected that these two generic brands may have a top coating layer different from other brands. To characterize the additional Raman peaks, the Raman spectra were collected of the RIC and UU tablets without the top coating layer. These scans of the interiors of the tablets were subtracted from the corresponding Raman spectra of the two brands with the coating layer. The representative result of the subtraction for the RIC is shown in Figure 4 and shows four main peaks, all appearing below 700 $cm^{-1}$. A similar result could be obtained for the UU. Performing a library spectral search identified the spectra as $TiO_2$, a conclusion further supported by the peak positions being highly consistent with those reported in the associated literature as presented in Table 3. The presence of $TiO_2$ as an ingredient is expected as it is present in the excipient lists for those the UU and RIC brands. The extra strength Tylenol replicate lots also include $TiO_2$ as an inactive ingredient but the Raman spectra did not exhibit the same peaks, likely due to the $TiO_2$ being homogeneously distributed throughout the tablet rather than in the coating layer. A small wt. % of $TiO_2$ distributed throughout the tablet would not appear strongly in the Raman spectrum, and therefore was not observed.

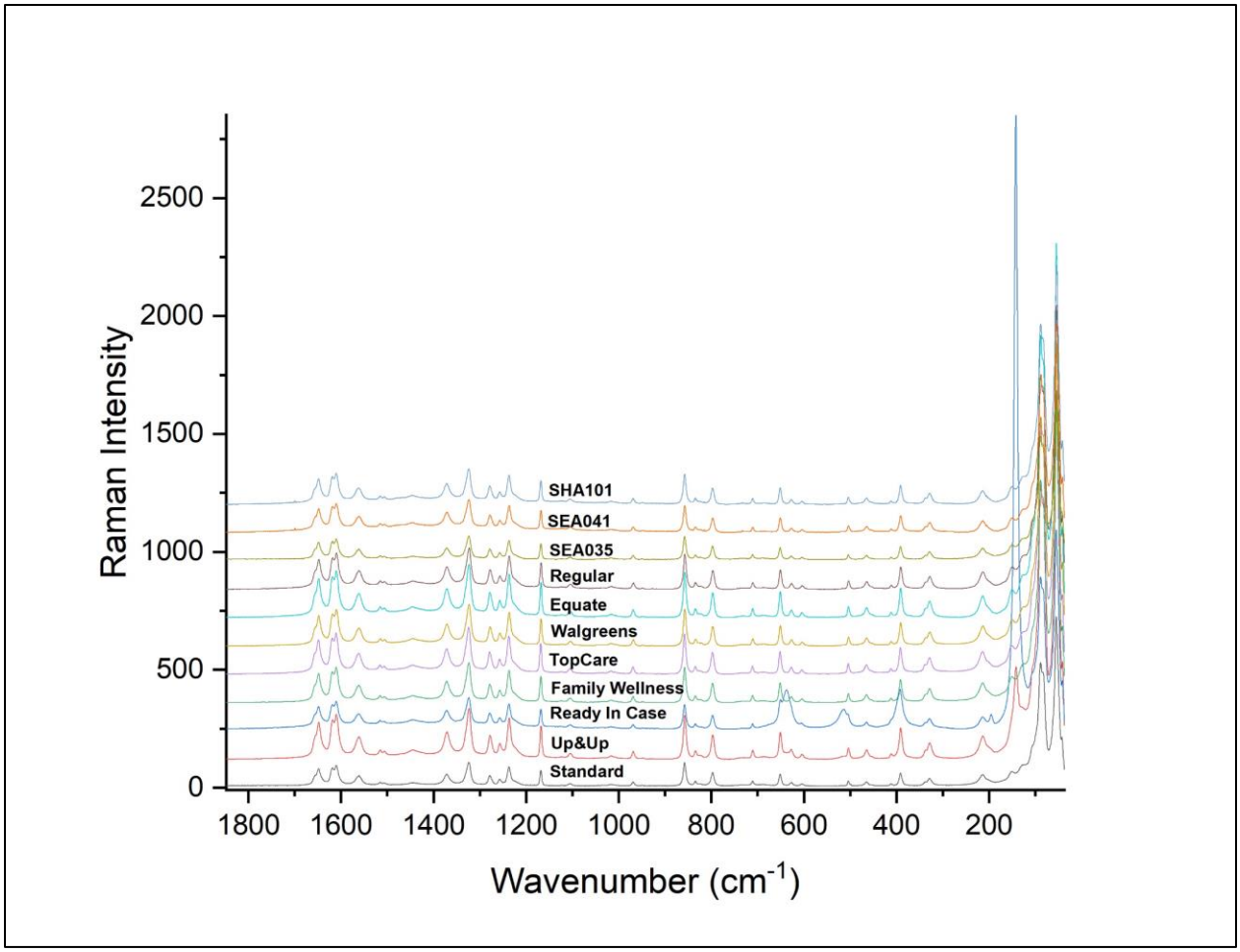



**Figure 3**. Representative Raman spectra of the intact acetaminophen tablet samples.

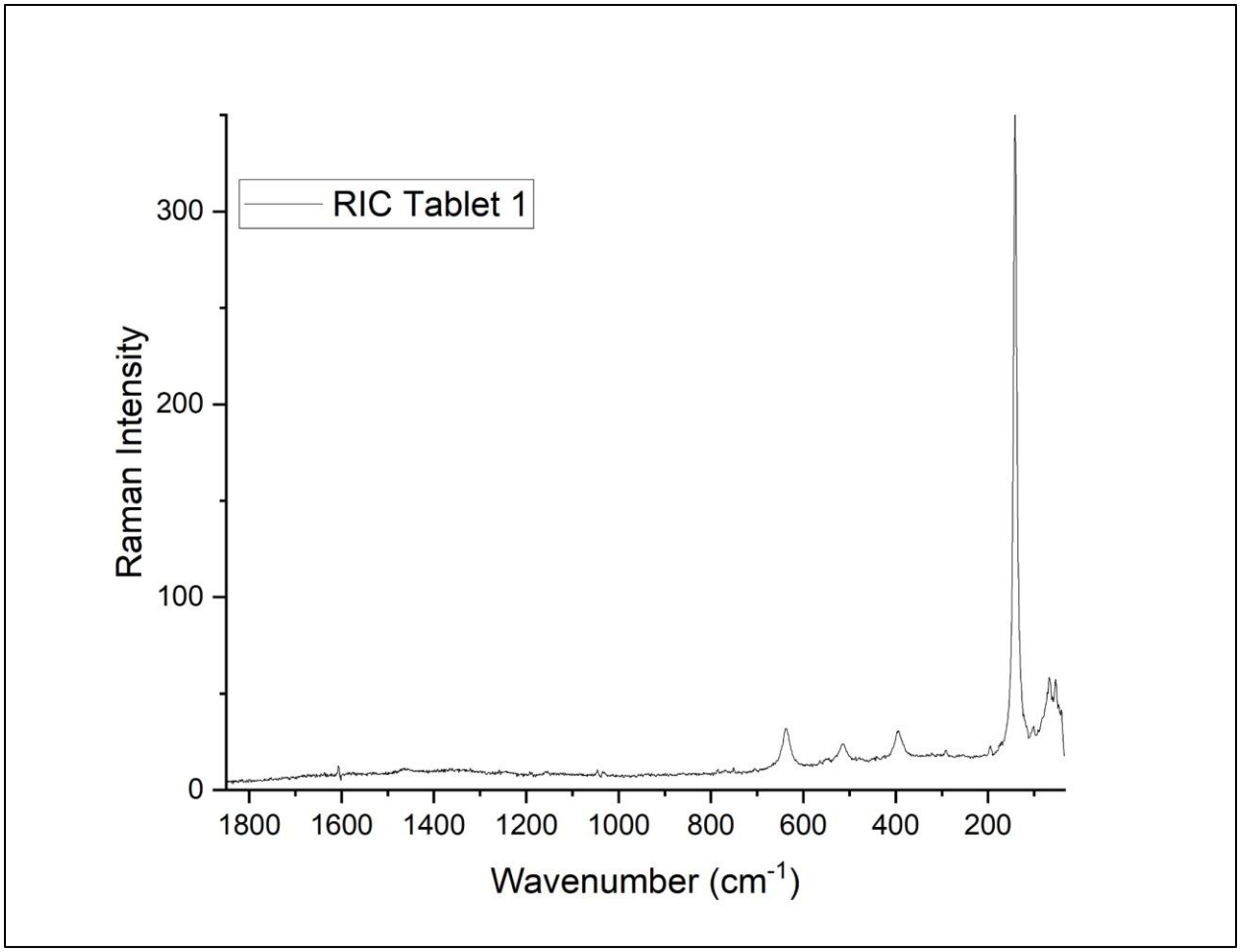



**Figure 4.** Results of the subtraction of the Raman spectra of the interior of a RIC brand tablet from that of the Raman spectra of the same intact tablet.

**Table 3.** Raman peaks of $TiO_2$ standard[29] and those observed in Figure 4.

| **$TiO_2$** | **Observed Peaks** |
|---|---|
| 144 $cm^{-1}$ | 142 $cm^{-1}$ |
| 394 $cm^{-1}$ | 394 $cm^{-1}$ |
| 514 $cm^{-1}$ | 513 $cm^{-1}$ |
| 634 $cm^{-1}$ | 637 $cm^{-1}$ |

Comparison of each of the UU and RIC intact tablets revealed several points of interest resulting from the presence of $TiO_2$ within the spectra. The 35 - 700 $cm^{-1}$ spectral range of the six tablets in which $TiO_2$ was observed are overlaid with a common scale in Figure 5. A primary point of interest is the variation in signal intensity, both between the tablets of like brands and between the two brands.

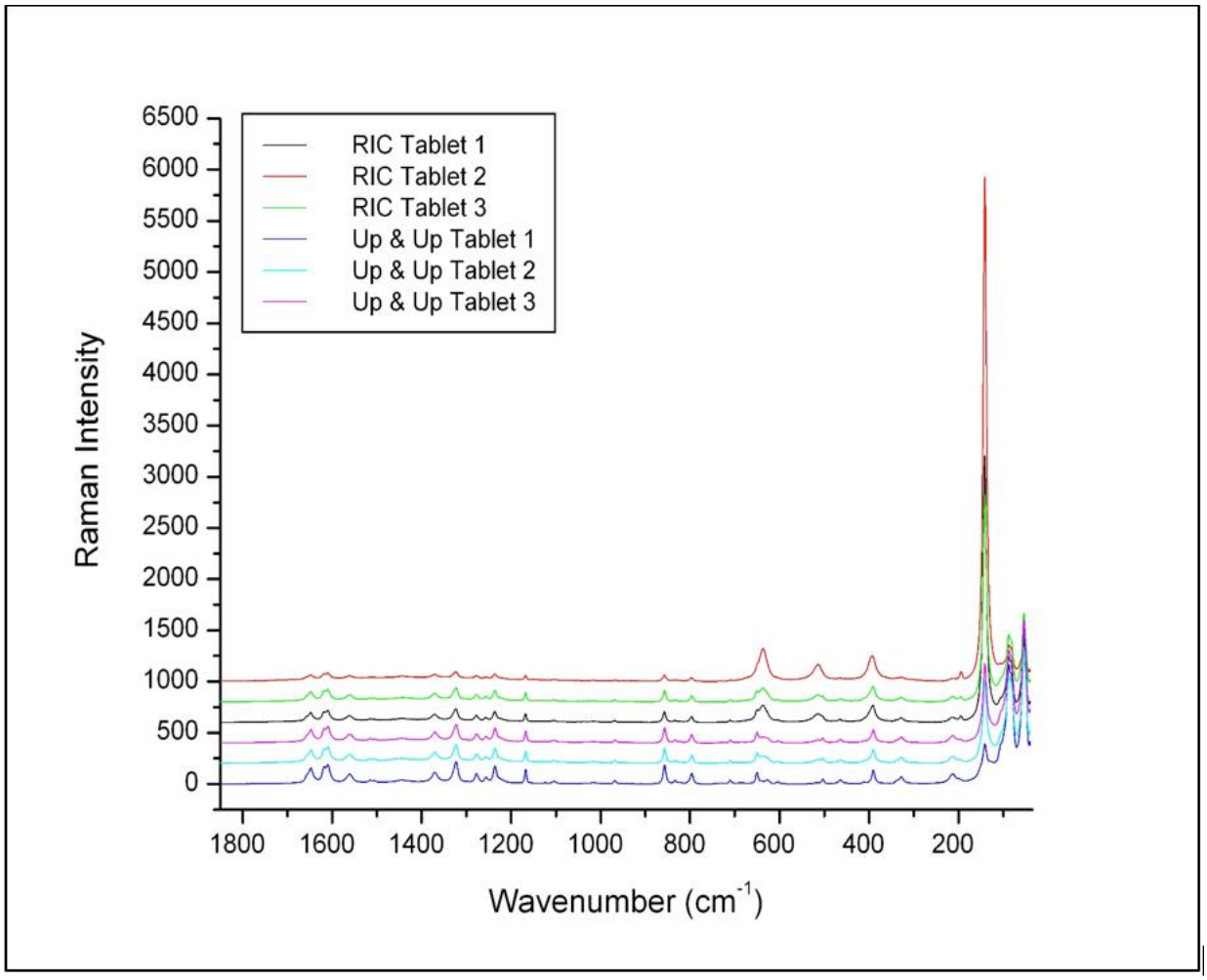



**Figure 5.** Raman spectra of all intact RIC and UU tablets tested.

The observed signal intensity between tablets of the same brand, while significant, was presumed to be due simply to inhomogeneity of the $TiO_2$ coating on the tablets. Evidence to support this was found first by examining the three replicate scans along the long axis of each

tablet which were averaged to produce the representative tablet spectra. As shown in Figure 6, these individual scans for RIC Tablet 1 reveal that there is signal intensity variation across the length of the tablet.

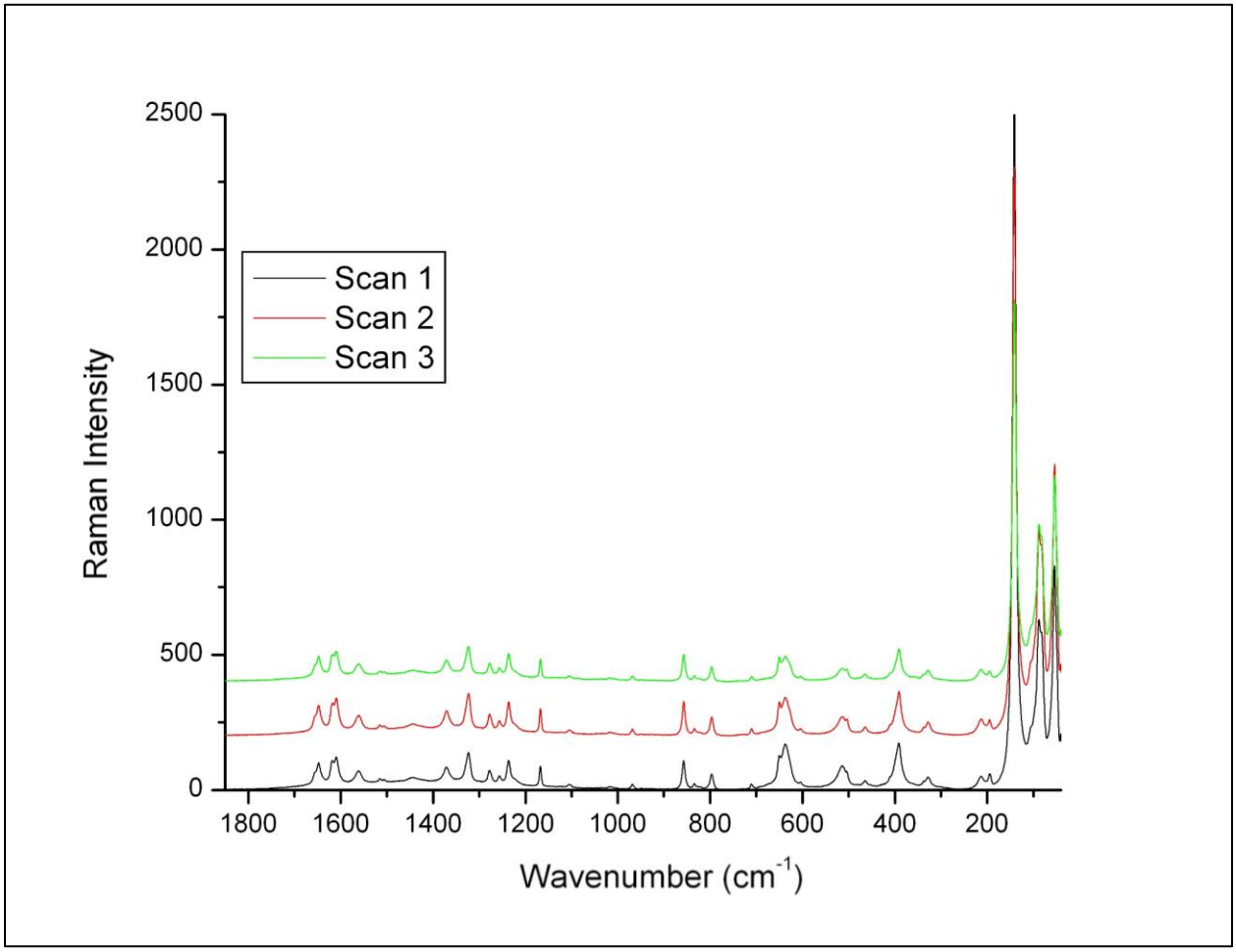



**Figure 6.** Raman spectra from three different locations along the length of a single RIC tablet.

To determine if this intensity variation is the result of $TiO_2$ inhomogeneity or photo-degradation of the sample by repeated exposure to the laser source, a RIC tablet was scanned multiple times in the same location. The data showed that there was no reduction in signal, suggesting that there would be no photo-degradation of the sample upon repeated exposure to the laser for the adjustments of optics prior to the Raman measurements or during the Raman

scanning. Therefore, it is expected that the $TiO_2$ coating varies in thickness both along individual tablets and between tablets of the same lot.

Even in light of the range of intensity within the brands, there was a clear distinction observed between the intensity of the $TiO_2$ peaks of RIC tablets and those of the UU tablets. The RIC tablets consistently presented higher signal intensity of $TiO_2$, so that these characteristic peaks effectively masked the characteristic acetaminophen peaks in the region lower than 200 $cm^{-1}$. The UU tablets, however, showed lower $TiO_2$ intensity and the characteristic acetaminophen profile is much more easily observed in these samples than in the RIC samples. These data suggest that they both use $TiO_2$ as a top coating layer but the RIC has a thicker $TiO_2$ layer than the UU.

Since the $TiO_2$ appears in only two tablet samples of the ten different tablet samples, this Raman profile of $TiO_2$ will contribute to a significant amount of independent variance. Based on the principles of PCA, this will be a likely convoluting feature. Two different approaches were attempted to see the dominant effect of $TiO_2$ peaks. First, the RIC and the UU brands were removed from the data set. Second, the wavenumber range of the Raman spectra in the data set was adjusted to exclude the characteristic Raman peaks of $TiO_2$, so that the Raman signals only between 700 and 1850 $cm^{-1}$ might be tested as shown in Figure 7.

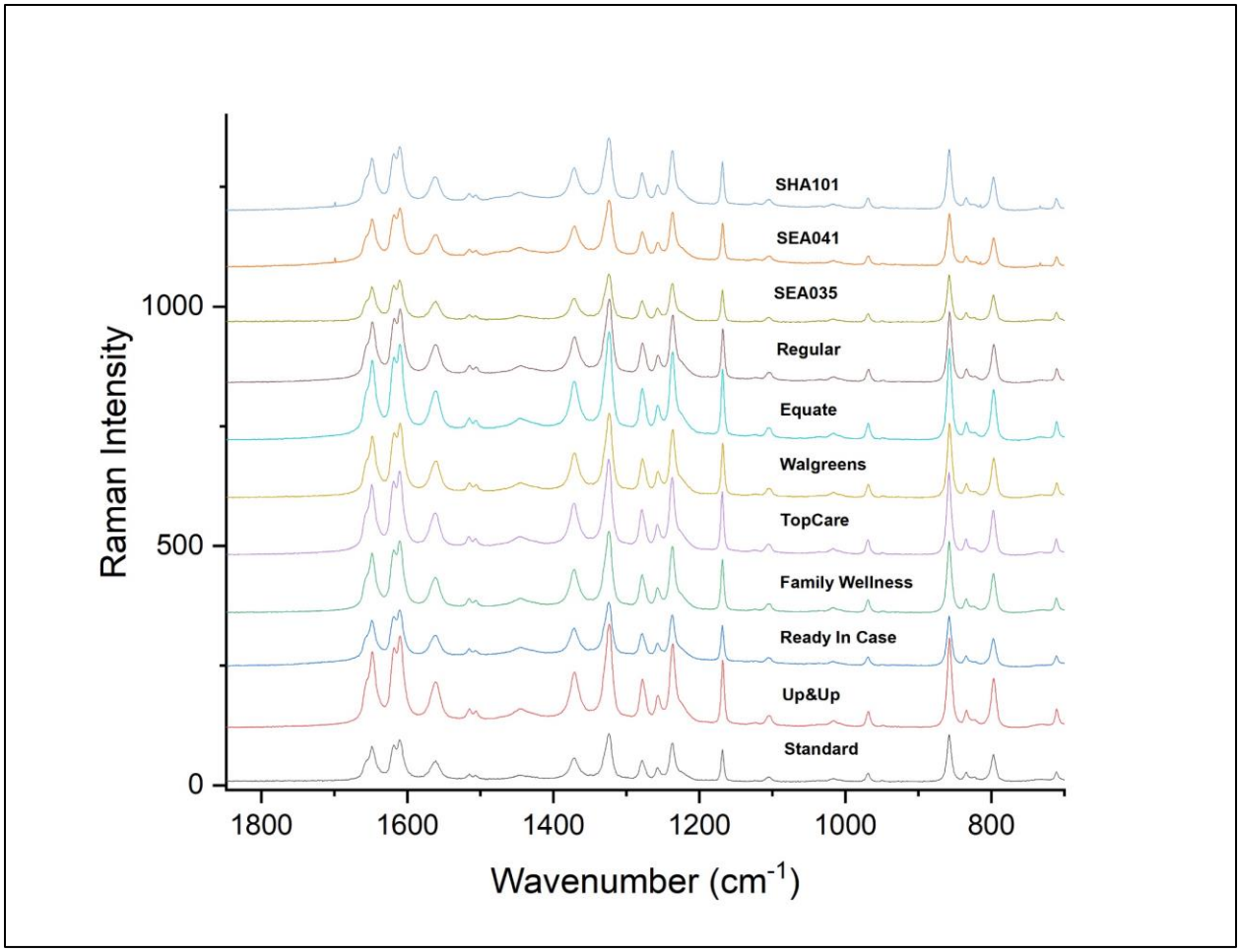



**Figure 7.** Upper spectral range (700 - 1850 $cm^{-1}$) of Raman spectra of intact acetaminophen tablets.

In order to further examine this spectral range, and to determine the variance observed between tablets within the same brands, the variance of the three average spectra for each brand was calculated using OMNIC software. The variance of each brand spectra is shown in Figure 8, with the brands ordered by decreasing variance. Examining these variance results we observe that all of the name (Tylenol) brand tablets demonstrate lower variance between tablets than the remaining brands, with the exception of Family Wellness, the only generic brand to show similar variance levels. This indicates that the tablets of the Tylenol brand, both extra and regular strength, are more consistent within the manufacturing lots, while the generic brands are more

varied in composition. These trends within the variance spectra may be compared with the PCA score plot. Larger variance between tablets is expected to correspond to a wider array of the tablets in the scores plots, whereas lower variance allows for closer groupings.

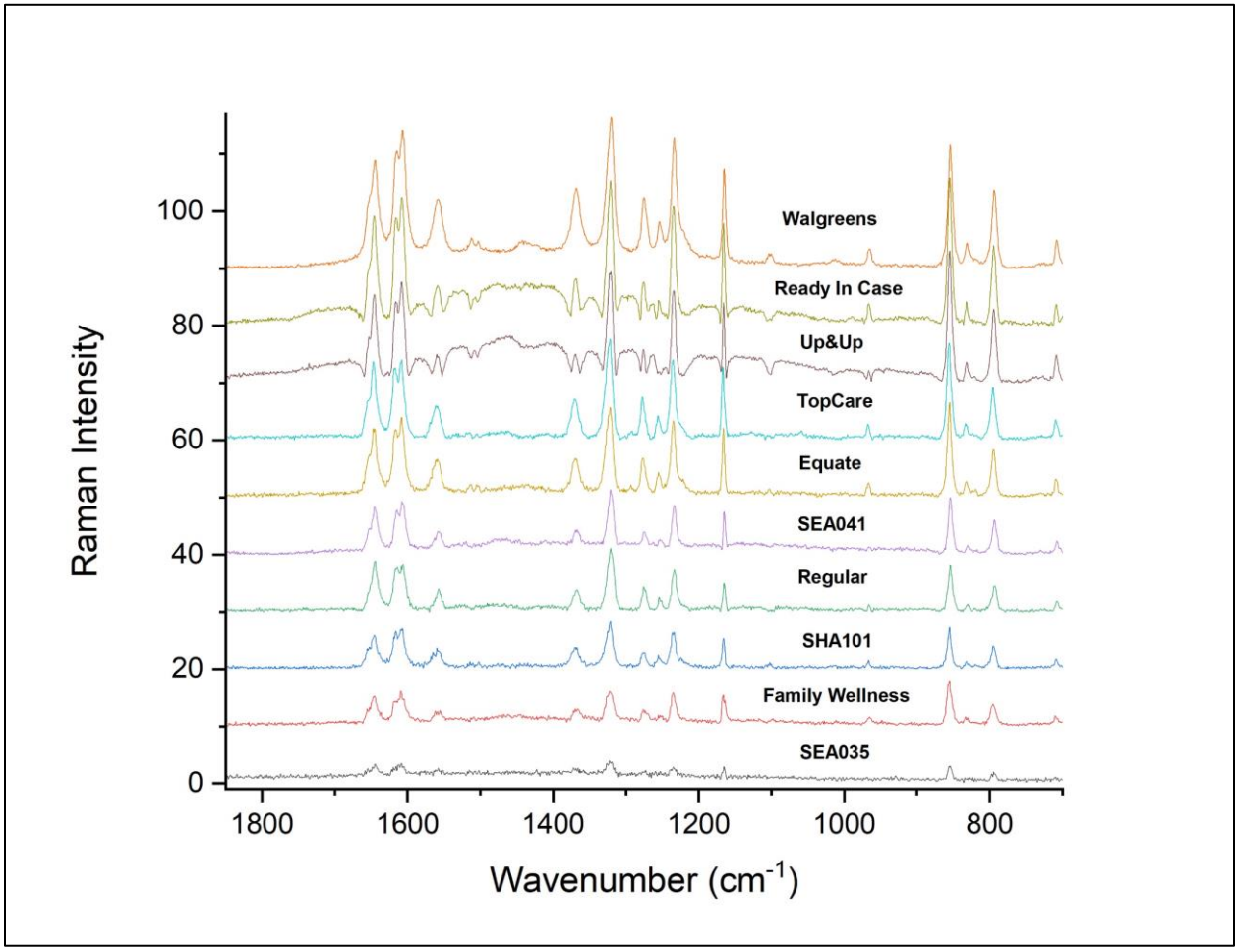



**Figure 8.** Variance of the Raman spectra of the intact surface of three tablets for each acetaminophen brand, arranged in order of decreasing variance.

### 3.2 Interior Tablet Spectral Results

The Raman spectra of the interior of tablet samples were obtained after the top layer of the tablets was removed. The representative Raman data of the samples are shown in Figure 9. These spectra look all similar to each other, but is unsurprising as the characteristic Raman peaks of $TiO_2$ of the two brands no longer appeared after the top layer of the tablets was removed. The Raman peaks in those spectra can be attributed to the active ingredients in the tablet. Although it

is not possible to differentiate the acetaminophen brands using this raw Raman data, it may be possible to do so by performing PCA projection on the Raman data.

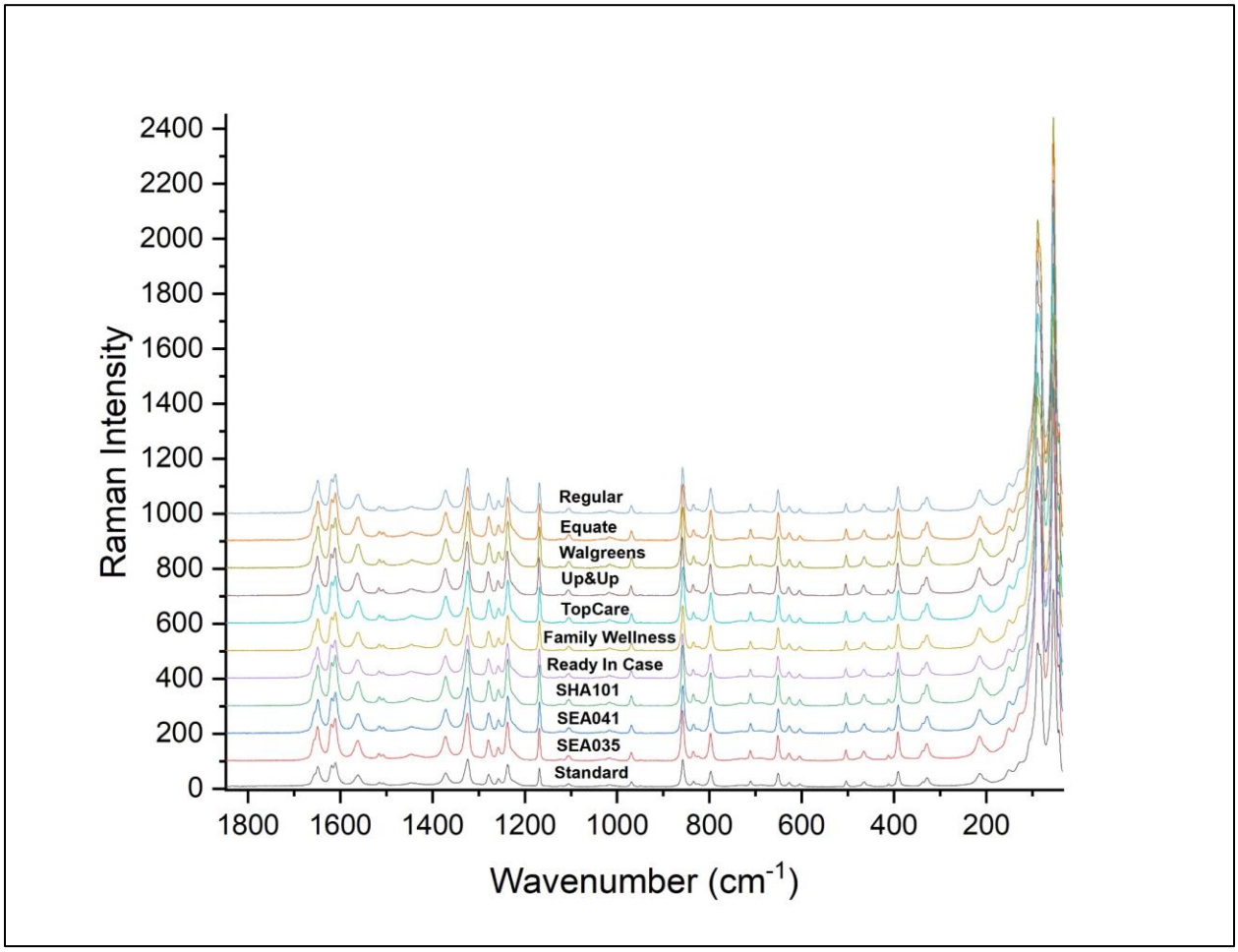



**Figure 9.** Representative Raman spectra for the interior surface of acetaminophen tablets from all acquired brands.

The variance within the interior tablets was assessed in the same way as the intact tablet samples, however, in this case the entire spectral range (35 - 1850 $cm^{-1}$) was examined. The results of the variance analysis are presented in Figure 10 and arranged in order of decreasing variance. There are some similarities in the trend between the variances of the Raman spectra of tablets with and without the coating layer, for example the Walgreens samples have the highest variance between intact tablets and the tablet interiors. However, there are also notable

differences in the rank order of the brand between the tablet interior scans and the intact tablets. The Tylenol brands are no longer grouped with similar variance, exhibit higher variance, and appear more inconsistent between lots in the interior than on the surface. Family Wellness tablets also appear to have much higher variance in their interiors, while Equate tablets have substantially lower variance. Overall, based on the variance results, the interior tablet scans should be more spread within the PCA space than the intact scans as there is a higher degree of variance between each tablet. This will be represented by distinct scores along the principal components.

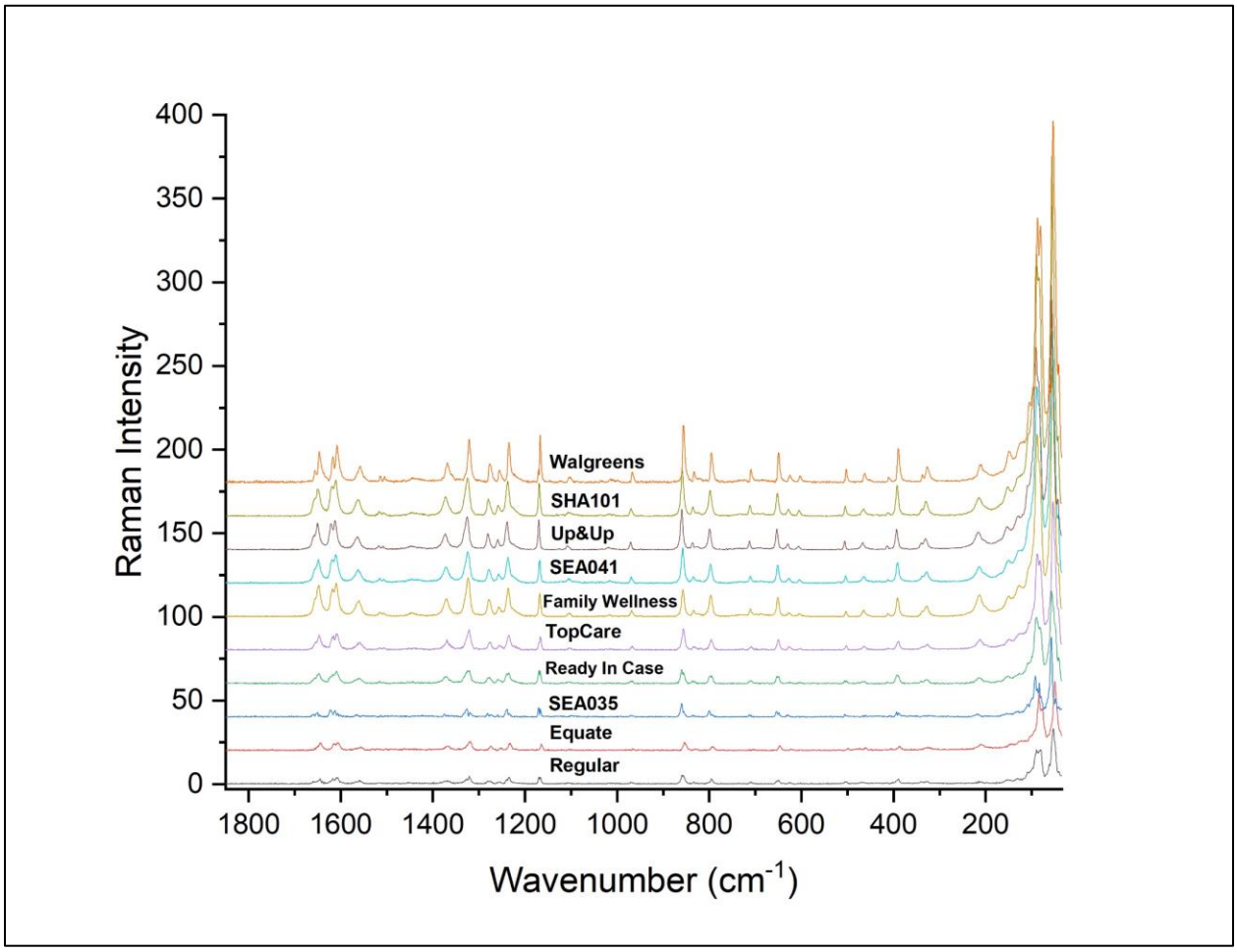



**Figure 10**. Variance of the Raman spectra of the interior surface of three tablets for each acetaminophen brand, arranged in order of decreasing variance.

The spread of tablets throughout the PCA scores plots is important for the model as the wider the range of scores for a certain brand is, the more likely it is to overlap with other brands scores along one or more PCs. This overlap, even slight, would limit the predictive power of the model; therefore, reducing the variance observed between tablets would be beneficial. The variance between tablets may be due to manufacturing, a proposition supported by the consistent low variance of the intact tablet scans of the Tylenol brand tablets, however the increased variance observed in the interior tablet spectra may be due to inconsistent tablet preparation for Raman measurements.

Variance between tablets could be introduced by the tablet preparation in several ways. Firstly, since the tablets were hand-sanded to remove the coating layers, the tablet surfaces were likely not of a consistent surface texture. While the surfaces were smoothed and focused in the instrument, the possible inconsistencies of texture may have contributed to overall variance between the tablets. Second, again due to the sanding being done by hand, each tablet may have been sanded to a different depth. If the tablet composition was not fully homogenous throughout the depth of the tablet, the interior surface scanned would be different for each tablet, contributing to increased variance. These impacts could be tested by mapping a cross section of the tablet to determine if there was inhomogeneous composition, and mitigated by ensuring that all tablets have consistent surface texture prior to analysis.

### 3.3 Intact Tablets: Full Spectral Range PCA

Data set #1 was compiled including all brands of intact tablet spectra and their full spectral range (35 - 1850 $cm^{-1}$) for the PCA analysis. A total of three PCs were selected to model 99.81 % variance within the data set based on the eigenvalue plot of this projection, as presented in Figure11. The full loadings and scores plots are provided in Appendix 3. The loading of PC 1 for this model (Figure 12) highly resembles the Raman spectra of $TiO_2$, with four main peaks

being highly loaded and very weak loadings of the remainder of the range. This indicates that the first PC is being overwhelmed by the unique variance being introduced by the $TiO_2$ coating present on the RIC and UU tablets. The loading plot of PC 2, as shown in Figure 13, resembles the acetaminophen spectrum, signifying that PC 2 is the principal component that is capturing the variance of the acetaminophen itself. This large source of known variance introduced by the presence of $TiO_2$ serves to overwhelm the first PC, and reduces the overall unknown variance that can be captured by the model. Only along PC 3, shown in Figure 14, did we observe unknown variance not reflecting the known components of the tablet samples.

Solely based on the loading plots it is expected that the RIC and UU tablets will show unique scores, and these observations are confirmed in examining the scores plot of PC 2 vs. PC 1 (Figure 15) in which the RIC and UU tablets are widely separated from the remaining brands along PC 1, and those remaining brands all score very closely on PC 1. The presence of the $TiO_2$ peaks leads to the high scores of RIC and UU tablets on PC 1, while the lack of these peaks in the remaining samples contributes to them scoring very consistently on PC 1. This overloading of PC 1 with the $TiO_2$ variance limits the ability of this model to capture small sources of variance within the dataset. However, some more unique groupings can be observed in Figure 16, which shows the scores plot of PC 3 vs. PC 2. This plot indicates that there is non-random grouping of brands together, and that tablets of like brands do more closely resemble their own brand than others. The most discrete groups were observed in the two of the extra strength Tylenol brand lots, SEA041 and SHA101, these showed unique scores on both PCs and were closely grouped. The Family Wellness and third extra strength Tylenol lot, SEA035, were not overlapping with other brands, however they showed a wide-spread between tablets, primarily along PC 2. The remaining brands all showed some degree of overlap with others.

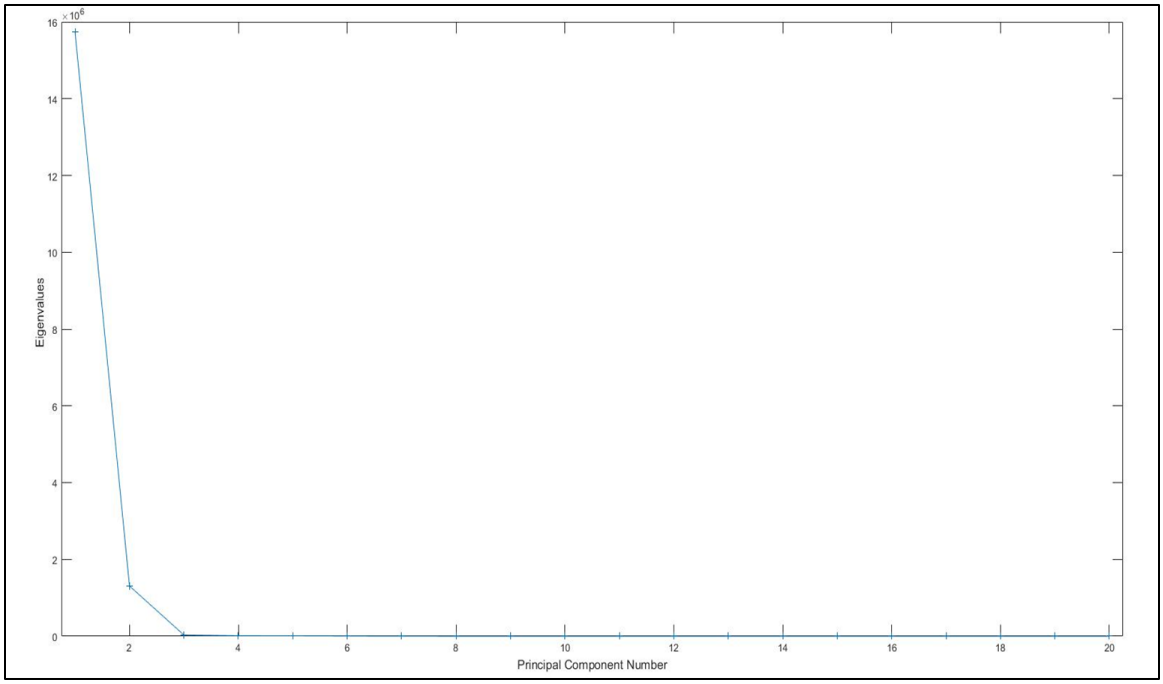



**Figure 11.** Eigenvalue plot of the PCA projection of the intact Raman spectra for all brands.

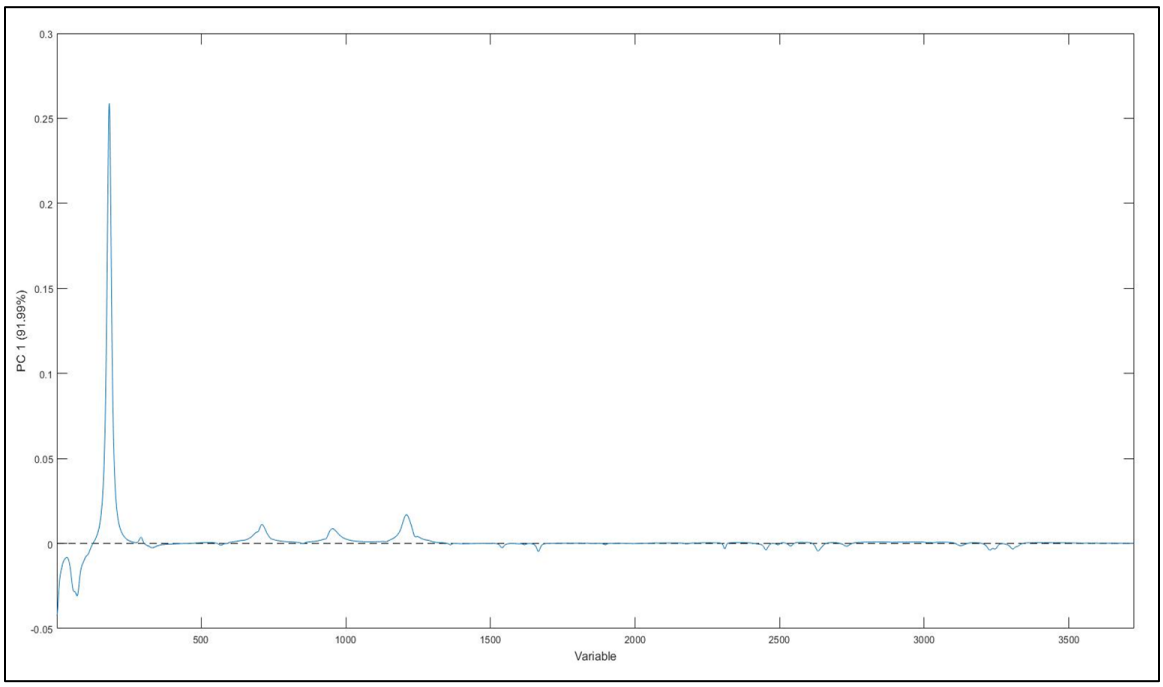



**Figure 12.** Loading plot of PC 1 of the PCA projection of the intact Raman spectra for all brands.

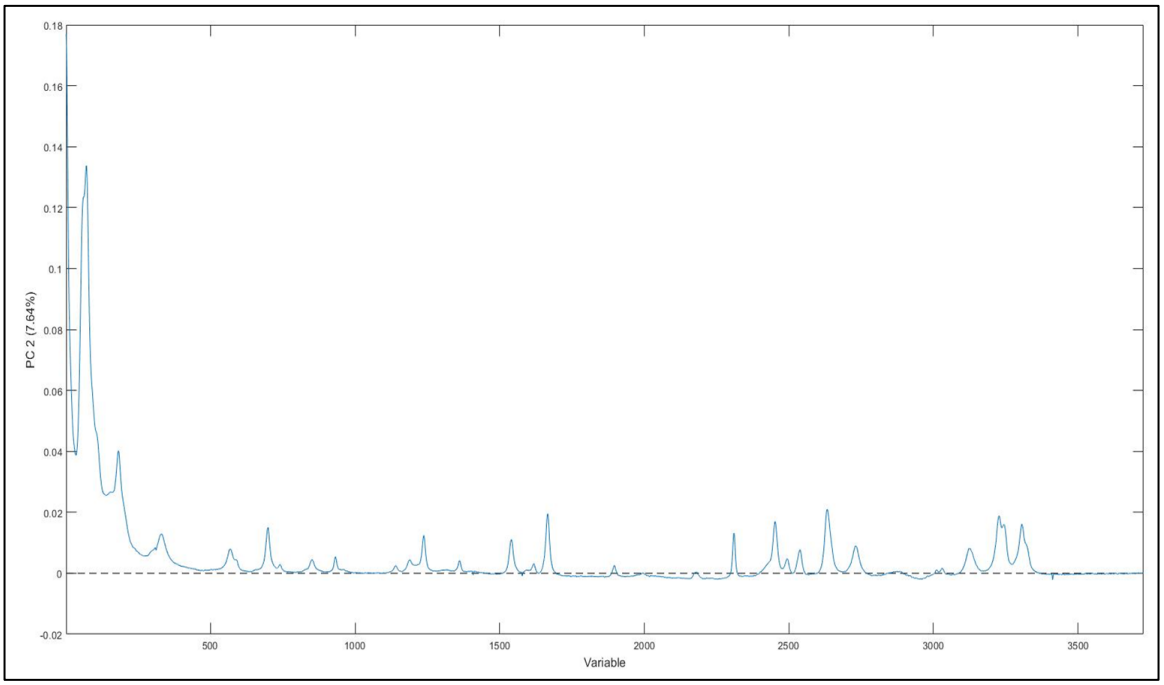



**Figure 13.** Loading plot of PC 2 of the PCA projection of the intact Raman spectra for all brands.

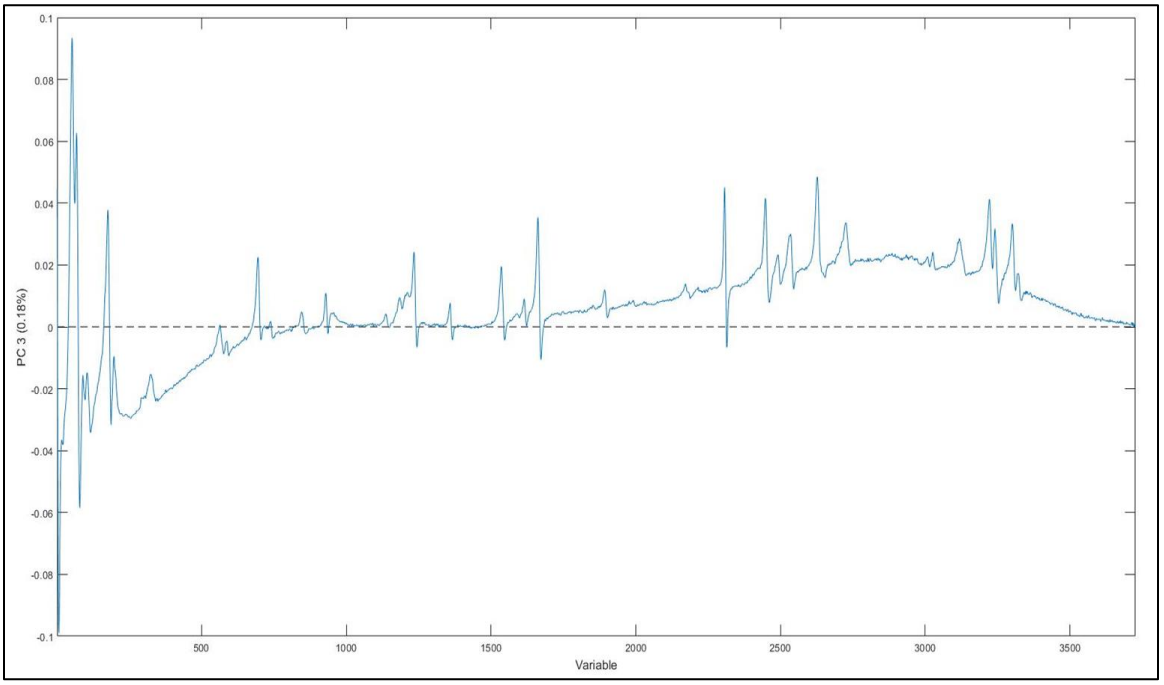



**Figure 14.** Loading plot of PC 3 of the PCA projection of the intact Raman spectra for all brands.

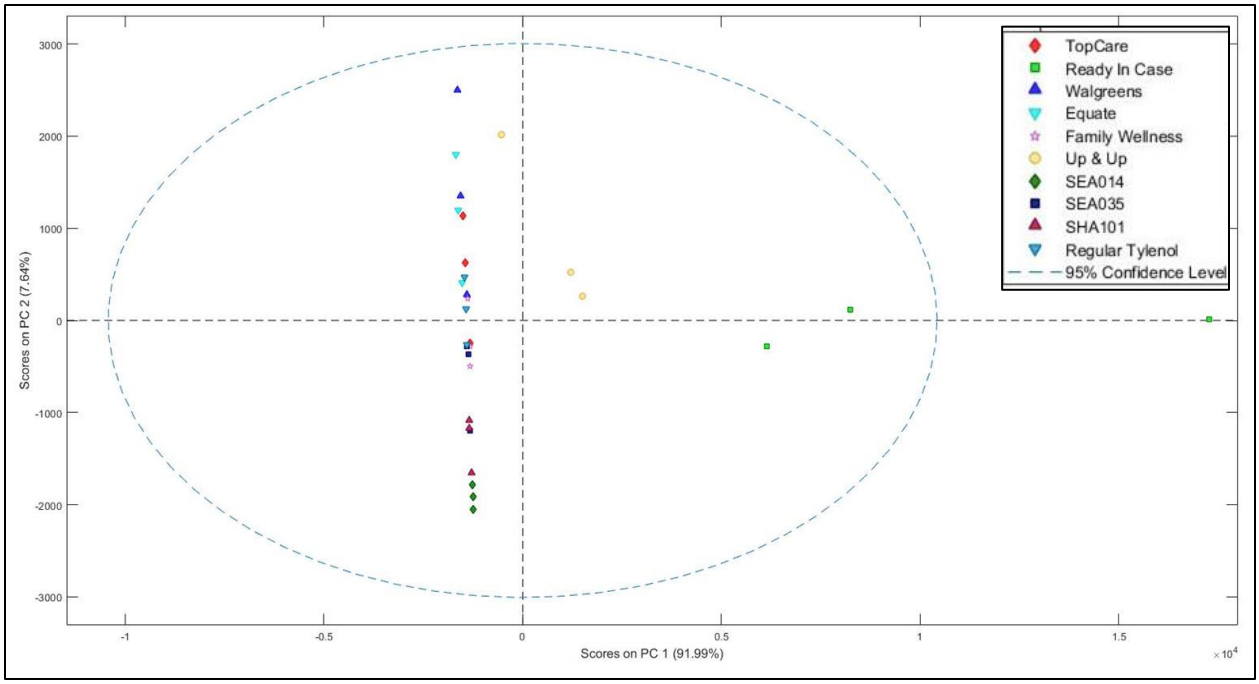



**Figure 15**. Scores plot of PC 2 vs. PC 1 of the PCA projection of the intact Raman spectra for all brands.

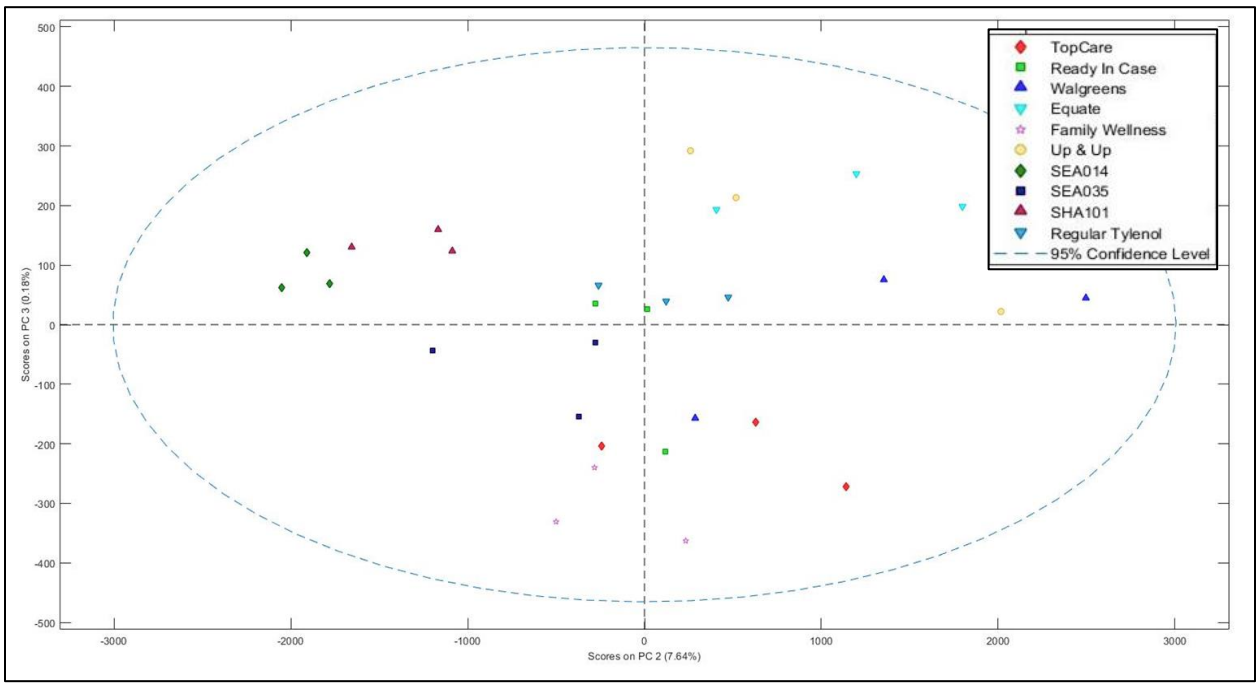



**Figure 16.** Scores plot of PC 3 vs. PC 2 of the PCA projection of the intact Raman spectra for all brands.

### 3.4 Intact Tablets: Full Spectral Range PCA with the RIC and UU Brands Excluded

In order to remove the convoluting variance of the $TiO_2$ in the Raman spectra, data set #2 was compiled by removing the two brands in which the compound was observed from the intact tablet PCA data set. The resultant eigenvalues for the PC projection are shown in Figure 17, and based on this figure, three PCs were chosen for the model and 99.61% variance was captured. The loading plot of PC 1 (Figure 18) was noted to resemble the loading plot of PC 2 of the PCA model of data set #1 (see Figure 15), while PC 2 (Figure 19) loading resembled that of PC 3 of the same model (see Figure 16). The loading of the current model's PC 1 resembles the acetaminophen spectra, while PC 2 and PC 3 (Figures 19 and 20) captured unknown variance within the data set.

The resemblance of the loading plots of PC 1 and PC 2 to the previous model also led to the score projection of these two PCs, as shown in Figure 21, to resemble that of the corresponding projection of PC 2 and PC 3 of the previous model (see Figure 16). Due to these similarities, the score plot of PC 1 and PC 2 do not provide any new information about grouping the tablets, but when projected with PC 2 and PC 3, we are able to observe new discrete groupings of the tablet brands, as seen in Figure 22. This plot has clear and discrete groupings of three generic brands, Family Wellness, Equate, and Top Care, as well as discrete groupings of all three lots of extra strength Tylenol together near the center of the plot. The only significant overlap observed is the regular strength Tylenol and the Walgreens brand. However, when including PC 1 in the analysis we can see clear differentiation in these brands, as seen in the 3D projection of scores in Figure 23. These results indicate that with a more robust training set of tablets, it is likely we could reliably differentiate between each of the distinct tablet brands and lots.

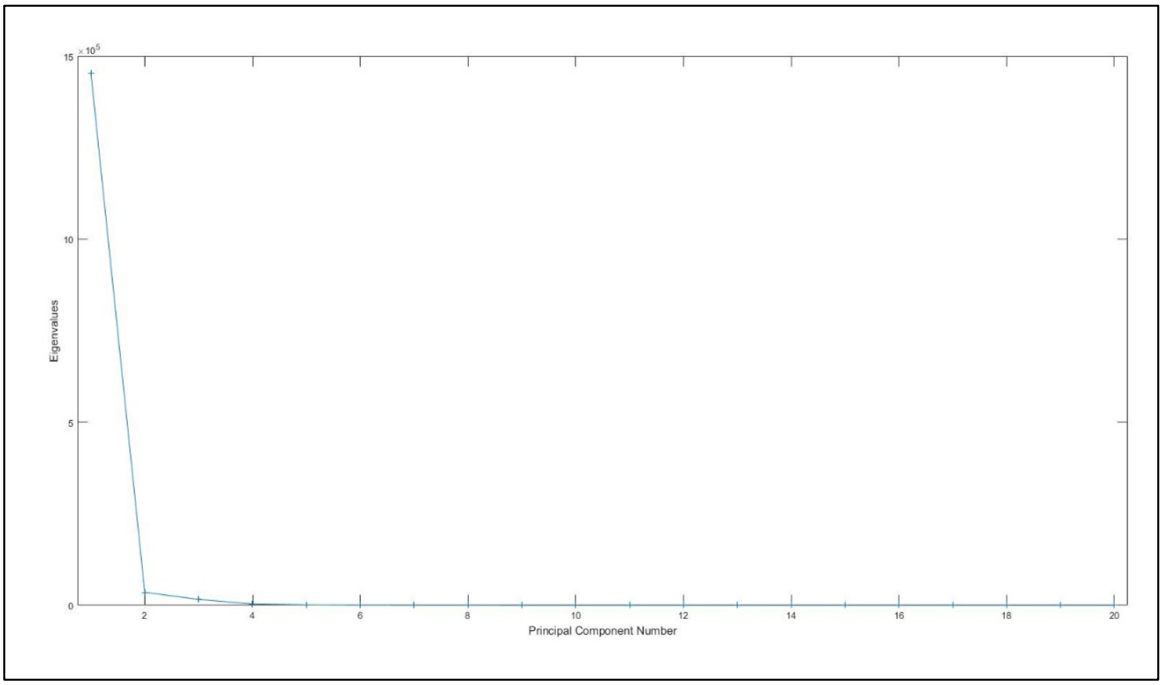



**Figure 17.** Eigenvalue plot of the PCA projection of the intact Raman spectra for brands excluding RIC and UU.

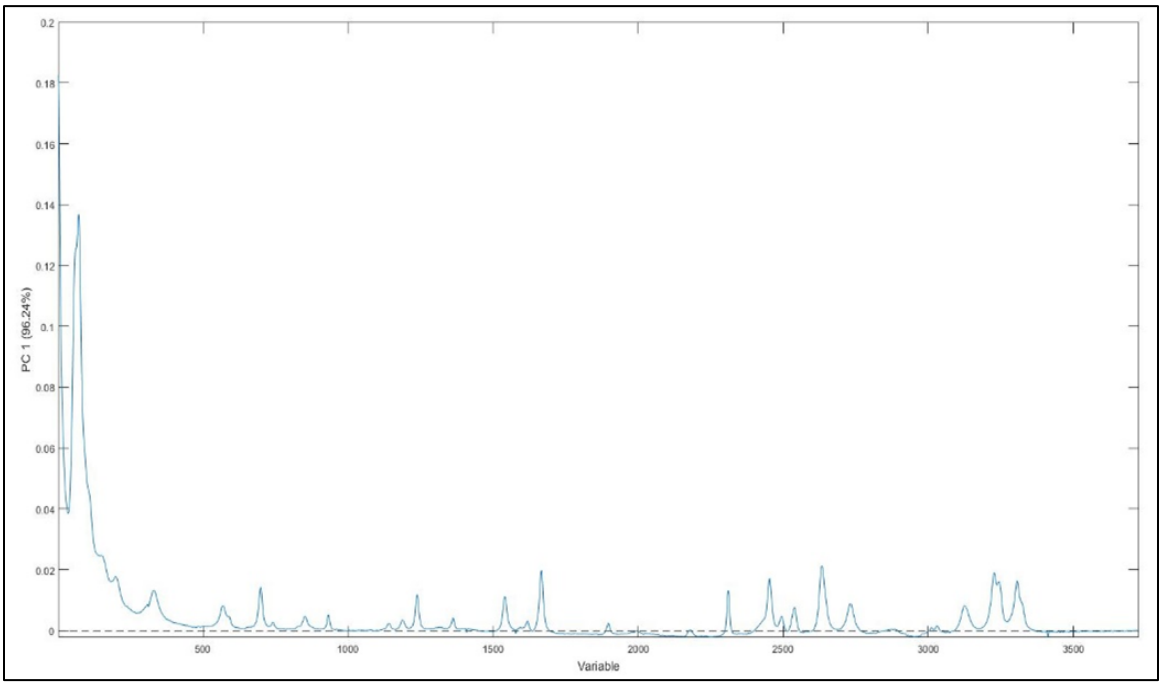



**Figure 18.** Loading plot of PC 1 of the intact Raman spectra for brands excluding RIC and UU.

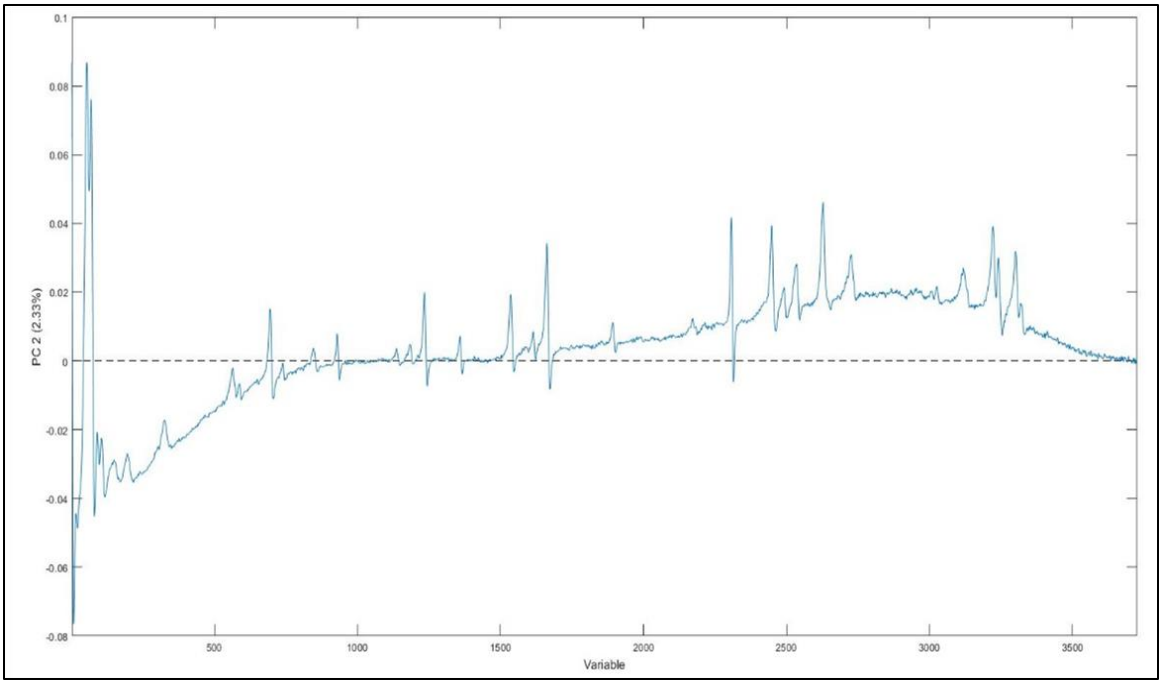



**Figure 19.** Loading of PC 2 of the intact Raman spectra for brands excluding RIC and UU.

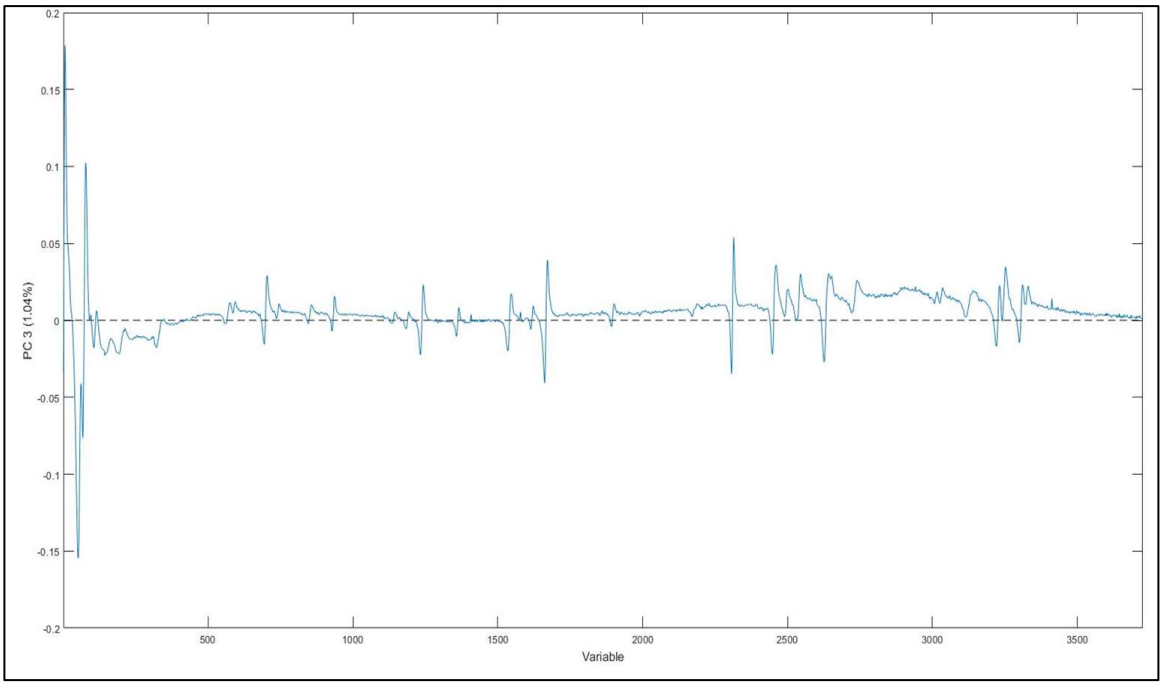



**Figure 20.** Loading of PC 3 of the intact Raman spectra for brands excluding the RIC and UU.

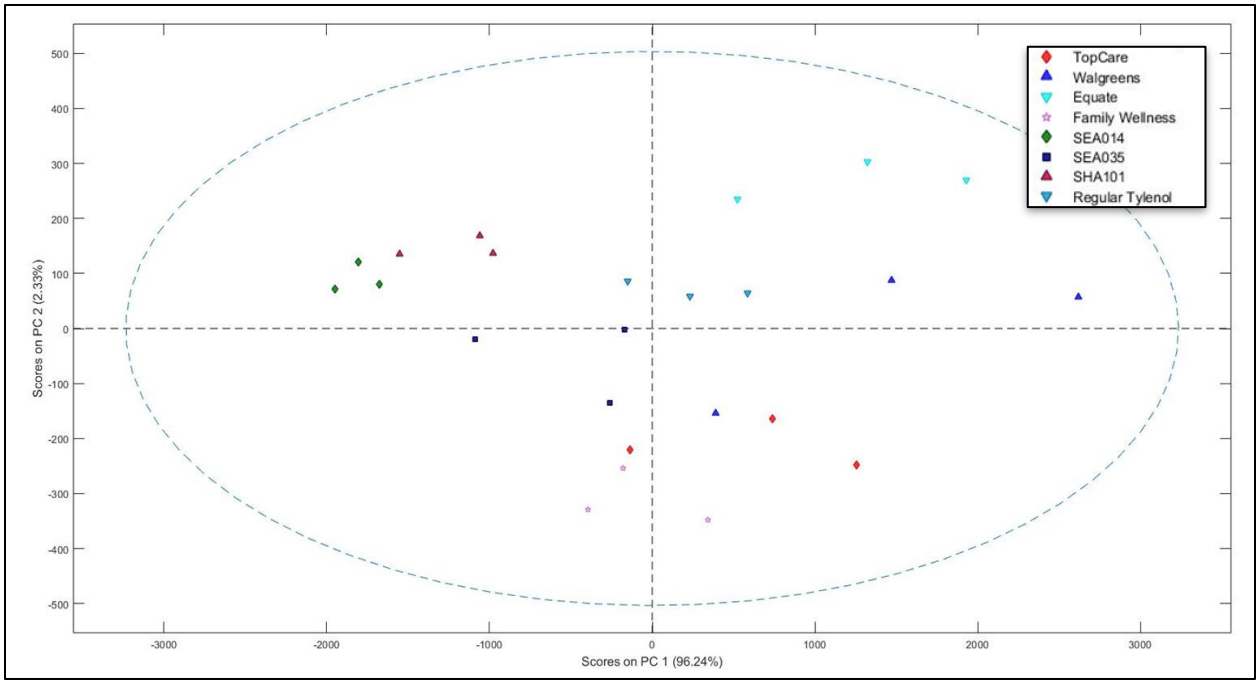



**Figure 21.** Scores plot of PC 2 vs. PC 1 of the intact Raman spectra for brands excluding RIC and UU.

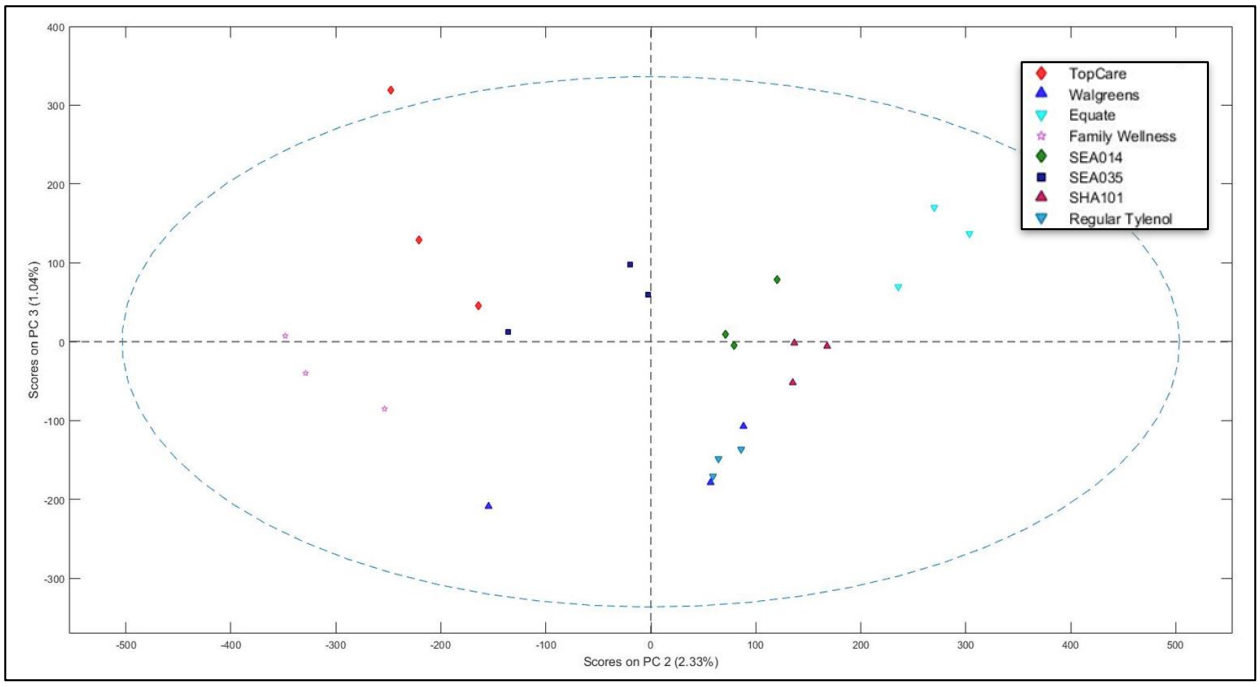



**Figure 22.** Scores plot of PC 3 vs. PC 2 of the intact Raman spectra for brands excluding RIC and UU.

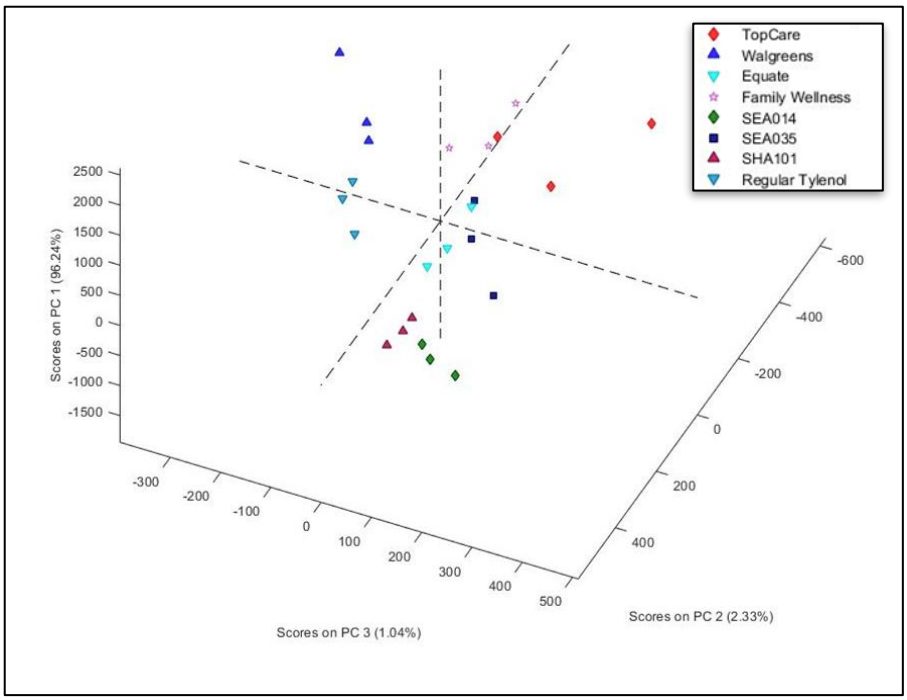



**Figure 23.** Three dimensional scores plot of PCs 1, 2, and 3 of the intact Raman spectra for brands excluding RIC and UU.

### 3.5 Intact Tablets: Upper Spectral Range PCA

The limitation of the previous model is that the brands in which $TiO_2$ was observed had to be removed prior to the calculation in order to achieve the results. In order to account for all of the brands in a single model while excluding any influence from $TiO_2$, data set #3, the upper spectral range (700 - 1850 $cm^{-1}$) of the intact tablets of all brands, was used for the PCA projection. The eigenvalues for this projection (Figure 24) indicated the choice of three PCs in order to capture a total of 98.66% of the variance within the data set.

The loading plot of PC 1 (Figure 25) resembles the upper range of the spectra, capturing the known variance within the sample, and the remaining PCs captured small sources of unknown variance. The 3D plots of the scores (Figures 26 and 27) show that the majority of the brands are grouped similarly to the model of data set #2, but with the RIC and UU brands are overlapping with each other as well as Equate, Walgreens, and Tylenol lot SEA035 tablets. This

overlap is in part due to the tablets being widely spread across the PC space, specifically along the PC 1. The source of this separation between tablets can be observed by examining the spectra of each of the tablets in which $TiO_2$ was observed (Figure 5) and noting the inverse relationship between the Raman intensity of $TiO_2$ peaks and the acetaminophen peaks. As PC 1 of this model is reflecting the known variance of the acetaminophen, the scores of the tablets along this dimension are dependent on the intensity of the acetaminophen peaks in the spectrum, so the variation in peak intensity between tablets created by the inhomogeneity of the $TiO_2$ coating thickness directly impacts the PC 1 scores. These variations in intensity are still contributing to known variance caused by the presence of $TiO_2$, although the characteristic peaks of this compound are not included. The 3D plots of the scores (Figures 26 and 27) indicate the distinct groupings of Family Wellness, Top Care, regular strength Tylenol, and extra strength Tylenol lots SEA041 and SHA101. The remaining brands showed overlapping scores, thus limiting the predictive power of this model. These results indicate the removal of the $TiO_2$ peaks by limiting the spectra range was not sufficient to remove interference of the coating layer of the RIC and UU tablets.

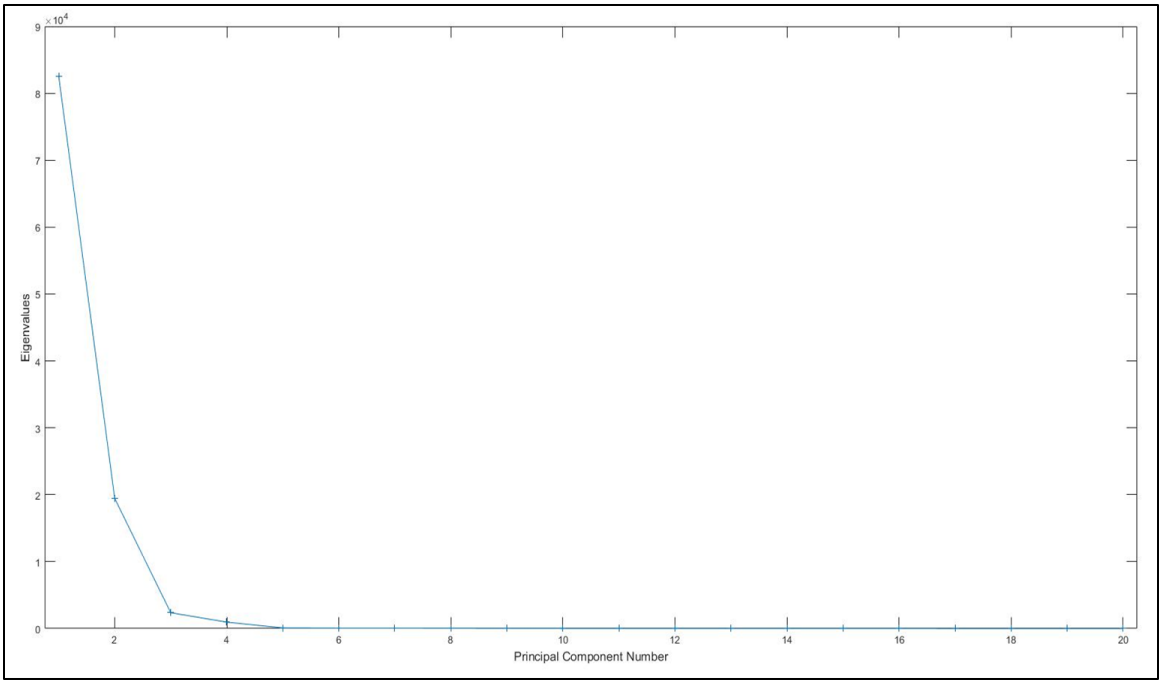



**Figure 24.** Eigenvalue plot of the PCA projection of the upper range (700 - 1850 $cm^{-1}$) Raman spectra of intact tablets for all brands.

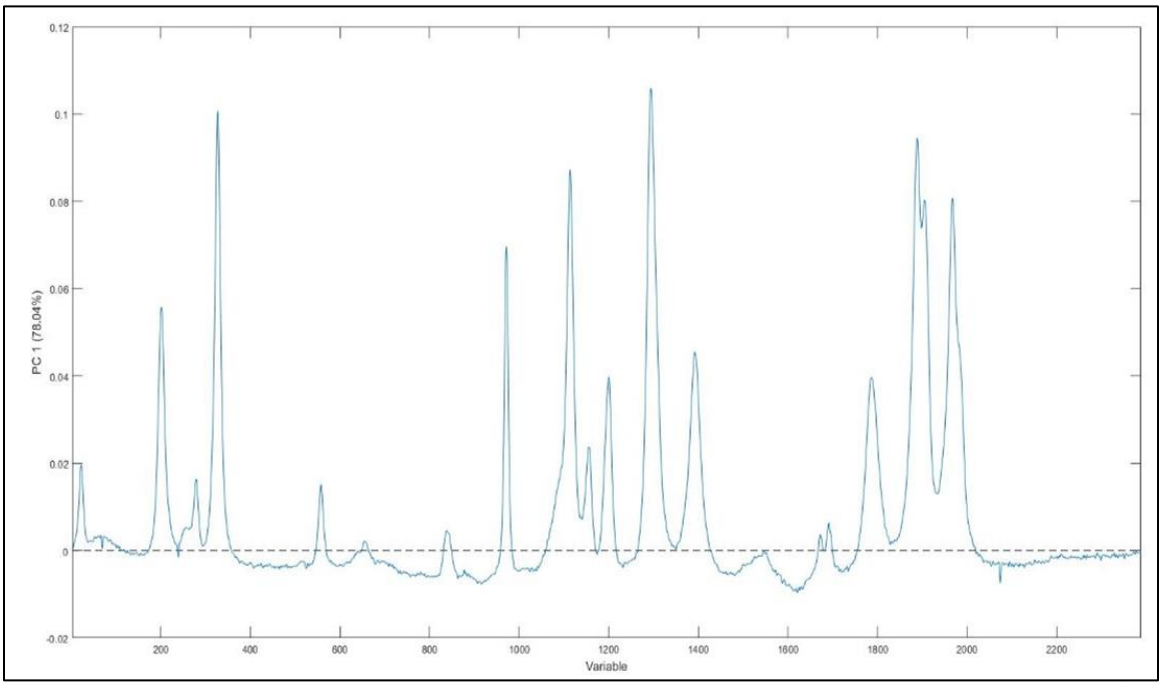



**Figure 25.** Loading plot of PC 1 of the PCA projection of the upper range (700 - 1850 $cm^{-1}$) Raman spectra of intact tablets for all brands.

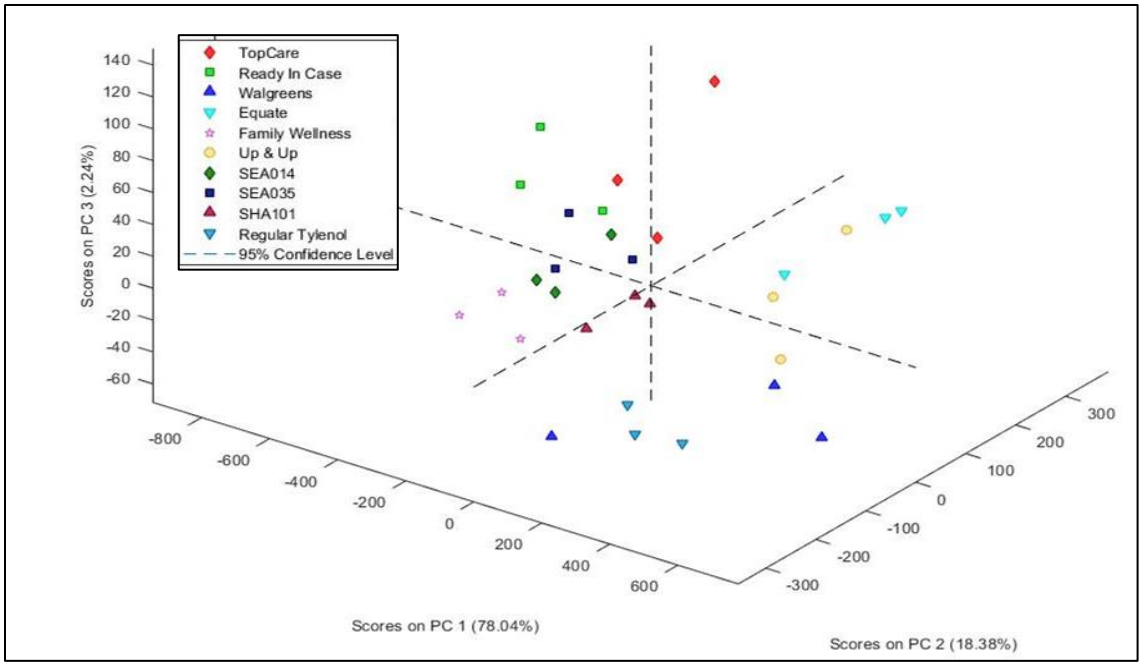



**Figure 26.** Three dimensional scores plot of PCs 1, 2, and 3 of the upper range (700 - 1850 $cm^{-1}$) Raman spectra of intact tablets for all brands - View 1.

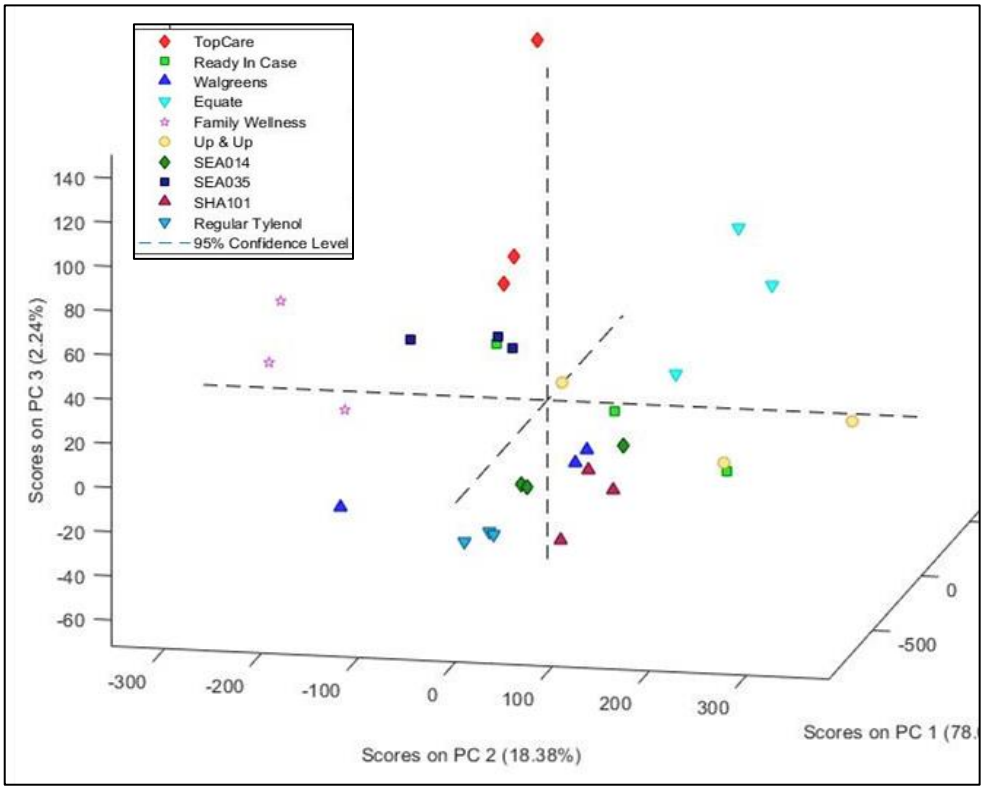



**Figure 27.** Three dimensional scores plot of PCs 1, 2, and 3 of the upper range (700 - 1850 $cm^{-1}$) Raman spectra of intact tablets for all brands - View 2.

### 3.6 Interior Tablets: Full Spectral Range PCA

The PCA projection for the Raman spectra of the tablet interiors (data set #4) was generated in the hope that by using the interior scans, in which no $TiO_2$ top layer was present, all brands could be discretely grouped and differentiated. The eigenvalues for the model are presented in Figure 28 and a total of three PCs were chosen in order to capture 99.38% of the total variance. The PC 1 loading (Figure 29) resembles the acetaminophen spectra as expected, while the remaining two PCs show unknown variance being captured (Figures 30 and 31). PC 3 loading, as apparent in Figure 31, shows strong loadings in the low wavenumber range (35-150 $cm^{-1}$).

The score plot of PC 2 vs. PC 1 (Figure 32) shows a non-random distribution of tablets, but no definable groups of any brands. All brands are overlapping along these two dimensions. Scores plots including PC 3 (Figures 33-34) show a unique score of the Equate tablets on PC 3. All three Equate tablets all show strong negative correlations along PC 3, creating a very distinct group. The Equate tablet spectra (Figure 35) showed no apparent unique signal which should produce these uncharacteristic scores within the lower wavenumber range and there was no unique Raman peak shifts or peak intensities observed. However, all other brands score very similarly along the model's PC 3, indicating that there is a unique variance within the Equate tablet not present in the remaining brands. The close scores of all other brands along PC 3 also creates significant overlap between the brands on this dimension.

This model has limited power for discrimination between brands as the delimitations of scores between tablets are more poorly defined. This is very likely a direct result of the increase in variance between spectra of the tablet interiors as compared to the intact tablet spectra. In the plot of PC 2 vs. PC 1 (Figure 32) there are few brands in which all tablets are grouped closely, namely Equate and regular Tylenol, while the UU, Walgreens, SHA101, SEA041, and FW

tablets are all widely spread across the PCs. These trends are consistent with the variance observed between tablets in Figure 10, as Equate and regular Tylenol show the lowest variance and the brands with the largest spread show the most. These brands showed closer groupings along PC 3, however this is in part an artificial effect of to the strong negative score of the Equate tablets along this dimension.

Therefore, one method of potentially improving the model is by isolating the chemical source of the unique PC 3 scores of the Equate tablets, which was not accomplished in the current work. Based on the PC 3 loading plot for the model, it is clear that the unique variance occurs within the low range of wavenumbers, as the loadings are very strongly positive for 35 - 100 $cm^{-1}$. However, there are no clear distinction of the Raman spectra of Equate tablets in this range as compared to any of the other brands tested, therefore the precise cause of the variance is unknown. The difficulty in directly identifying the unique features is understandable as the third principal component is only capturing 0.74 % of the total variance in the model, therefore it follows that the specific features of the graph being represented may not be readily apparent. These small variations between brands would not be observable by direct comparison of the spectral data, but the PCA is able to model them as significant and allow for reliable distinction between the Equate tablets and other brands of acetaminophen tested.

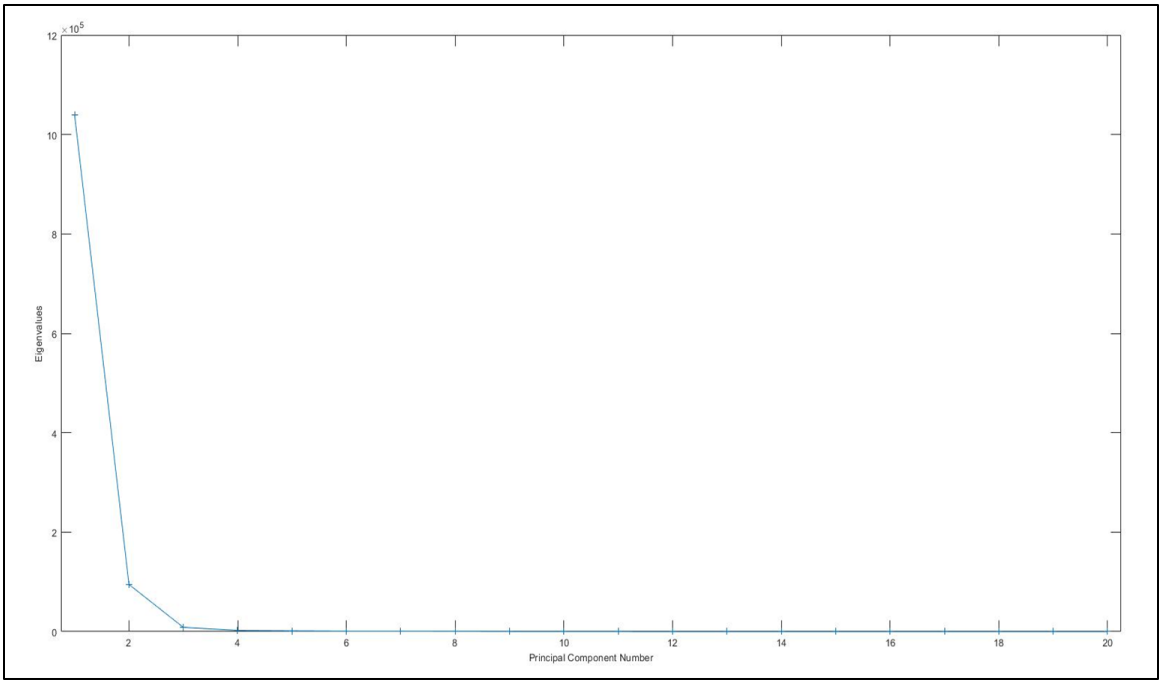



**Figure 28.** Eigenvalue plot of the PCA projection of the Raman spectra of the tablet interiors for all brands.

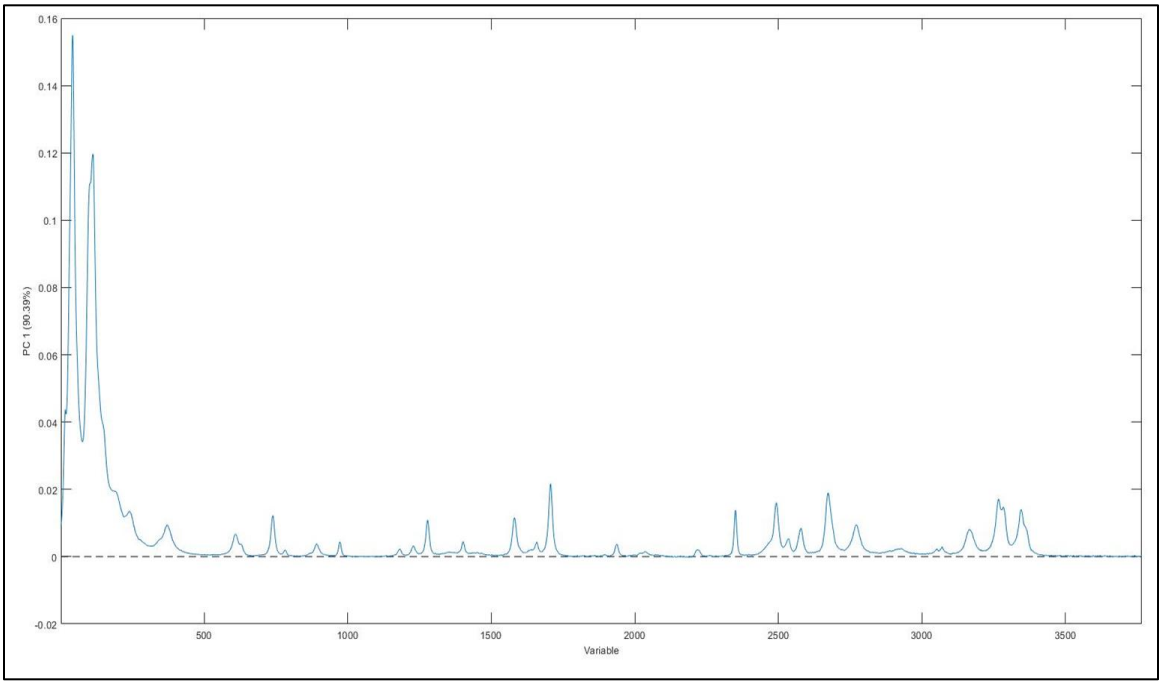



**Figure 29.** Loading plot of PC 1 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

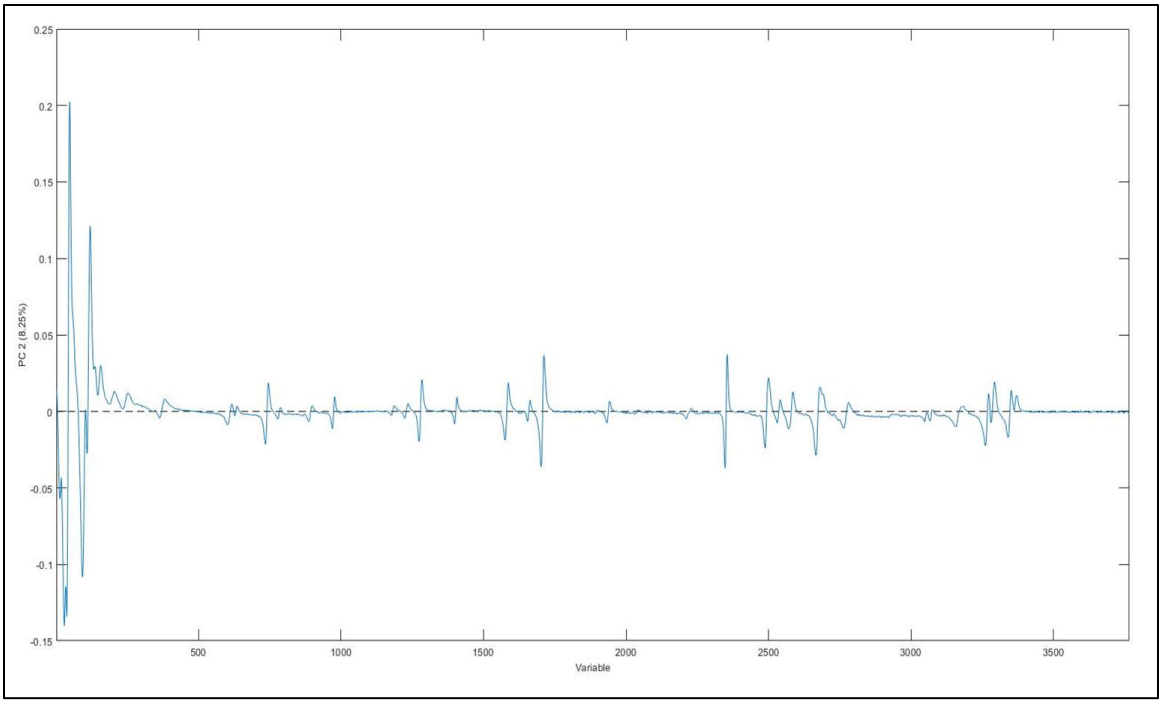



**Figure 30.** Loading plot of PC 2 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

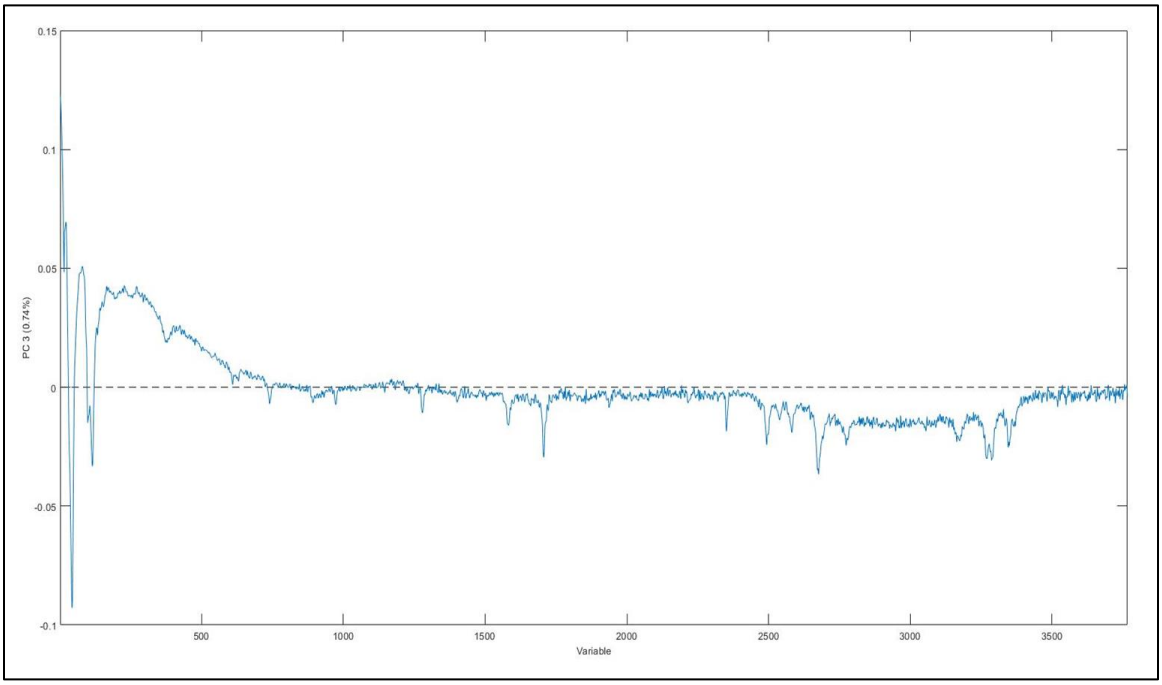



**Figure 31.** Loading plot of PC 3 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

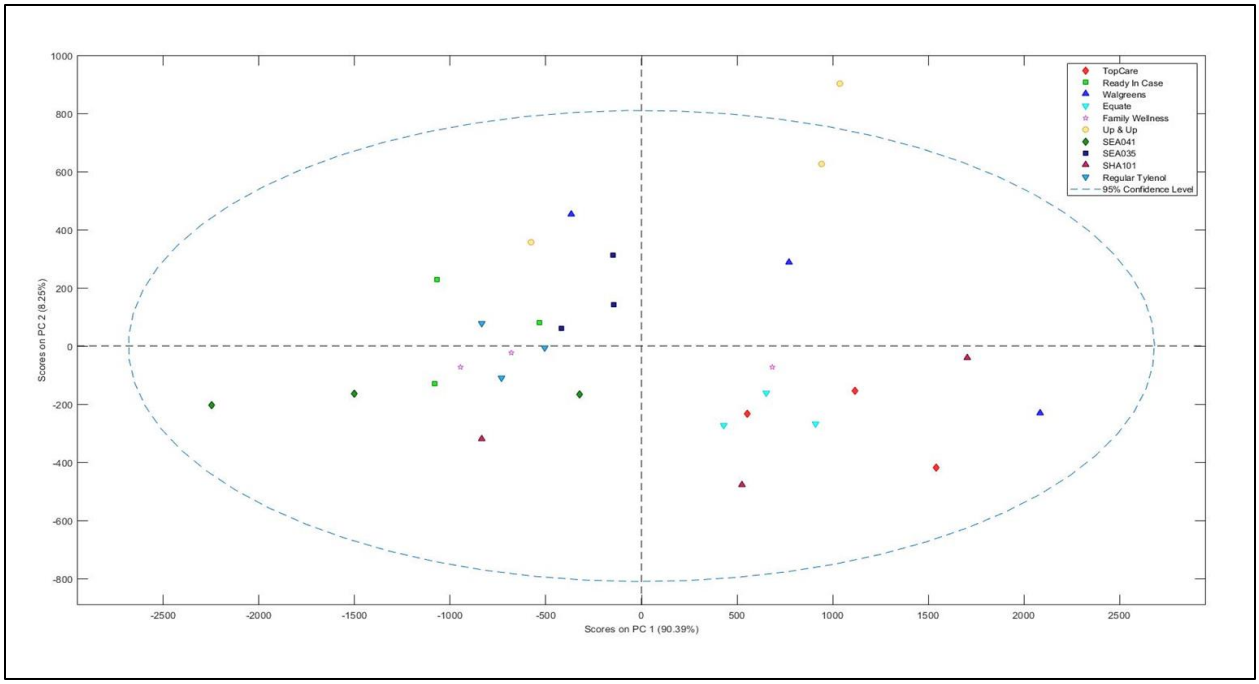



**Figure 32.** Scores plot of PC 2 vs. PC 1 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

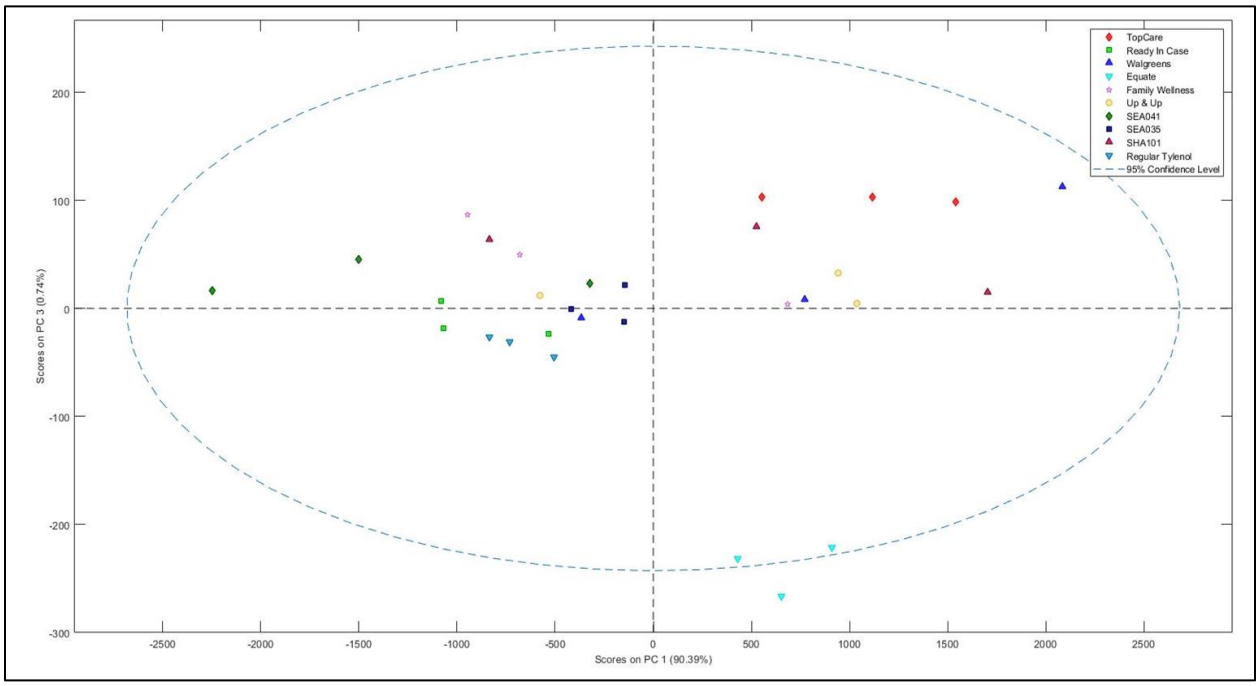



**Figure 33.** Scores plot of PC 3 vs. PC 1 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

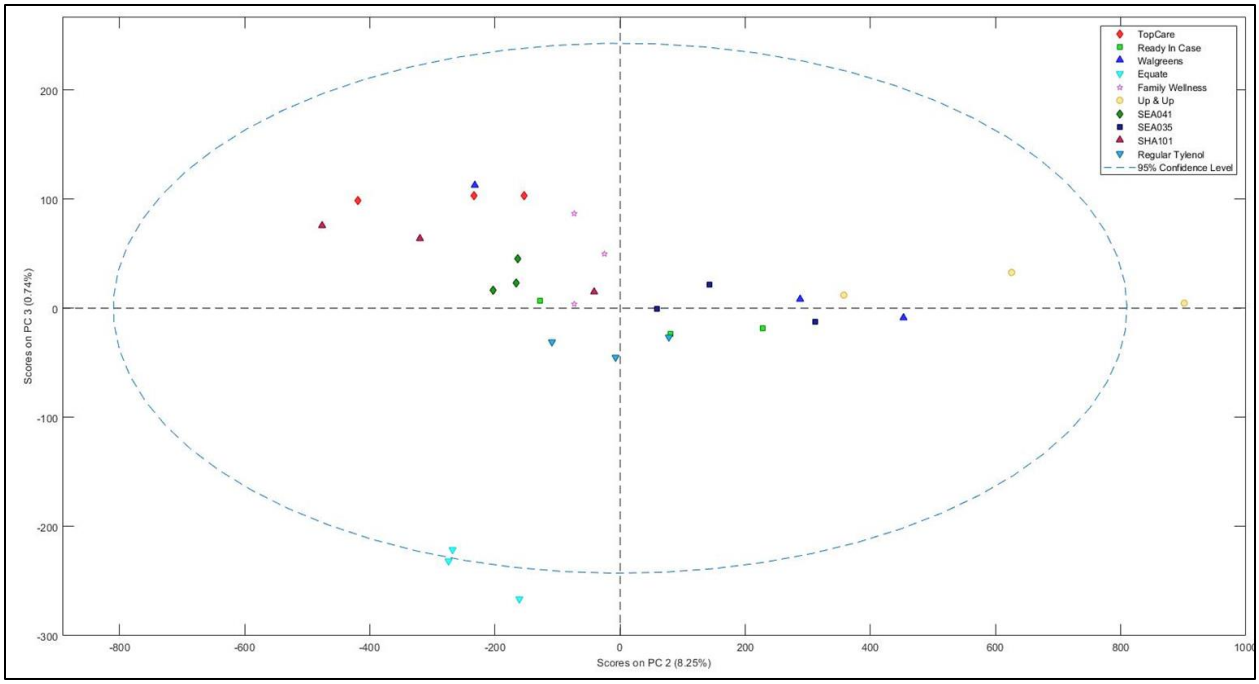



**Figure 34.** Scores plot of PC 3 vs. PC 2 of the PCA projection of the Raman spectra of the tablet interiors for all brands.

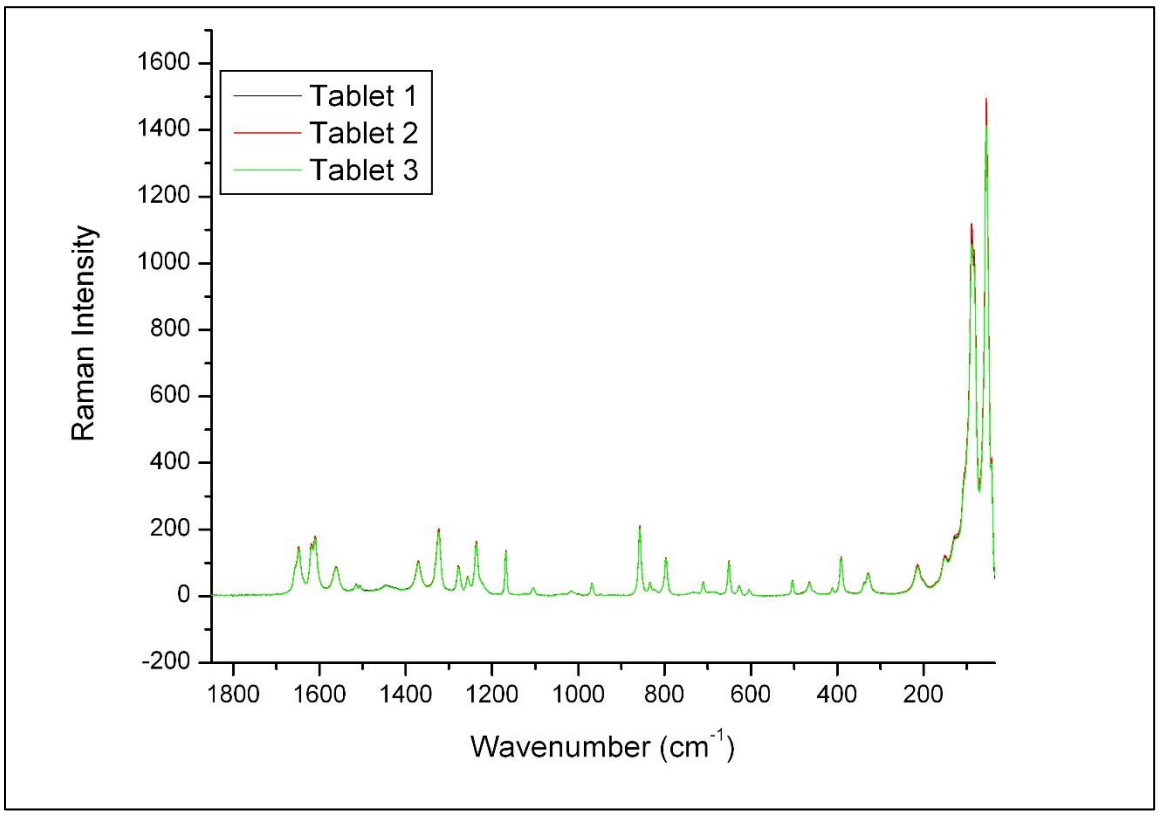



**Figure 35.** Average Raman spectra for each Equate tablet scanned.

## V. Conclusions and Future Directions

This work represented the successful proof of concept of the use of Raman scattering spectroscopy with PCA chemometric projection for the differentiation of various brands and lots of genuine acetaminophen. A total of ten acetaminophen brand samples were tested using the method which was developed for the Raman scanning of the surface of intact acetaminophen tablets and the interior surface of the same tablets. The gathered Raman spectra did provide some useful information about the tablet brands, most notably the Raman scans of the intact tablets revealed the presence of a $TiO_2$ layer on the surface of RIC and UU brand tablets, easily differentiating them from each other. Additionally, the peak intensity of the $TiO_2$ varied significantly between the two brands, allowing them to also be differentiated from each other. The remaining tablet spectra were highly similar, making it difficult to differentiate the remaining brands, however the use of PCA of the spectra allowed the slight sources of variance within the spectra to be clearly modeled.

Revisiting the stated goals of the study, the PCA projections were, as a whole, able to achieve the desired results. Most promising, and revealing the most significant information, was the PCA model of the intact tablets, excluding the RIC and UU brands. This model, by excluding the known variance of $TiO_2$ within the spectra, resulted in discrete groups of each of the acetaminophen tablets by brand, and differentiated manufacturing lots of extra strength Tylenol and the two strengths (regular and extra strengths) of Tylenol brand tablets. These groups were not tested for their predictive power; however, the use of a similar model for prediction should be a main goal for future work on this project.

Any future studies performed on the basis of these results should also be adjusted to address some key limitations of this work. A main limitation is the number of brands tested and the lack of replicate lots of each of the generic brands. The results presented represent only

singular lots of the six generic brands, and therefore cannot be generalized to the product as a whole. Additionally, the method for Raman scanning should be adjusted in order to ensure non-biased results. The current results were gathered by scanning all tablets of the same brand one after another, which may have created bias within scans and contributed to the grouping of tablets within the PCA. This possible bias could be minimized by creating a random order of tablet scanning so that all brands are not grouped, or by including an internal standard sample of acetaminophen in the sample set for normalization.

Once a more accurate and representative data set has been acquired, the use of chemometrics should also be expanded by using additional projections and predictive tools in order to thoroughly test the power of the differentiation. The addition of hierarchical cluster analysis, boosted tree models, or multilinear regressions may all provide additional significant data and more clearly represent the variance for interpretation and prediction of tablet identity.

Finally, models similar to those presented here may have potential to application to other pharmaceuticals and products. These applications could assist in consumer awareness, forensic applications, and product tracing. An overall increase in transparency surrounding pharmaceutical products is important as it allows the public to be more mindful and aware of their consumption, and manufacturers to hold themselves to a higher standard of consistency and safety.

## VI. References


1. Been, F.; Roggo, Y.; Degardin, K.; Esseiva, P.; Margot, P., Profiling of counterfeit medicines by vibrational spectroscopy. *Forensic Sci. Int.* **2011,** *211*, 83-100.
2. Neuberger, S.; Neusüß, C., Determination of counterfeit medicines by Raman spectroscopy: Systematic study based on a large set of model tablets. *J. Pharm. Biomed. Anal.* **2015,** *112*, 70-78.
3. Conti, R. M.; Berndt, E. R.; Kaygisiz, N. B.; Shivdasani, Y., We Still Don't Know Who Makes This Drug. Health Affairs, 2020.
4. Conti, R. M.; Berndt, E. R., Who Makes This Drug? Disclosing The Identity Of Generic Drug Manufacturers. Health Affairs: 2015.
5. Storme-Paris, I.; Rebiere, H.; Matoga, M.; Civade, C.; Bonnet, P. A.; Tissier, M. H.; Chaminade, P., Challenging Near InfraRed Spectroscopy discriminating ability for counterfeit pharmaceuticals detection. *Anal. Chim. Acta* **2010,** *658*, 163-174.
6. Ali, H.; Ullah, R.; Khan, S.; Bilal, M., Raman spectroscopy and hierarchical cluster analysis for the ingredients characterization in different formulations of paracetamol and counterfeit paracetamol. *Vib. Spectrosc.* **2019,** *102*, 112-115.
7. Kwok, K.; Taylor, L. S., Analysis of counterfeit Cialis® tablets using Raman microscopy and multivariate curve resolution. *J. Pharm. Biomed. Anal.* **2012,** *66*, 126-135.
8. Vajna, B.; Farkas, A.; Pataki, H.; Zsigmond, Z.; Igricz, T.; Marosi, G., Testing the performance of pure spectrum resolution from Raman hyperspectral images of differently manufactured pharmaceutical tablets. *Anal. Chim. Acta* **2012,** *712*, 45-55.
9. Khanmohammadi, M.; Garmarudi, A. B.; Moazzen, N.; Ghasemi, K., Qualitative Discrimination Between Paracetamol Tablets Made by Near Infrared Spectroscopy and Chemometrics With Regard to Polymorphism. *J. Struct. Chem.* **2010,** *51*, 663-669.
10. Wabuyele, B. W.; Sotthivirat, S.; Zhou, G. X.; Ash, J.; Dhareshwar, S. S., Dispersive Raman Spectroscopy for Quantifying Amorphous Drug Content in Intact Tablets. *J. Pharm. Sci.* **2017,** *106*, 579-588.
11. Beyer, T.; Day, G. M.; Price, S. L., The Prediction, Morphology, and Mechanical Properties of the Polymorphs of Paracetamol. *J. Am. Chem. Soc.* **2001,** *123*, 5086-5094.
12. Morse, H. N., Ueber eine neue Darstellungsmethode der Acetylamidophenole. *Chem. Ber.* **1878,** *11*, 232-233.
13. Smith, S. J.; Bishop, M. M.; Montgomery, J. M.; Hamilton, T. P.; Vohra, Y. K., Polymorphism in paracetamol: evidence of additional forms IV and V at high pressure. *J Phys Chem A* **2014,** *118*, 6068-77.
14. Katdare, A.; Chaubal, M., *Excipient development for pharmaceutical, biotechnology, and drug delivery systems*. CRC Press: 2006.
15. Esbensen, K. H.; Román-Ospino, A. D.; Sanchez, A.; Romañach, R. J., Adequacy and verifiability of pharmaceutical mixtures and dose units by variographic analysis (Theory of Sampling) – A call for a regulatory paradigm shift. *Int. J. Pharm.* **2016,** *499*, 156-174.
16. Granger, R. M.; Yochum, H. M.; Granger, J. N.; Sienerth, K. D., *Instrumental Analysis*. Oxford University Press: 2017.
17. Gao, Q.; Liu, Y.; Li, H.; Chen, H.; Chai, Y.; Lu, F., Comparison of several chemometric methods of libraries and classifiers for the analysis of expired drugs based on Raman spectra. *J. Pharm. Biomed. Anal.* **2014,** *94*, 58-64.

18. Griffen, J. A.; Owen, A. W.; Burley, J.; Taresco, V.; Matousek, P., Rapid quantification of low level polymorph content in a solid dose form using transmission Raman spectroscopy. *J. Pharm. Biomed. Anal.* **2016,** *128*, 35-45.
19. Cooper, J. B., Chemometric analysis of Raman spectroscopic data for process control applications. *Chemom. Intell. Lab. Syst.* **1999,** *46*, 231-247.
20. Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R., Chemometrics in analytical chemistry-part I: history, experimental design and data analysis tools. *Anal. Bioanal. Chem.* **2017,** *409*, 5891-5899.
21. Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R., Chemometrics in analytical chemistry-part II: modeling, validation, and applications. *Anal. Bioanal. Chem.* **2018,** *410*, 6691-6704.
22. Anderson, M. J., Success with DOE. *Quality* **2000,** *39*, 38-44.
23. Ralbovsky, N. M.; Lednev, I. K., Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer. *Spectrochim. Acta, Part A* **2019,** *219*, 463-487.
24. Yazgan Karacaglar, N. N.; Bulat, T.; Boyaci, I. H.; Topcu, A., Raman spectroscopy coupled with chemometric methods for the discrimination of foreign fats and oils in cream and yogurt. *J. Food Drug Anal.* **2019,** *27*, 101-110.
25. De Bleye, C.; Chavez, P. F.; Mantanus, J.; Marini, R.; Hubert, P.; Rozet, E.; Ziemons, E., Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications. *J. Pharm. Biomed. Anal.* **2012,** *69*, 125-132.
26. Jolliffe, I. T.; Cadima, J., Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **2016,** *374*, 20150202.
27. Shlens, J., A tutorial on principal component analysis. *arXiv:* 2014, *1404.1100*, ver. 1.
28. Bellia, G. F.; Gil-Ramirez, G.; Baron, M. G.; Croxton, R.; Gonzalez-Rodriguez, J., Analysis of omeprazole and esomeprazole obtained from traditional pharmacies and unlicensed internet websites using Raman spectroscopy, 1H-NMR and chemometric analysis. *Vib. Spectrosc.* **2020,** *106*, 102996.
29. Hardcastle, F., Raman Spectroscopy of Titania (TiO2) Nanotubular Water-Splitting Catalysts. *JAAS* **2011,** *65*, 43-48.

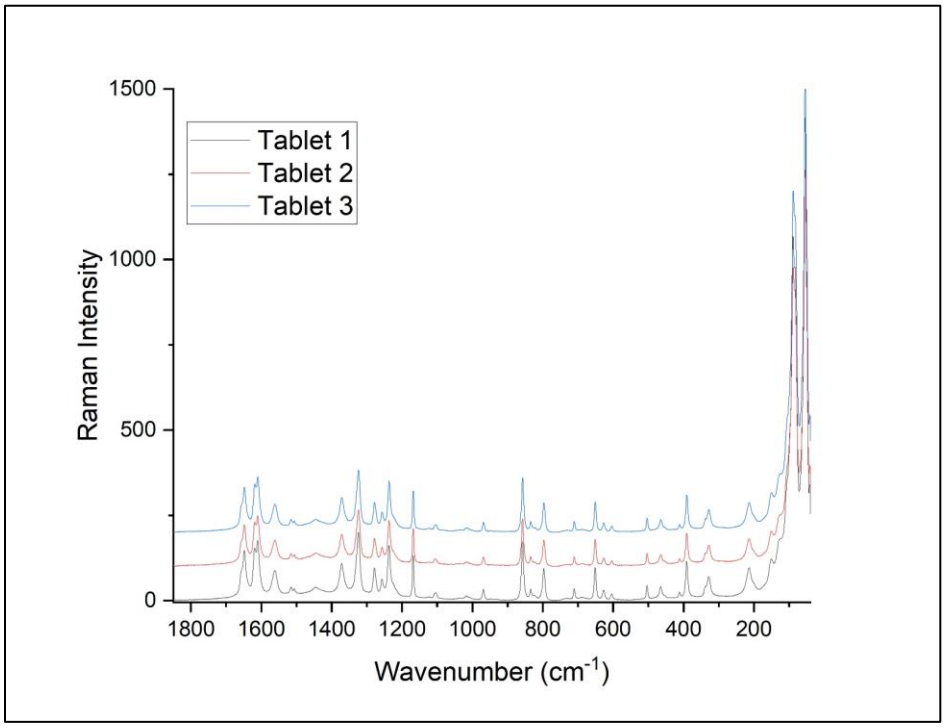
## Appendix 1

## Inactive Ingredients of Acetaminophen Samples

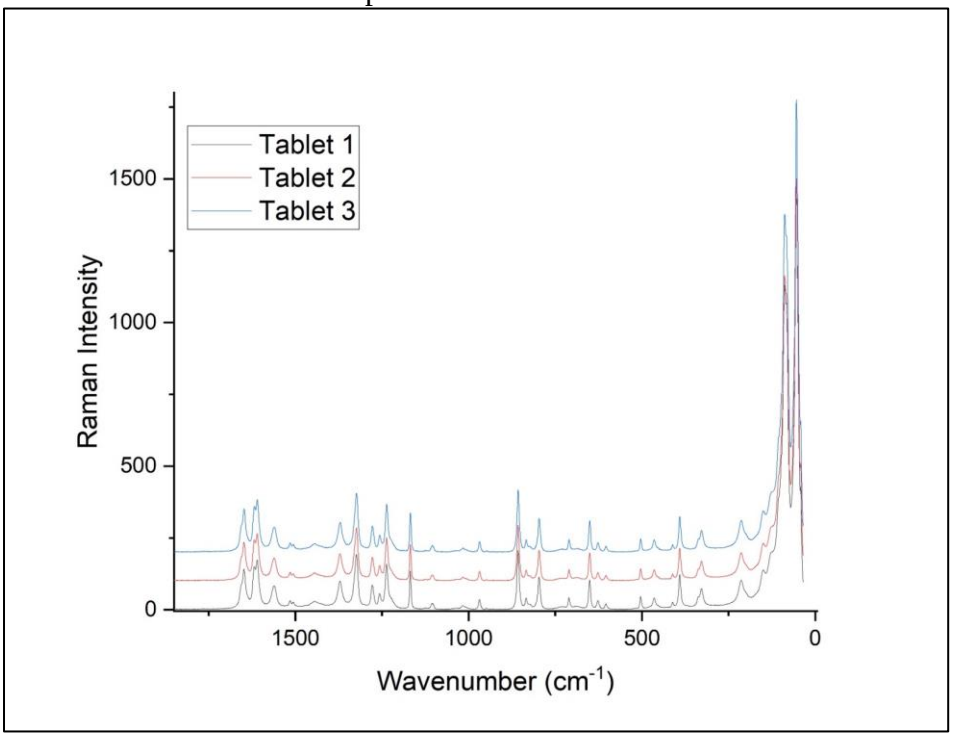
| Brand | Abbreviation | Inactive Ingredients |
|---|---|---|
| TopCare Health | TC | Carnauba wax, corn starch*, croscarmellose sodium*, hypromellose, polyethylene glycol, povidone, pregelatinized starch, sodium starch glycolate*, Stearic Acid. *contains one or more of these ingredients |
| Ready In Case | RIC | Corn starch, croscarmellose sodium*, hypromellose*, lactose monohydrate*, magnesium stearate*, maltodextrin*, medium-chain triglycerides*, mineral oil*, polyethylene glycol polydextrose*, polyvinyl alcohol*, providone, purified water*, sodium starch glycolate*, stearic acid*, talc*, titanium dioxide. *contains one or more of these ingredients |
| Walgreens | Wal | Castor oil, hypromellose, povidone, sodium starch glycolate, starch, stearic acid |
| Equate | EQ | Carnauba wax, corn starch*, croscarmellose sodium*, hypromellose, polyethylene glycol, povidone, pregelatinized starch, sodium starch glycolate*, stearic acid. *contains one or more of these ingredients |
| Family Wellness | FW | Carnauba wax, corn starch*, croscarmellose sodium*, hypromellose, polyethylene glycol, povidone, pregelatinized starch, sodium starch glycolate*, stearic acid. *contains one or more of these ingredients |
| Up&Up | UU | Hypromellose, mineral oil, povidone, pregelatinized starch, sodium starch glycolate, stearic acid, titanium dioxide |
| Tylenol | SEA041 | Carnauba wax*, corn starch*, fd&c red no. 40 aluminum lake, hypromellose, magnesium stearate, modified starch*, polyethylene glycol*, powdered cellulose, pregelatinized starch, propylene glycol, shellac, sodium starch glycolate, titanium dioxide *contains one or more of these ingredients. |
| Tylenol | SEA035 | Carnauba wax*, corn starch*, fd&c red no. 40 aluminum lake, hypromellose, magnesium stearate, modified starch*, polyethylene glycol*, powdered cellulose, pregelatinized starch, propylene glycol, shellac, sodium starch glycolate, titanium dioxide *contains one or more of these ingredients. |
| Tylenol | Reg | Magnesium stearate, modified starch, powdered cellulose, pregelatinized starch, sodium starch glycolate |
| Tylenol | SHA101 | Carnauba wax*, corn starch*, fd&c red no. 40 aluminum lake, hypromellose, magnesium stearate, modified starch*, polyethylene glycol*, powdered cellulose, pregelatinized starch, propylene glycol, shellac, sodium starch glycolate, titanium dioxide *contains one or more of these ingredients. |

## Appendix 2

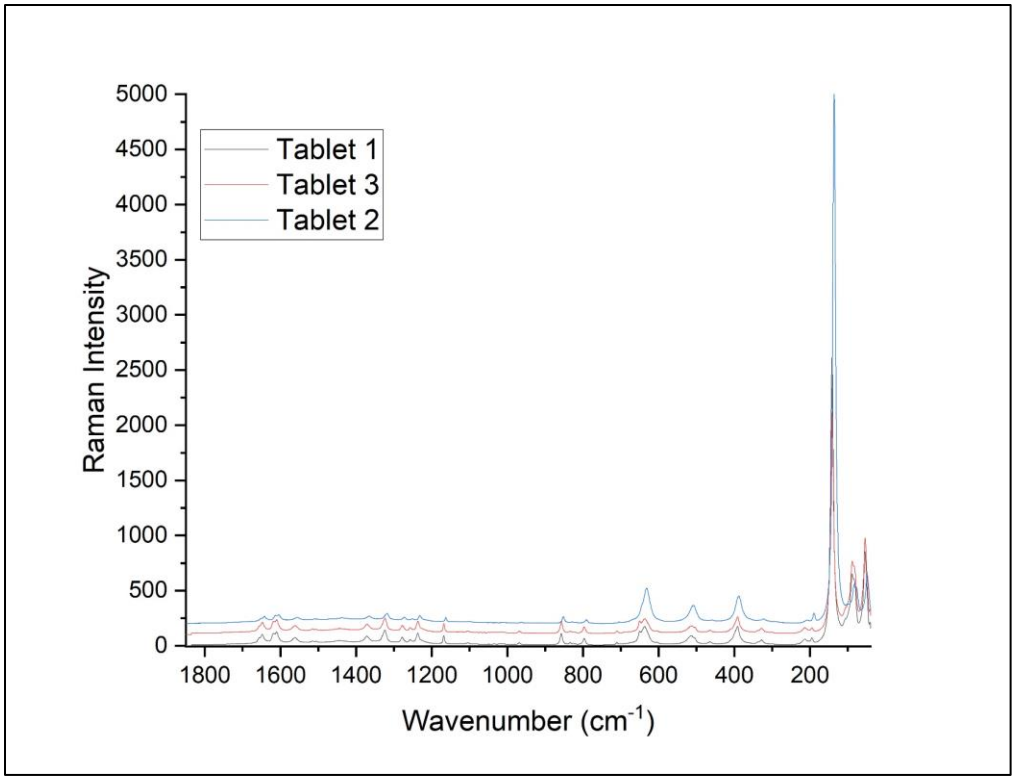
## Average Raman Spectra of Acetaminophen Tablets



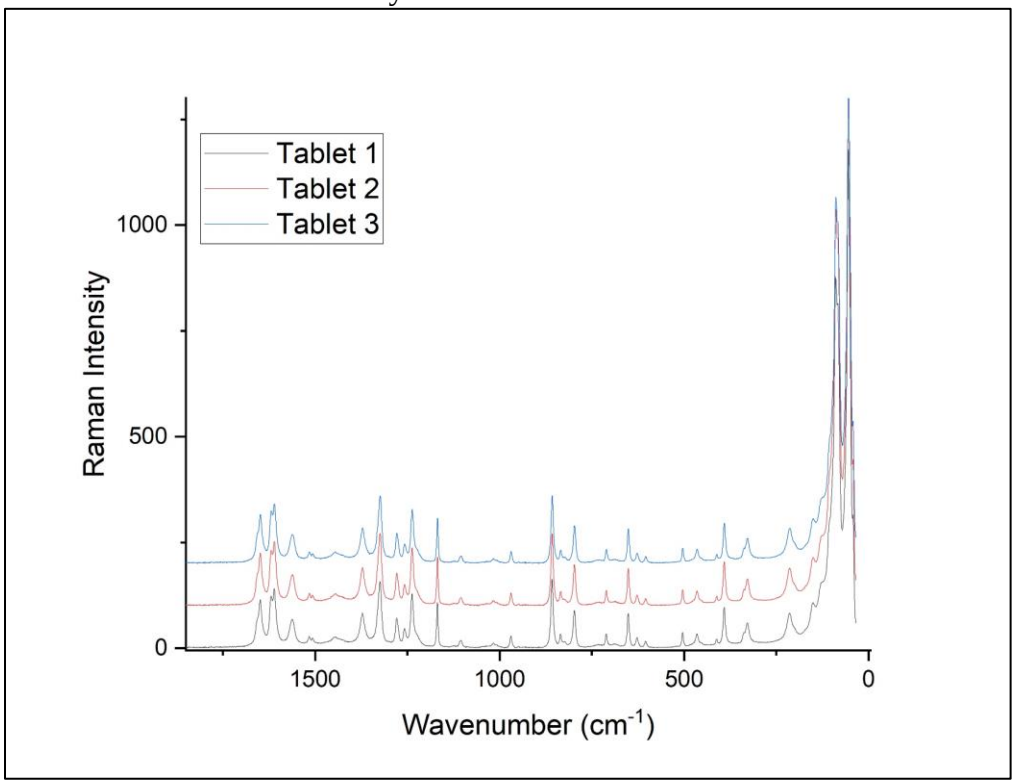

TopCare Intact Tablets



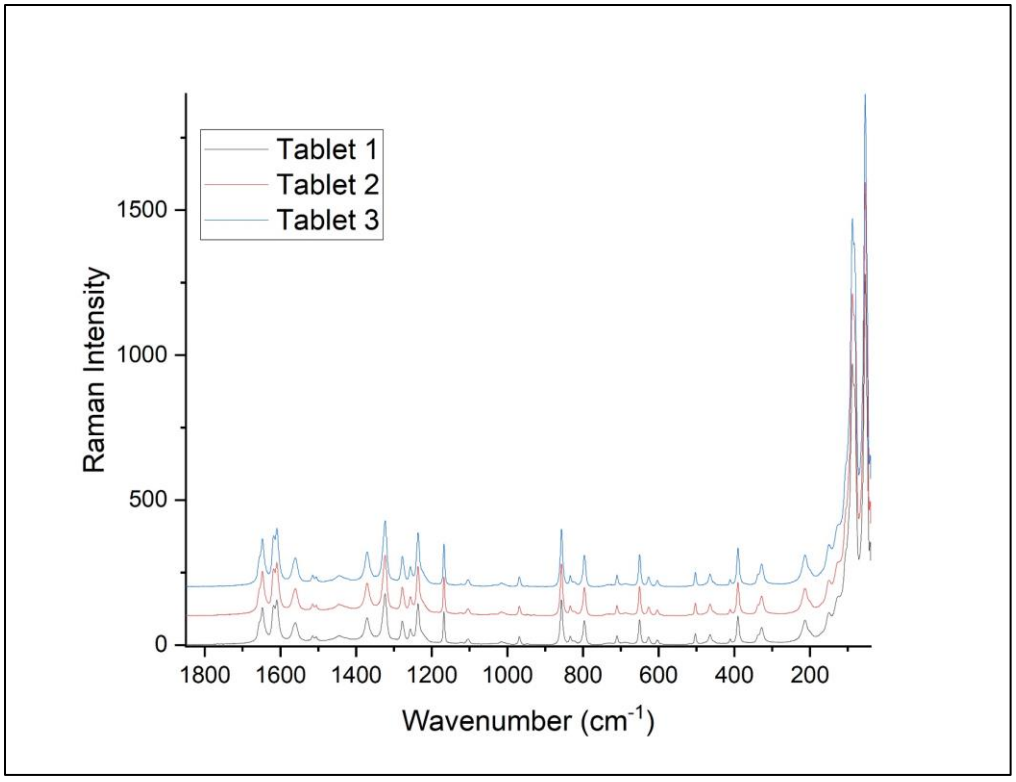

TopCare Tablet Interior



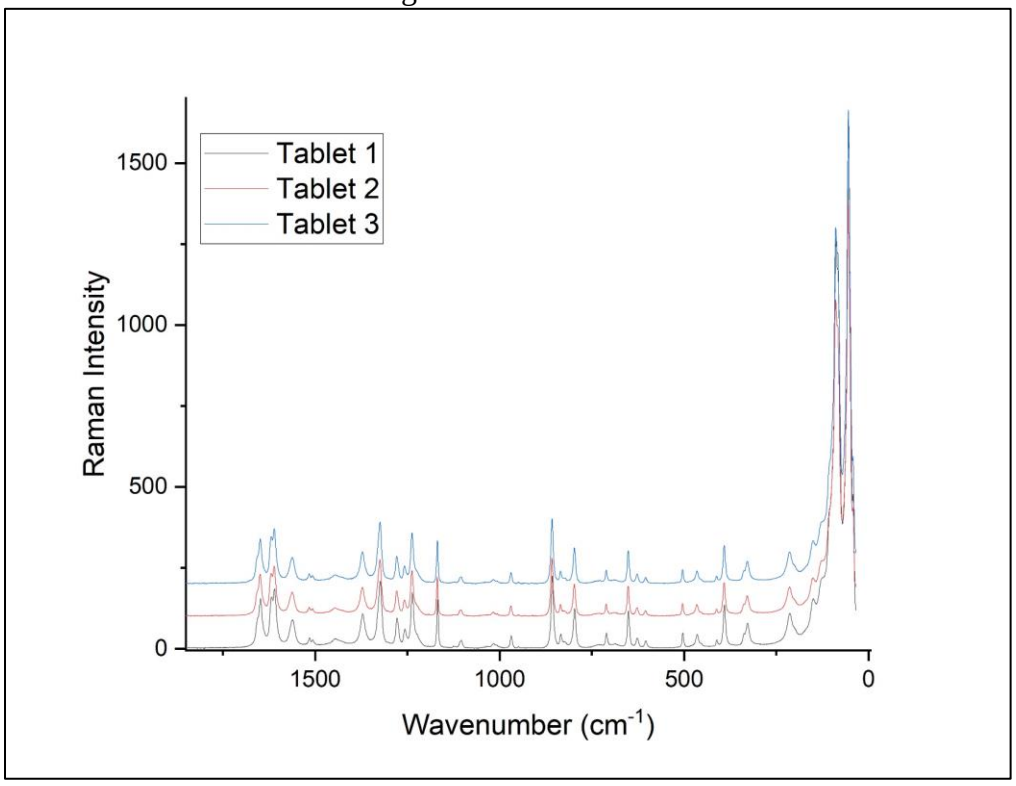

Ready In Case Intact Tablets



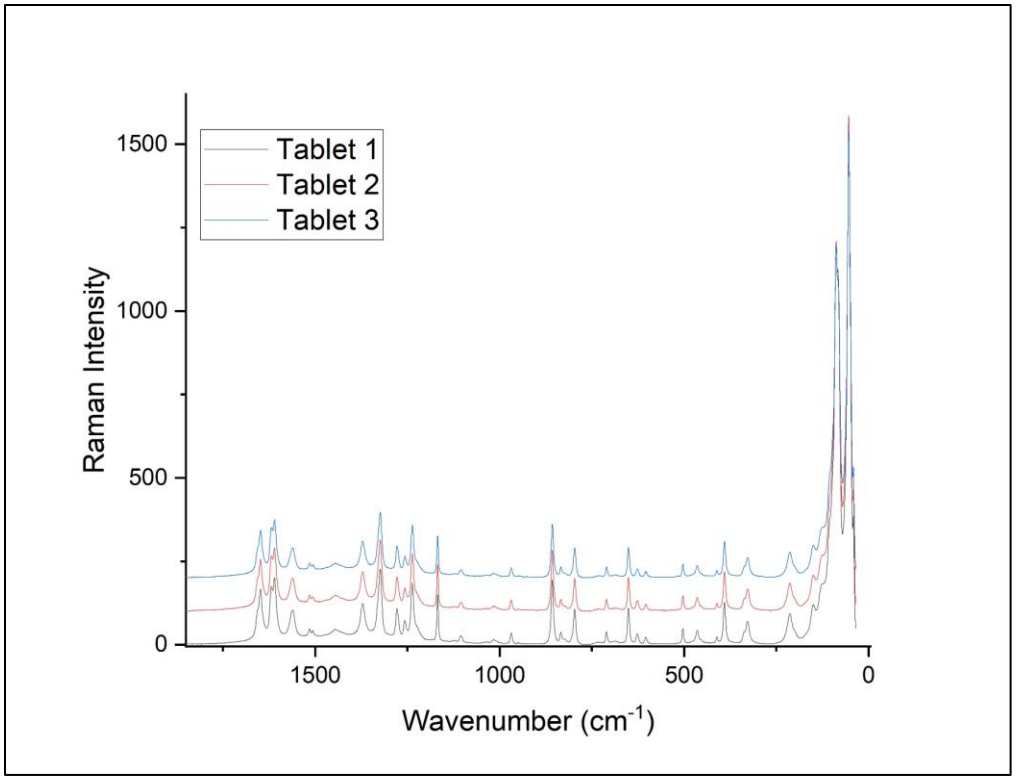

Ready In Case Tablet Interior



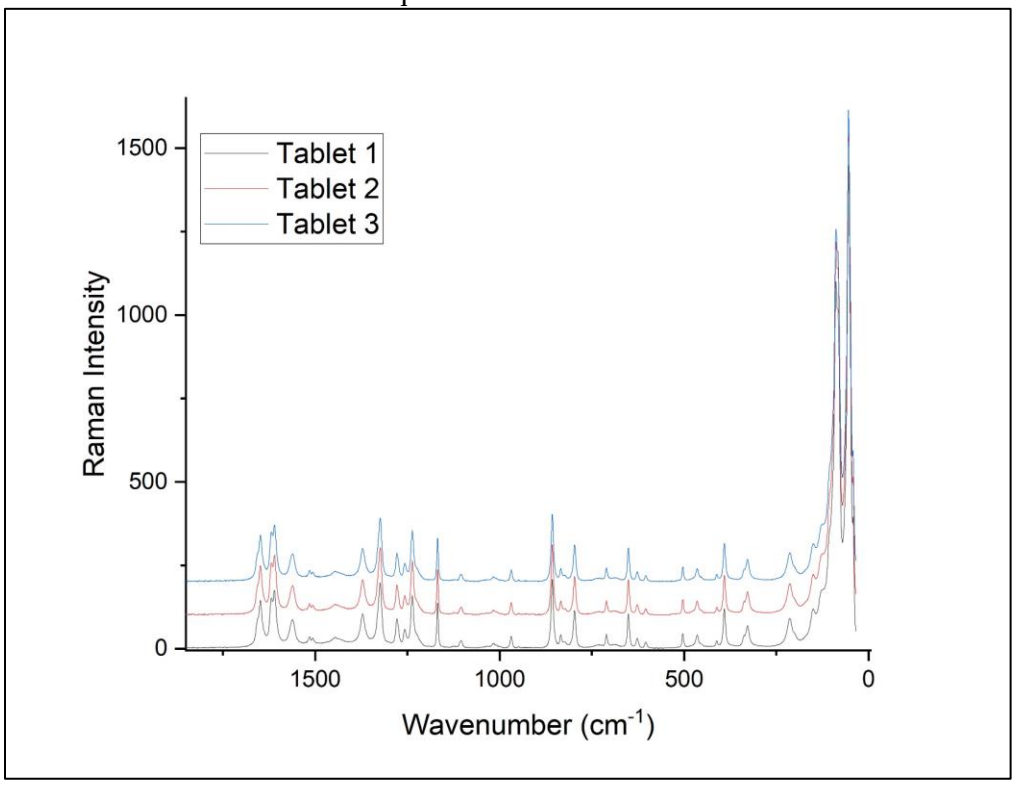

Walgreens Intact Tablets



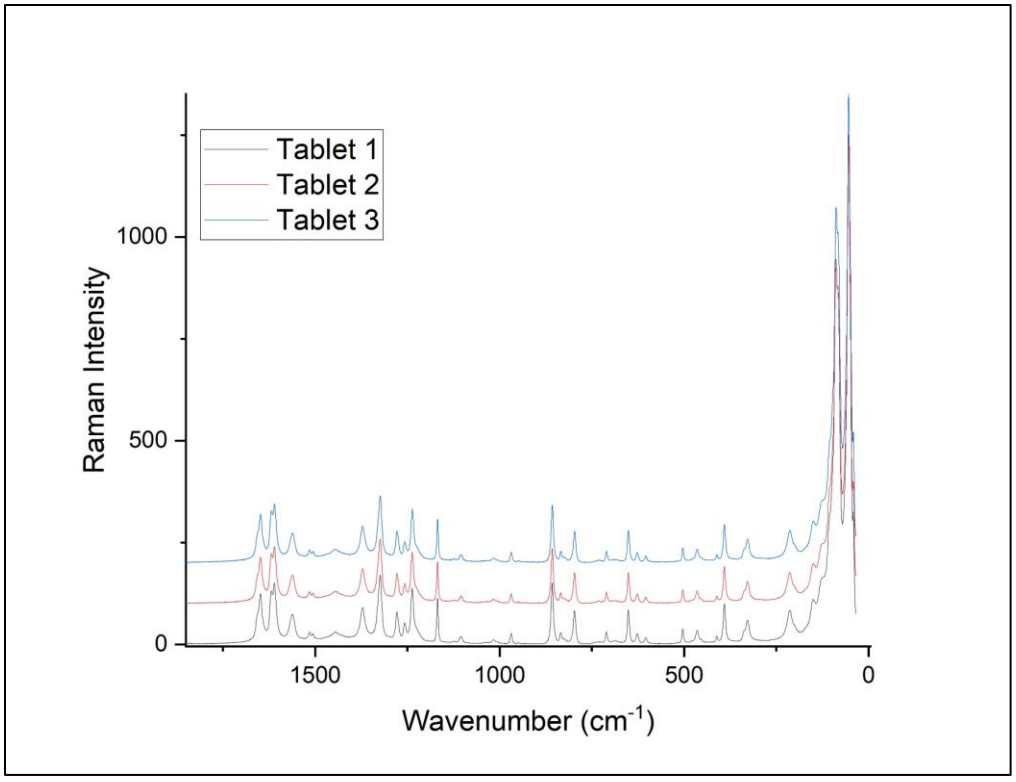

Walgreens Tablet Interior



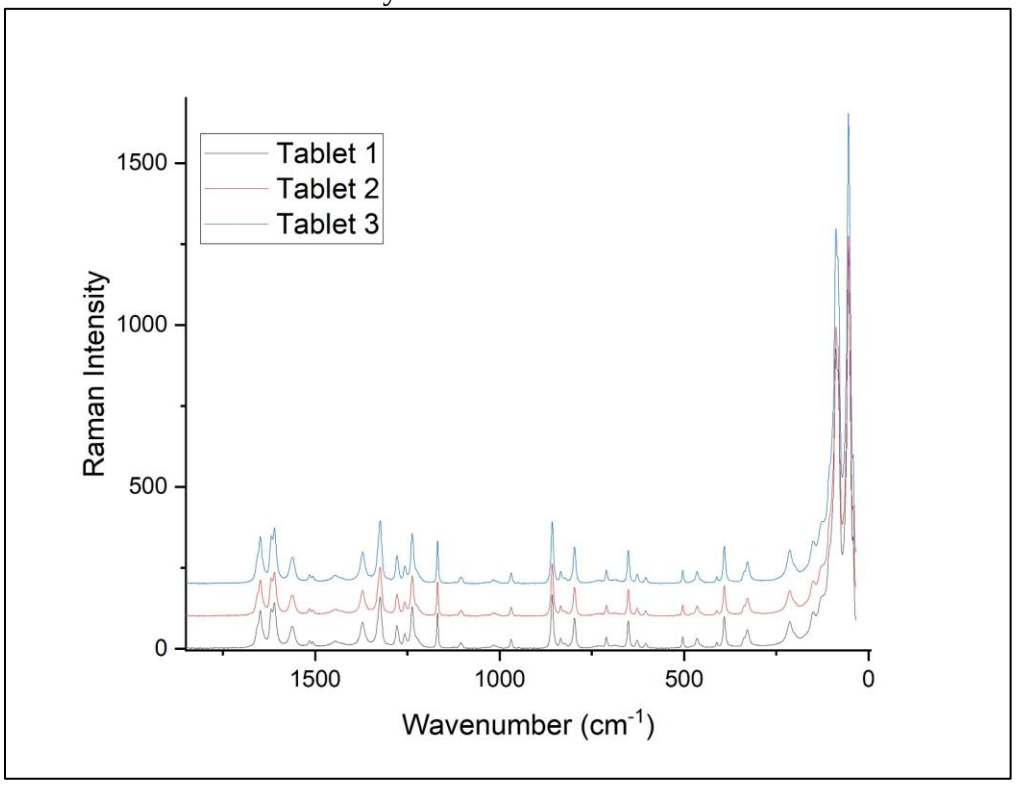

Equate Intact Tablets



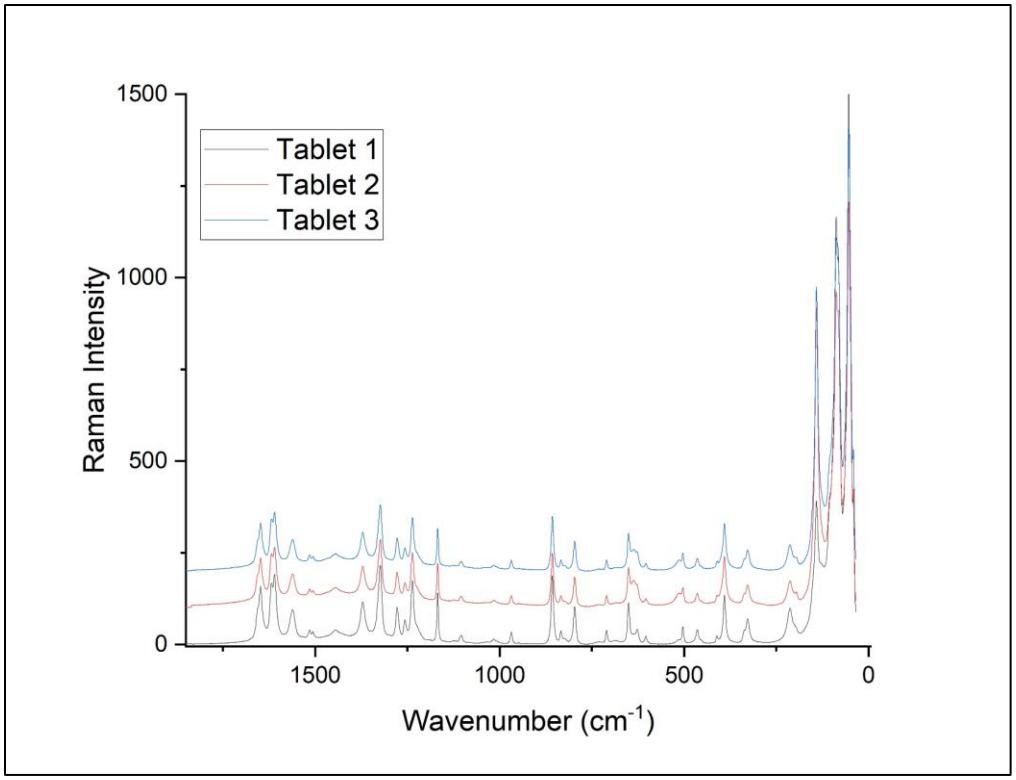

Equate Tablet Interior



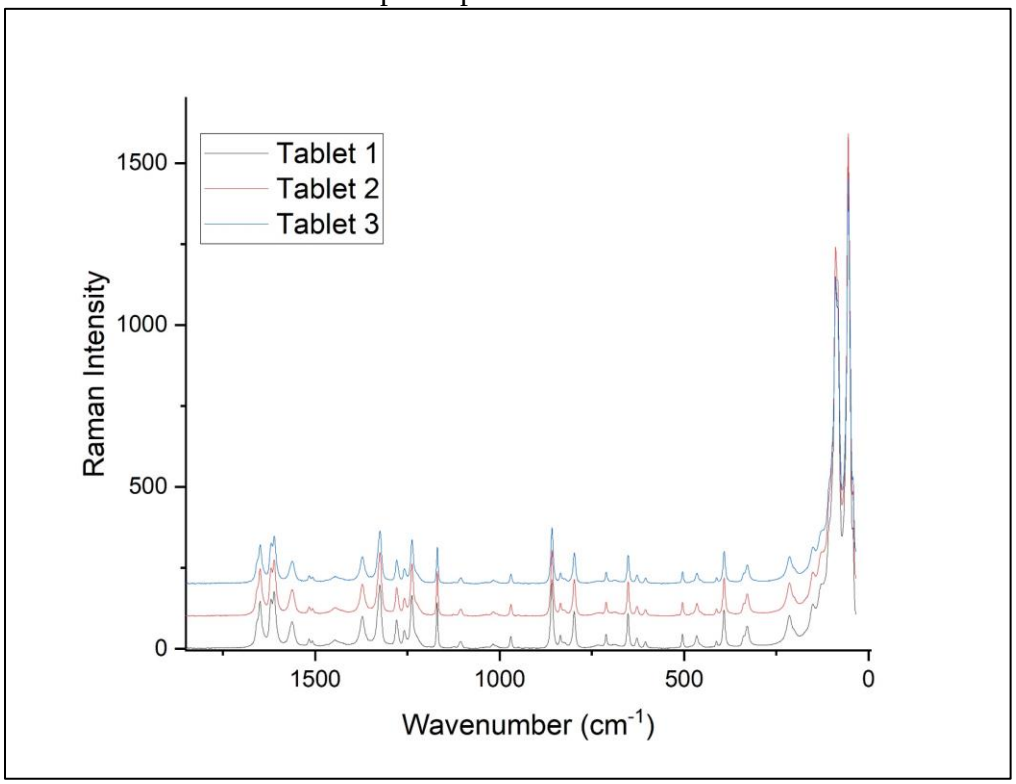

Family Wellness Intact Tablets



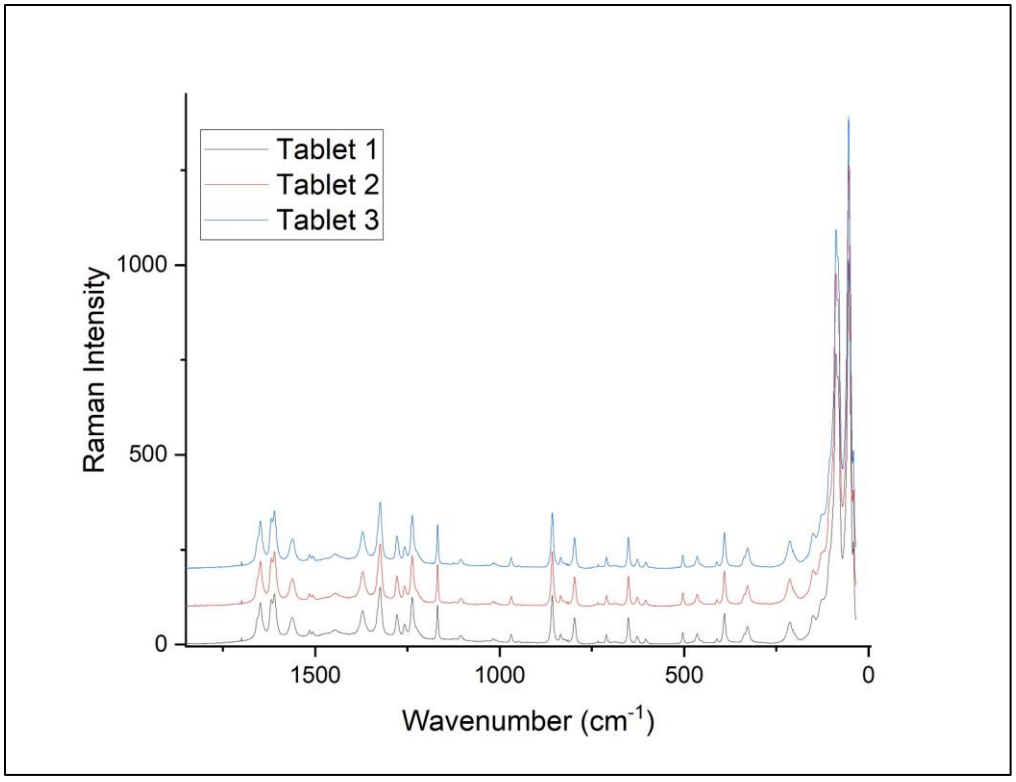

Family Wellness Tablet Interior



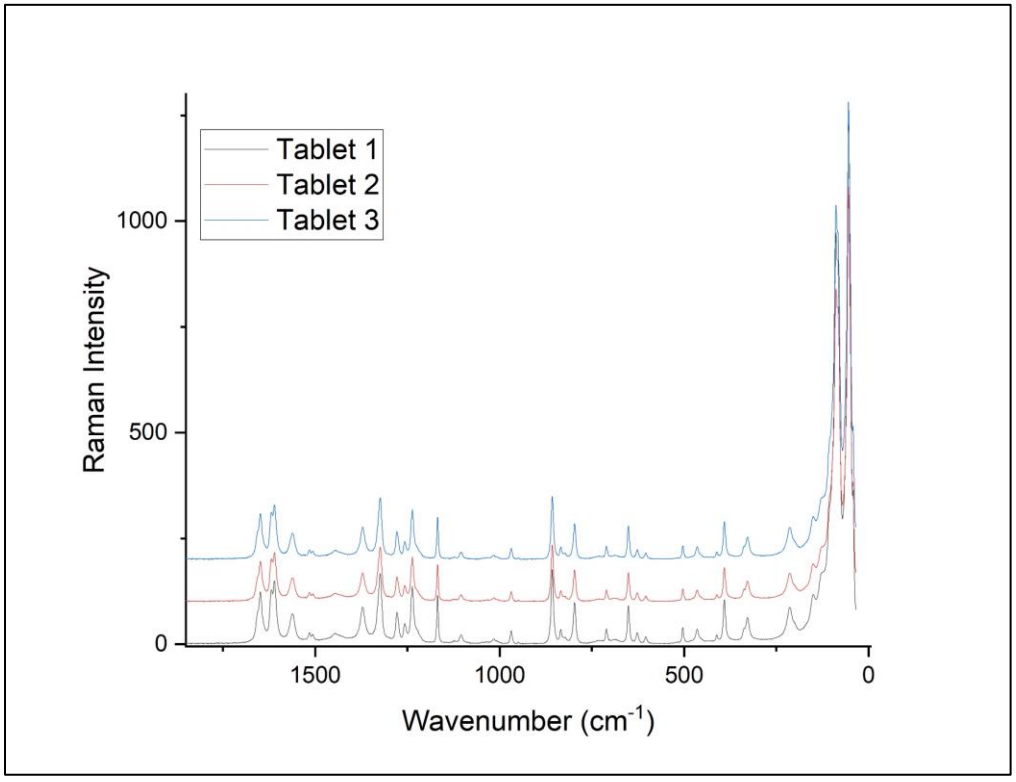

Up & Up Intact Tablets



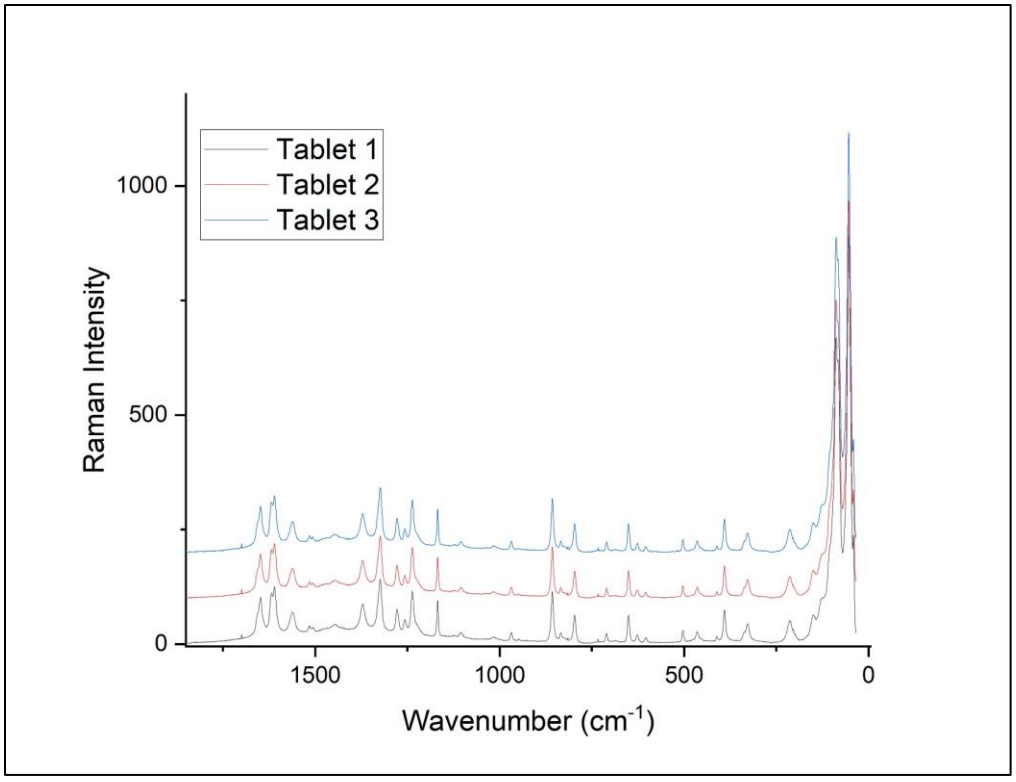

Up & Up Tablet Interior



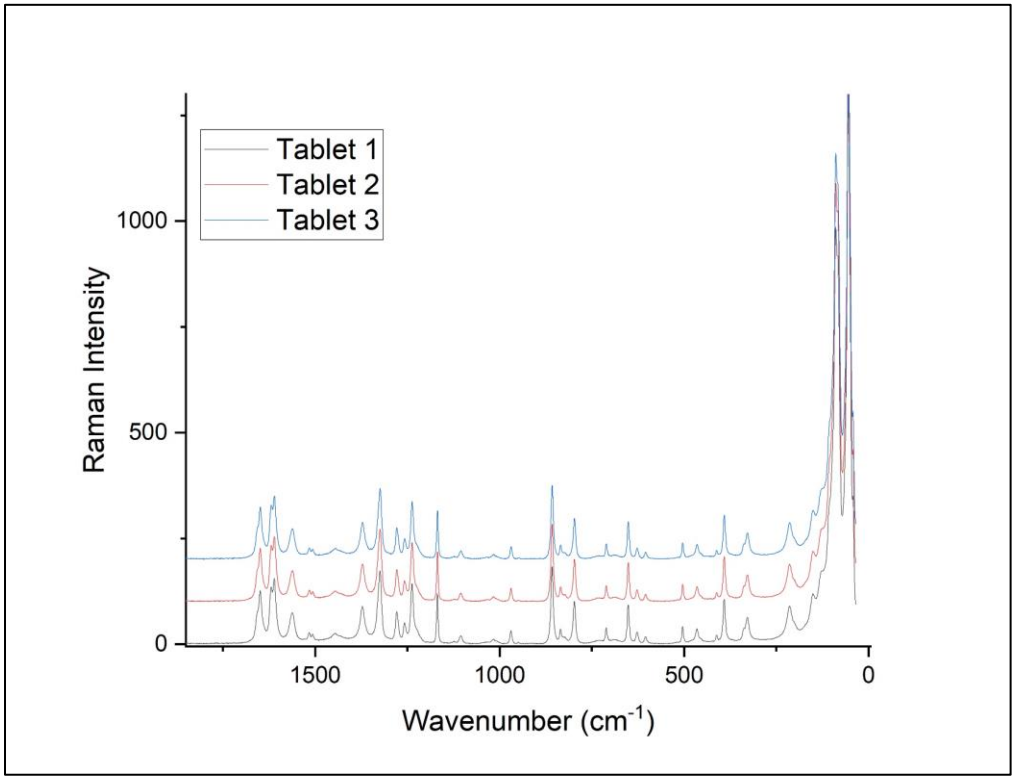

SEA041 Intact Tablets



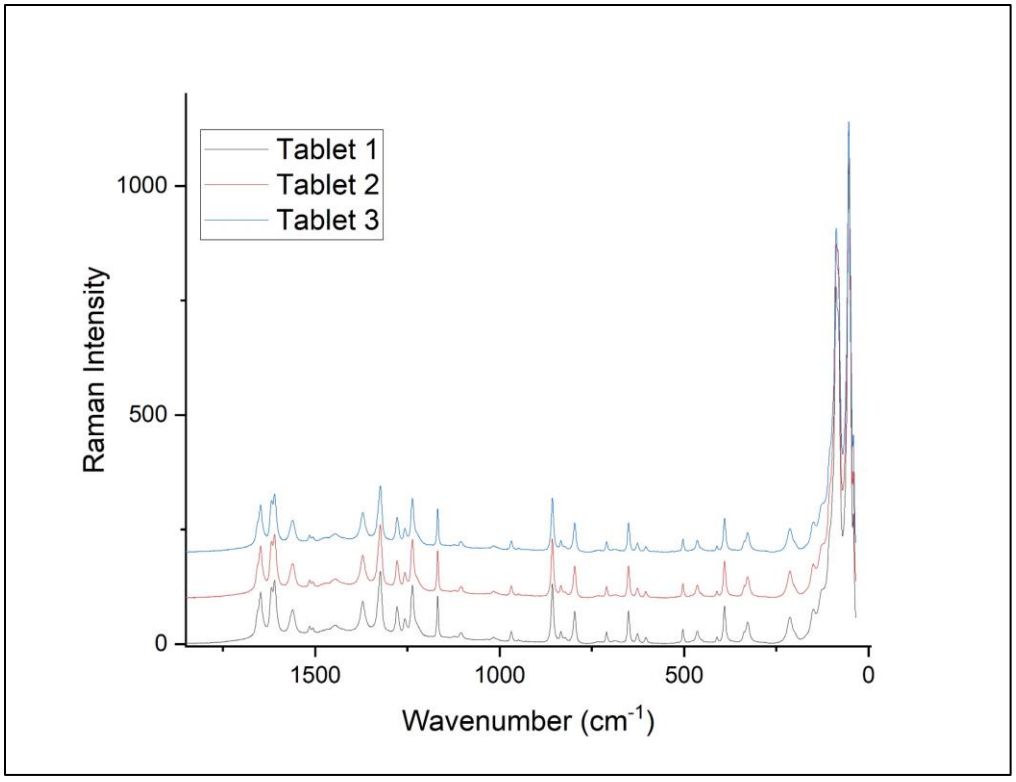

SEA041 Tablet Interior



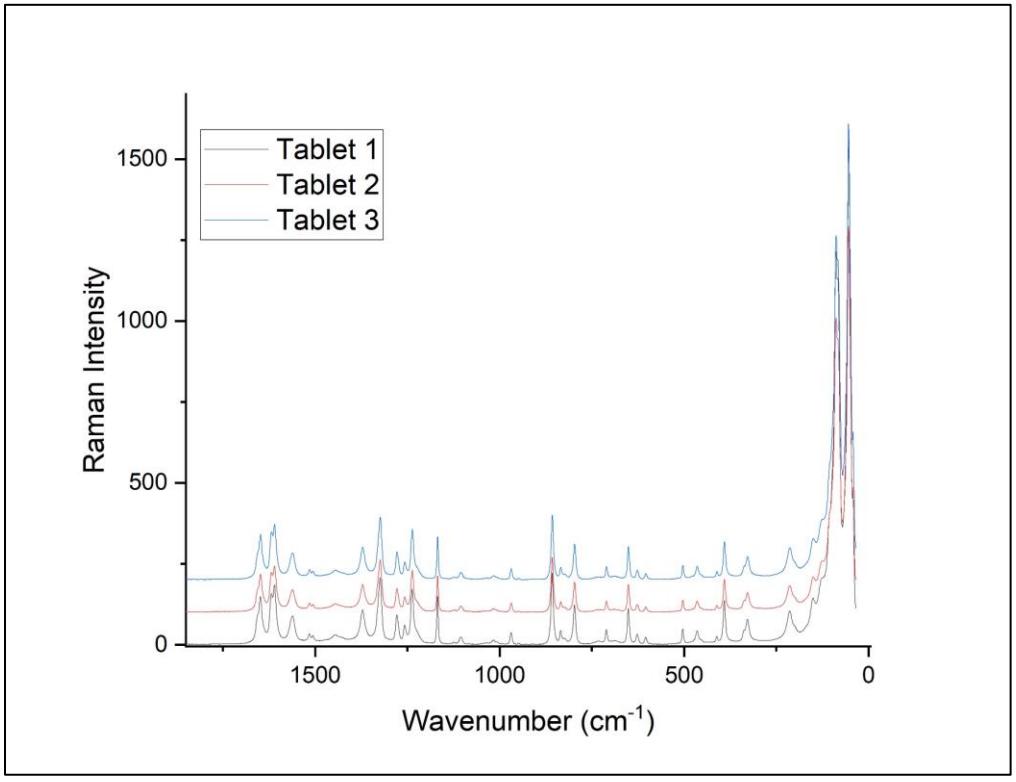

SEA035 Intact Tablets



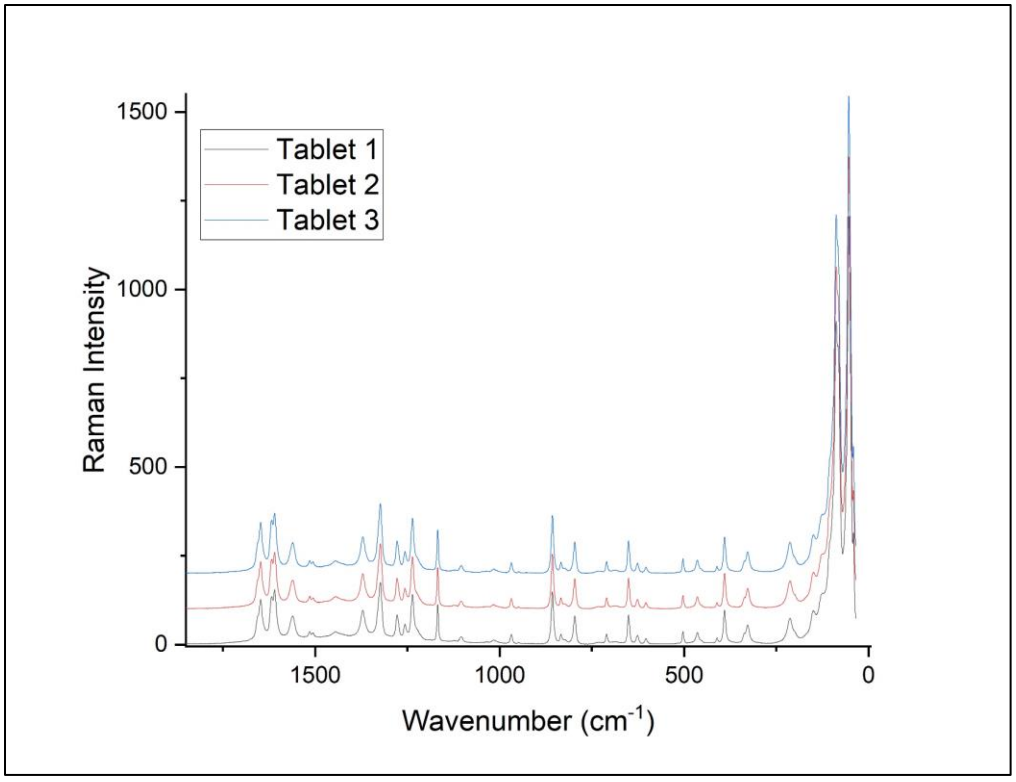

SEA035 Tablet Interior



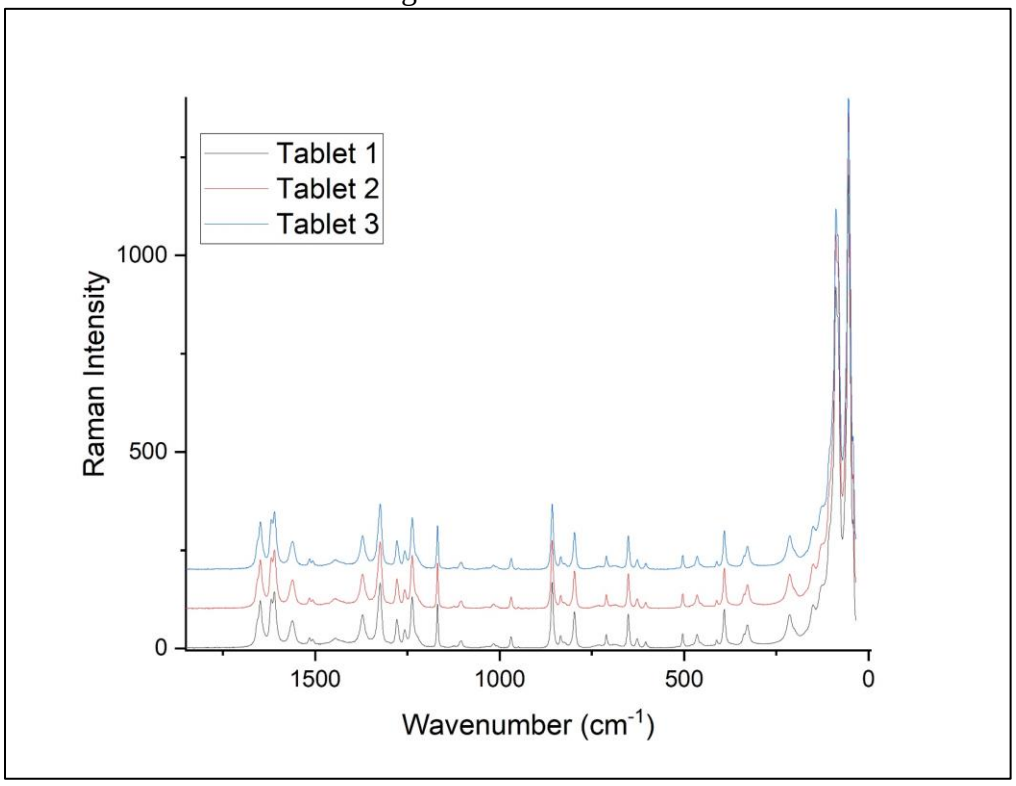

SHA101 Intact Tablets




SHA101 Tablet Interior




Regular Intact Tablets




Regular Tablet Interior