Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Jeanna Marie Feldmann

Entitled
Discrimination of Color Copier/Laser Printer Toners by Raman Spectroscopy and Subsequent Chemometric Analysis.

For the degree of Master of Science

Is approved by the final examining committee:

John V. Goodpaster Chair	
Jay A. Siegel	
Lei Lee	
Barry B. Muhoberac	

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): John V. Goodpaster

Approved by: Christine Picard 03/22/2013
Head of the Graduate Program Date

DISCRIMINATION OF COLOR COPIER/LASER PRINTER TONERS BY RAMAN SPECTROSCOPY AND SUBSEQUENT CHEMOMETRIC ANALYSIS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jeanna Marie Feldmann

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2013

Purdue University

Indianapolis, Indiana

For my family and friends who have supported me and helped me get to this point in my life. For my parents who have sacrificed so much to provide me with the best opportunities in life; I could not have achieved all I have without your unconditional love and support. For my siblings who have always helped me keep things in perspective and see the bigger picture. Para o meu amor que ficou do meu lado e encorajou os meus sonhos sem nunca deixar que a distância se tornar um obstáculo, eu te amo.

ACKNOWLEDGEMENTS

I would like to thank Dr. Siegel for giving me this opportunity and providing guidance throughout my graduate career. I would also like to thank Dr. Goodpaster for stepping in as my second advisor and giving me additional guidance. I am grateful for the wealth of knowledge that both my advisors have provided. I would like to thank Joe Stephens and the United States Secret Service Laboratory for providing the samples for this study. In addition, I would like to thank Eric Reichard for his assistance with XLSTAT, and the rest of the Goodpaster research group for their support and feedback throughout my research. I am deeply appreciative of all those who have positively impacted my graduate career.

TABLE OF CONTENTS

Page

LIST OF TABLES vi
LIST OF FIGURES... vii
ABSTRACT ix
CHAPTER 1. INTRODUCTION..1
  1.1 Copier Toners and Their Analysis 1
    1.1.1 Background 1
    1.1.2 Current Methods of Analysis/Research 3
  1.2 Chemometric Techniques for Data Analysis 4
    1.2.1 Preprocessing..6
    1.2.2 Principal Component Analysis 9
    1.2.3 Agglomerative Hierarchical Clustering 12
    1.2.4 Discriminant Analysis...15
CHAPTER 2. RAMAN SPECTROSCOPY 19
  2.1 Review of Raman Spectroscopy 19
  2.2 Blue Toners..21
    2.2.1 Materials and Methods..21
      2.2.1.1 Instumental Analysis 21
      2.2.1.2 Data Analysis 23
    2.2.2 Results and Discussion..24
      2.2.2.1 Statistical Results 24
      2.2.2.2 External Validation..31
      2.2.2.3 Formation of Classes 31
    2.2.3 Conclusions 34

Page
2.3 Yellow Toners....35
2.3.1 Materials and Methods....35
3.3.1.1 Instumental Analysis....35
3.3.1.2 Data Analysis....36
2.3.2 Results and Discussion....38
3.3.2.1 Statistical Results....38
3.3.2.2 External Validation....42
2.3.3 Conclusions....42
2.4 Magenta Toners....43
2.4.1 Materials and Methods....43
2.4.2 Conclusions....44
CHAPTER 3. OVERALL CONCLUSIONS OF THE STUDY....45
CHAPTER 4. FUTURE RESEARCH AND RECOMMENDATIONS....50
LIST OF REFERENCES....53

APPENDICES

Appendix A. Blue Toner Spectra by Raman Spectroscopy....59
A.1 Training Samples....59
A.2 External Validation....107
A.3 External Validation Key....120
Appendix B. Yellow Toner Spectra by Raman Spectroscopy....121
B.1 Training Samples....121
B.2 External Validation....165
B.3 External Validation Key....178

LIST OF TABLES

Table Page

Table 2.1. Eigenvalues and variability associated with each principal component (PC) 26

Table 2.2. Confusion matrix for cross-validation results from DA based on three classes 30

Table 2.3. Confusion matrix for the external validation results of the supplemental data from DA 31

Table 2.4. Manufacturer information separated by class 33

Table 2.5. Confusion matrix for cross-validation results from DA based on four classes 41

Table 2.6. Confusion matrix for the external validation results of the supplemental data from DA 42

Table 3.1. Discrimination of samples using combined class information for each sample’s results utilizing two colors 49

LIST OF FIGURES

Figure Page

Figure 1.1. Example of a scores plot from PCA 10

Figure 1.2. Example of PCA scree plot (courtesy of Eric Reichard) 12

Figure 1.3. Example of an AHC dendrogram 13

Figure 1.4. Example of an observations plot from DA 17

Figure 2.1. Schematic demonstrating formation of Stokes and anti-Stokes lines............ 20

Figure 2.2. Spectral Comparison illustrating lack of paper interference in blue toner spectrum.. 23

Figure 2.3. Observations plot from PCA for the blue toners with three groups/classes shown 25

Figure 2.4. Scree plot of principal component factor scores F1-F50 for blue toners 28

Figure 2.5. Observations plot from DA based on three classes of blue toners 30

Figure 2.6. Blue toner central objects of each class determined by DA.......................... 34

Figure 2.7. Spectral comparison illustrating the interference of paper in the yellow toner spectrum 37

Figure 2.8. Example of a yellow toner resultant spectrum after spectral subtraction of paper 37

Figure 2.9. Dendrogram from AHC of yellow toner replicates forming four classes 38

Figure 2.10. Central objects of the four yellow classes as determined by AHC.............. 39

Figure 2.11. Scree plot of principal component factor scores F1-F35 for yellow toners 40

Figure 2.12. Observations plot from DA based on four classes of yellow toners............ 41

Figure Page
Figure 3.1. Blue pigments (a). copper phthalocyanine, (b). victoria blue, (c). azo-pigment 46
Figure 3.2. Yellow pigments (a). acetamide, (b). azo-pigment........ 46
Figure 3.3. Micro pictograph of colored documents under 20X magnification 48

ABSTRACT

Feldmann, Jeanna Marie. M.S., Purdue University, May 2013. Discrimination of Color Copier/Laser Printer Toners by Raman Spectroscopy and Subsequent Chemometric Analysis. Major Professors: John Goodpaster and Jay Siegel.

Toner analysis has become an area of increased interest due to the wide availability of laser printers and photocopiers. Toner is most often encountered on paper in questioned document analysis. Because of this, it is important to develop methods that limit the interference of paper without damaging or destroying the document. Previous research using Fourier transform infrared spectroscopy (FTIR) has differentiated toners based on their polymer resin components. However, Raman spectroscopy and chemometric analysis are not typically used for the examination of this material.

Raman spectroscopy is a popular tool for the chemical analysis of pigmented samples and was used to characterize cyan, yellow, and magenta toners. Analyses were performed using a dispersive micro-Raman spectrometer equipped with a 785nm diode laser, a CCD detector, and an objective at 20X magnification. One hundred samples of each color toner were collected. Three different and separate methods were developed for cyan, yellow, and magenta toners on paper to optimize results. Further analysis of the magenta toners was excluded due to a weak signal and significant paper interference. The data collected from the analyses of the blue and yellow toners was then processed using a combination of statistical procedures, including principal component analysis (PCA), agglomerative hierarchal clustering (AHC), and discriminative analysis (DA). Ninety-six blue toners were analyzed by PCA and three classes of spectra were suggested. Discriminant analysis showed that the three classes were well-differentiated with a cross-validation accuracy of 100% for the training set and 100% cross-validation accuracy for the external validation set. Eighty-eight yellow toners were analyzed by AHC and four classes of spectra were suggested. Discriminant analysis showed good differentiation between the classes with a cross-validation accuracy of 95.45% for the training set, but showed poor differentiation for the external validation set with a cross-validation accuracy of 72%. While these toners were able to be discriminated, no correlation could be made between the manufacturer, printer make and model, and the toner sample.

CHAPTER 1. INTRODUCTION

The purpose of this study was to develop a method to analyze color laser printer/copier toners in situ via Raman Spectroscopy. A quick, non-destructive method that would limit any interference from the paper substrate was the primary focus. The second aim of this study was then to discriminate these toners using multivariate statistical methods often referred to as chemometrics.

1.1. Copier Toners and Their Analysis

1.1.1. Background

There are two main types of printer cartridges: inkjet and xerographic toner. The latter is used in laser printer and xerographic copier processes. These toners are the focus of this study. Unlike inkjet cartridges which use liquid ink, toner cartridges use a dry powder containing a variety of components. These components include a fusible copolymeric resin, iron oxide, carbon black, dyes or pigments, charge control agents, amorphous silica, paraffin wax, and surfactants.[1,2] Each of these components serve a specific function in the xerographic printing process.

Copolymeric resins make up anywhere from 65-80% of the mixture that is used in toner. Exact formulations vary from manufacturer to manufacturer and are considered trade secret; however, polyester, polystyrene, or polyacrylate are the base polymers of most of these copolymers.[2] The role of these specially formulated copolymers is to act as a binder for the dyes, pigments, and other components needed to adhere toner to its paper substrate.

Magnetite is an iron oxide that is used as a medium in toner to carry electrostatic charges necessary in the xerographic process. Its use can also double as a pigment in black toner.[2] Iron oxide is needed to carry the toner to the metal revolving drum, but it is not transferred or fused to the paper.

Dyes and pigments are used to add color to the toner making up the familiar cyan, yellow, and magenta colors. Nigrosine, Victoria blue, methyl violet, pthalocyanines, azo-pigments, acetamides, and quinacridones are some of the pigments known to be added to toner. The charge control agents are often complexorganometallic compounds, which can also act as dyes, or quaternary ammonium salts (both aromatic and aliphatic).[1]

Surfactants, amorphous silica, and paraffin wax are all components involved in the adhesion process. Surfactants are usually fluorinated compounds used to help ease the surface tension between the paper and the toner. Amorphous silica comes from naturally occurring silica. Amorphous basically means that the compound is not in crystalline form and has a two dimensional molecular geometry. It is exclusively used in toner.[2] Paraffin wax is used in the xerographic process to provide a medium to fuse the toner onto the paper with the help of heat from the printing device. It is either a colorless or white, in some cases translucent, wax composed of solid straight chain hydrocarbons.[2]

Toner cartridges work by using three main parts: the toner hopper which holds the toner powder, the developer unit which is an assortment of negatively charged magnetic beads, and a metal drum that the beads are attached to. The revolving drum coats the entire sheet of paper with a positive electric charge. Then a laser removes the positive charge in the places where the image is going to be printed, leaving behind a negatively charged electrostatic image. Since the toner contains compounds that carry a positive charge, namely iron oxide, the negatively charged beads pick up the toner from the hopper. As it is being rolled over the paper, the toner is attracted to the places where the laser created a negative image. Before the page is printed it out, it goes through a pair of heated rollers called a fuser which melts the toner onto the page.[2]

Unlike ink, which absorbs into the fibers of the paper, toner remains on the surface of the paper, due to the physical process described above. In quadra-color (BCYM) printing processes, the colors are placed down in layers. This can make it difficult to pinpoint a specific color toner in documents containing multiple colors. However, separation of colors can be seen along the edges of the images, making analysis of individual color toners possible when a microscopic aperture is employed.

1.1.2. Current Methods of Analysis/Research

For general information, several reviews on the forensic analysis of photocopies have been written by Totty.[3,4] Nondestructive techniques have become of particular focus for the analysis of forensic evidence with the use of optical techniques, such as infrared luminescence, infrared reflectance, and laser luminescence, for the examination of color photocopies.[5-7] Scanning electron microscopy coupled with energy dispersive X-ray (SEM-EDX) has been used to study the surface morphology and elemental composition of photocopy toner on documents.[8,9] Other important analytical techniques used in the analysis of toners are gas chromatography/mass spectrometry (GC/MS),[10-17] laser ablation inductively coupled plasma time-of-flight mass spectrometry (LA-ICP-TOF-MS),[18] infrared spectroscopy [IR, Fourier transform infrared (FTIR)],[15,16,19-24] and diffuse reflectance (DR).[5,25-27] Several studies have also explored the use of Raman spectroscopy for the analysis of various pigments.[28,29] Most recently, some studies have emerged examining color toners using Raman spectroscopy;[30,31] however, these studies involved a low number of samples, lacked chemometric analysis, and varied in parameters when compared with this work.

A brief study using Differential Scanning Calorimetry (DSC) showed potential usefulness for thermal analysis but was not pursued due to the time-consuming nature of the technique.[9] DSC was also initially considered in this study. The aim was to sample the various toners in situ. Minimal destruction to the original document occurred by the use of microplug samples. However, the paper proved to be too large of an interferent in the thermogram. An attempt to spectrally subtract the paper was explored, but proved unsuccessful. The paper to toner ratio was too large, resulting in a minimal toner sample which could not be detected. The sample size was limited to the sample holder, thus preventing a larger sample from being analyzed. Extraction of the toner was also attempted using a variety of solvents, however not all components were able to be removed, limiting the information that could ultimately be obtained. In addition, the extraction showed inconsistent results between documents and colors. The extraction process required a large area of the document, destroying the sample in question. The use of a microDSC with modulation capabilities was pursued and showed potential results. The toner was able to be minimally detected on paper, as this instrument had greater sensitivity and lower levels of detection. However, this technique was ultimately unsuccessful due to the dominance of paper interference, its time consuming nature, and instrument unavailability.

1.2. Chemometric Techniques for Data Analysis

Chemometrics is the application of multivariate statistical analysis to chemical data, e.g. spectra and chromatograms. Multivariate statistical analysis is becoming more common in forensic chemistry settings where data interpretation and comparison is standard practice. Identifying patterns and interpreting differences in data is typically done visually by the analyst, but chemometrics has made this task more accurate, objective, and manageable.[32] Chemometric analysis is particularly useful in analyses involving large data sets with large quantities of variables, as is the case with this study. Multivariate statistics have already been applied to many types of trace evidence since it was first introduced into the forensic discipline, including accelerants, inks, fibers, document examination, ammunition, gun powder, glass and paint.[32]

As just mentioned, forensic scientists often rely on visual comparisons of spectra to determine the possibility of a common source of origin in known and unknown analyses. However, this results in conclusions with no statistical basis, thus raising concerns of admissibility into a court of law. These concerns were raised in *Daubert v. Merrell Dow Pharmaceuticals* and again brought forth by the National Academy of Sciences in their report: “Strengthening Forensic Science in the United States: A Path Forward.”[33,34] Incorporating chemometric analysis into forensic casework can address two of the recommendations set forth in the NAS report. Recommendation 3 addresses the accuracy, reliability, and validity of trace evidence analysis, while recommendation 5 addresses human observer bias and sources of human error that occur during trace evidence analysis.[34]

Chemometric methods are utilized to (1) reduce the complexity of data, (2) sort and group variables of large data sets, and (3) investigate the dependence or correlation of variables with one another, thus predicting placement of unknown samples, or constructing hypotheses.[35] Chemometric techniques are a highly effective tool when large, complex data sets have been acquired.[32] After preprocessing the data, three chemometric techniques were utilized in this study: Principal Component Analysis (PCA), Agglomerative Hierarchical Clustering (AHC), and Discriminant Analysis (DA). The underlying principles of some of these techniques are not new, but rather, have been around since the early 20th century.[32] The theory behind PCA was first introduced by Pearson in 1901. However, the algorithm for computing principal components was not introduced until 1933 by Hotelling. In 1936, Fisher developed DA and Mahalanobis developed the distance measurement bearing his name which would be utilized in DA.[32]

1.2.1. Preprocessing

Preprocessing is defined as any mathematical manipulation of the data prior to the primary analysis.[36] Preprocessing data is used to reduce or remove any irrelevant sources of variation that could cause confusion in the primary modeling tool and complicate data interpretation. However, preprocessing can sometimes negatively impact the data, so techniques should be chosen and applied carefully based on known characteristics of the data. Preprocessing techniques can be applied to either the samples or the variables in a data set. Four common preprocessing techniques were employed in this study: background correction, smoothing, baseline correction, and normalization.

Background correction is used to keep variation in background levels from creating confusion during interpretation. In Raman spectroscopy, fluorescence is often problematic and can dominate the background of the spectrum.[32] Background correction can be accomplished in several ways. The first way is to subtract a straight line or polynomial from the baseline of the spectrum. A Savitzky-Golay algorithm exists for background correction. This algorithm replaces each data point with the derivative of the smoothing polynomial at that given data point. In addition, background correction can be done by replacing sample vectors with their first derivative.[32,36]

Data smoothing can increase the signal-to-noise ratio by removing unnecessary noise from the spectrum. However, smoothing can have adverse effects on a spectrum causing distortions in peak height and width, impair resolutions of peaks, and result in the loss of some features.[32] Smoothing can take place in one of several ways. Mean smoother, running mean smoother, running median smoother, and running polynomial smoother are discussed here.[32,36] A mean smoother is used to decrease the number of variables in a sample vector. Running mean and median smoothers cause “end effects” because the ends of the vector cannot be smoothed in the same manner as the other points. These running smoothers are complementary and sometimes a combination of the two are used, as one is more effective at reducing noise and the other more effective at removing arbitrary spikes. The running polynomial smoother, including the Savitsky-Golay algorithm, is most commonly used and accomplishes its task by using a low order polynomial to fit to the points in a given window. These methods all require a window width, where the points inside the window are considered during the calculations. Therefore, the window width chosen is very important.[36]

Baseline correction accounts for systematic variation and varying background levels that may cause confusion during interpretation.[32,36] The sample vector can be written as a function equal to the signal of interest plus some background features, as presented in Equation 1.1, where $\tilde{r}(x)$ is the signal of interest and the remaining coefficients are the baseline features.

$$r(x)=\tilde{r}(x)+\alpha+\beta x+\gamma x^2+\delta x^3+\cdots \qquad \text{Equation 1.1}$$

Therefore, the baseline can be accounted for by estimating the necessary coefficients and subtracting them from each element in the sample. If an offset baseline is present, as is the case in this study, then Equation 1.1 becomes simplified as the signal of interest and the first coefficient term, α, demonstrated by Equation 1.2.

$$r(x)=\tilde{r}(x)+\alpha \qquad \text{Equation 1.2}$$

Then the baseline can be accounted for by estimating α and subtracting it.[32]

Normalization of spectra eliminates variations due to sample size, concentration, amount, and instrument response.[32,36] It typically takes place after smoothing, background, and baseline correction are completed. Normalization of spectra places all the samples in the data set on the same scale allowing for easier comparison. Samples can be normalized to unit area or unit length. The latter was performed in this study. Normalizing the data set to unit area is achieved by dividing each variable in the sample by the sum of the absolute value of all variables in the sample. To normalize to unit length each variable in the sample is divided by the square root of the sum of all values squared. Equation 1.3 represents normalizing to unit length.[36,37-39]

$$x_{norm} = \frac{x_i}{\sqrt{\sum_{i=1}^{n} x_i^2}} \qquad \text{Equation 1.3}$$

Another approach is to normalize to the maximum value, which is accomplished by dividing each variable in the sample by the maximum value among all variables in the sample.

1.2.2. Principal Component Analysis

Principal component analysis is a dimensionality reduction technique that takes advantage of the fact that the variables may not be independent of one another, but are often correlated because of the underlying information.[32] The original variables (e.g., wavenumbers) are reduced into a lower number of orthogonal and uncorrelated variables, referred to as principal components (PCs) or factor scores, that have maximum variance. The first principal component will account for the greatest variance in the data set. The second principal component accounts for the next greatest variance in a direction perpendicular to the first PC.[32] Each successive PC represents a portion of the remaining variance in the data set and is always orthogonal to the previous PC. The total number of possible principal components is the smaller of the number of samples or variables.[32] The majority of the variance is captured in the first few PCs making those containing relatively small amounts of variance negligible.

The information provided by PCA can be visualized using a scores plot and a factor loadings plot. The scores plot, shown in Figure 1.1, plots the factor score of one PC against the factor score of another for each sample. This allows one to view the information in multiple dimensions, sometimes revealing the separation or grouping of samples based on similarities. A factor loadings plot allows the analyst to view which variables contribute to the respective principal component. The contribution of each variable to the new principal component can be defined as the cosine of the angle between the variable axis and the principal component axis. These cosine values are called factor loadings, which can have the value between ±1. The factor loadings are plotted against each variable (i.e., wavenumber). When the value is positive, the variable and PC are positively correlated. If the value is negative, then a negative correlation exists between the variable and the PC. Areas where the value is close to zero have no correlation between the variable and the PC.[38]

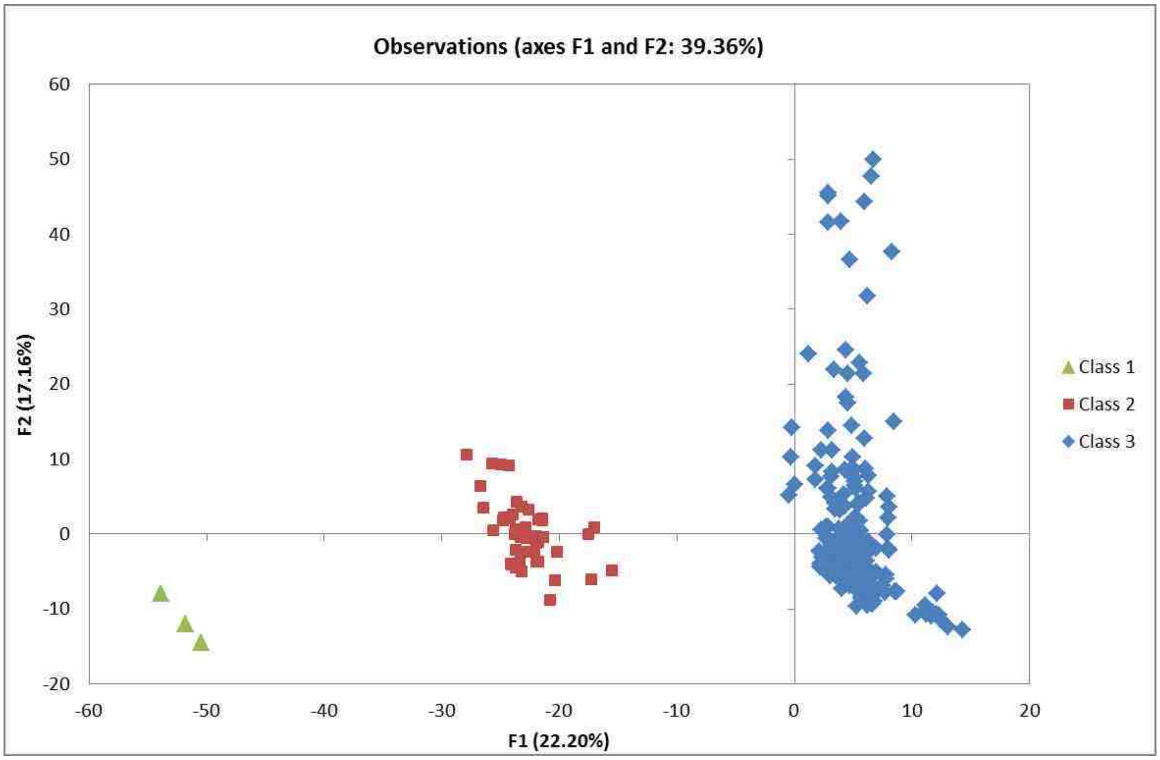


Figure 1.1. Example of a scores plot from PCA.

Principal components have eigenvalues associated with them that reflect the variance, percent variance, and cumulative variance for the principal component. The eigenvalue refers to the sum of squares of each principal component or score.[32] These eigenvalues can be used to determine the appropriate number of principal components to represent the data set in further analysis, such as DA. The first method for this determination uses a scree plot which plots eigenvalues versus factor numbers (shown in Figure 1.2). A sharp decrease in the eigenvalues is seen followed by a steady decline forming an approximate 108 degree angle. This sudden change in direction indicates the number of significant principal components. Anything to the right of this location is considered "factorial scree," or debris.[40] The second method is referred to as the Kaiser criterion, introduced in 1960 by Kaiser, which states that only factors with eigenvalues greater than one are significant.[40] This method can often result in too many PC's. The third and final method is to set a value of cumulative variance (i.e. 95%) that must be retained. The PC's that explain 95% of the total variance are retained and all remaining PC's are discarded. The first method was chosen for this study because it resulted in a fewer number of factors compared to the other two methods, introducing less noise into subsequent discriminant analysis.

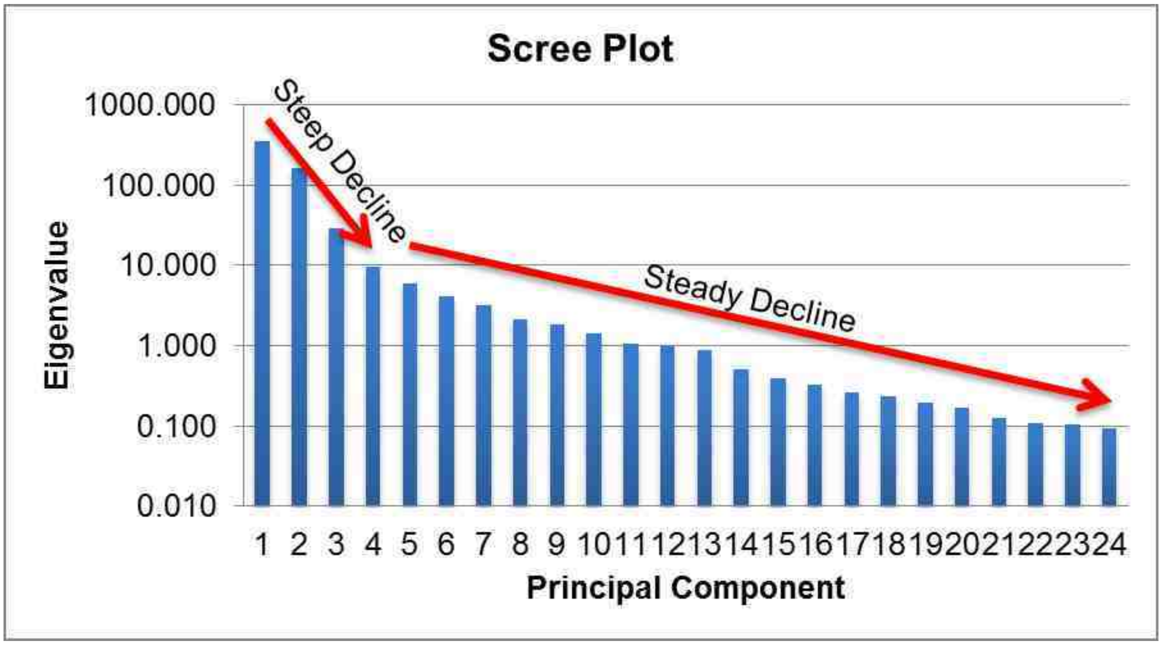


Figure 1.2. Example of PCA scree plot (courtesy of Eric Reichard).

PCA is one of the most common multivariate statistical methods used as it allows one to manage large, complex data sets by reducing the number of variables. In addition, PCA provides information as to which variable contribute to the most variance in the data set. PCA has been used in the analysis of counterfeit coins,[41] electrical tape,[42,43] gasoline,[44,45] hair dye,[46] ignitable liquids,[47] illicit drugs,[48] soil,[49] paints,[50] and inks.[51]

1.2.3. Agglomerative Hierarchical Clustering

Cluster analysis is an unsupervised technique that examines the interpoint distances between all of the samples and represents that information in the form of a two-dimensional plot called a dendrogram (shown in Figure 1.3).[32] This analysis allows one to view groupings of individual samples, detect outliers, validate the data set, and evaluate any underlying behavior of the data set. There are two main types of hierarchical cluster analysis techniques: divisive (DHC) and agglomerative (AHC).[35] These techniques begin by determining the similarity or dissimilarity between objects (i.e., the distance of each sample to all the remaining samples). In DHC all samples begin in a single cluster and are divided into smaller clusters until all samples are their own group. The sample with the largest distance separates from the group first.[35,52] In AHC, the opposite occurs with all samples beginning as their own group and the samples with the closest distance are clustered together until a single cluster remains.[35,36,52] The latter was performed in this work.

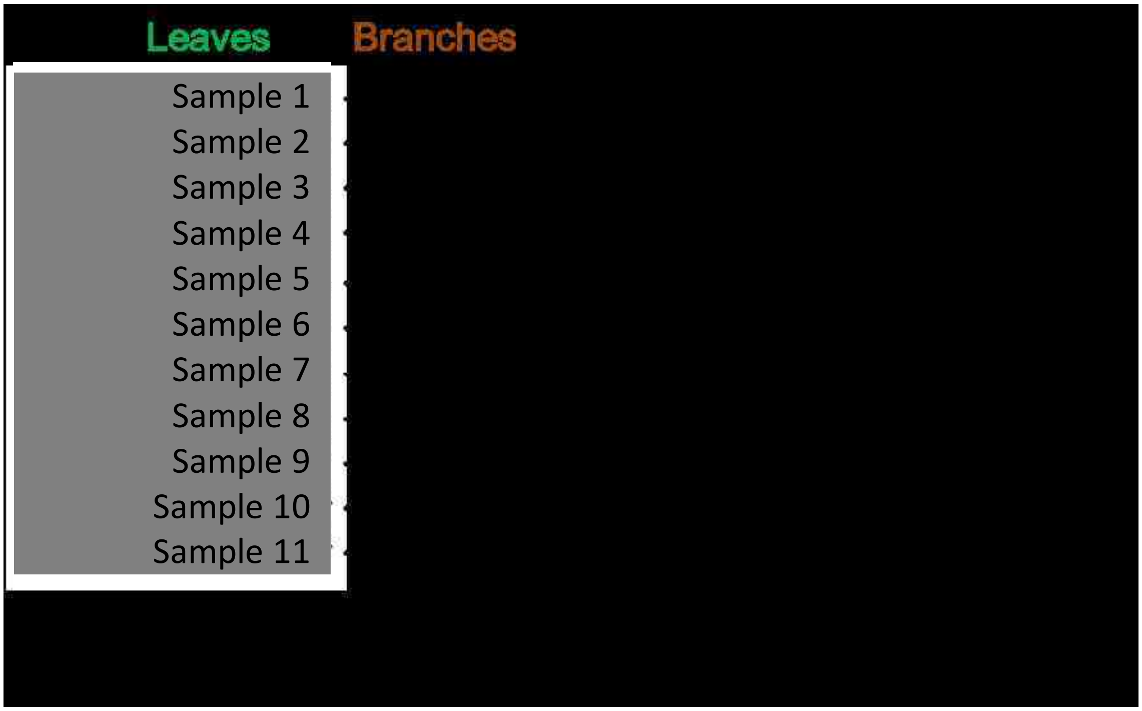

Leaves
Branches
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
Sample 9
Sample 10
Sample 11

Figure 1.3. Example of an AHC dendrogram.

The distances between objects can be measured as similarity or dissimilarity using one of several mathematical approaches. Euclidean, Manhattan, and Mahalanobis distance are three ways to calculate the dissimilarity between samples. Euclidean distance, or ruler distance, is based on the Pythagorean theorem and is calculated using Equation 1.4, where (x-y)' is the transpose of the matrix (x-y) and d_{xy} is the distance between them.[32,35,38] The smaller the value of d_{xy}, the more similar the two samples are.[38]

$$d_{x,y} = \sqrt{(x-y)'(x-y)} \qquad \text{Equation 1.4}$$

Euclidean distance is the most common method utilized and the one used in this study. While Euclidean distance represents the length of the hypotenuse of a right triangle, Manhattan distance represents the distance along the other two sides of the triangle, thus always greater than Euclidean distance.[38,39] Equation 1.5 represents the Manhattan distance.[40]

$$d_{x,y} = \sum_i |x_i - y_i| \qquad \text{Equation 1.5}$$

The Mahalanobis distance is the final method which can be used to calculate dissimilarity. This method, unlike the other two, takes into account that some variables may be correlated and therefore uses the inverse of the variance-covariance matrix as a scaling factor. It is calculated according to Equation 1.6 where C is the variance-covariance matrix.[38,39]

$$d_{x,y} = \sqrt{(x-y)C^{-1}(x-y)'} \qquad \text{Equation 1.6}$$

However, this method is inappropriate if the number of variables exceeds the number of samples. In addition to these three methods of determining dissimilarity, there are several other methods to determine similarity not discussed here.[38,39]

Once the distance is calculated, the samples are linked together using one of several methods. These methods include, but are not limited to, nearest neighbor, furthest neighbor, and Ward's method. Nearest neighbor, or single linkage, links clusters based on the distance between the two closest samples of each respective group. Farthest neighbor, or complete linkage, links clusters based on the distance between the two farthest members of each group.[53] Ward's method, the method used in this work, seeks to minimize the "loss of information," or an increase in the error sum of squares, when linking two clusters. An error sum of squares is determined by measuring the sum of squared deviation of each sample from the mean of the cluster. This method must consider all possible linkages of clusters during every step and the two clusters whose linkage results in the smallest sum of squares are linked.[38,39]

Overall, AHC is a valuable technique for initially analyzing a large data set for relationships between samples based on either similarity or dissimilarity. However, AHC does not provide information about which variables have the greatest influence on these relationships. Cluster analysis, including the AHC method, has previously been applied to the analysis of counterfeit coins,[41] electrical tape,[42,43] hair dye,[46] lighter fuel,[54] illicit drugs,[48] soil,[49,55] paint,[50] pen ink,[51] and polymers.[56]

1.2.4. Discriminant Analysis

Linear discriminant analysis, also called canonical variates analysis, is a dimensionality reduction technique as well as a pattern recognition technique. DA is a supervised technique because knowledge of group membership or class for each sample is required.[57] Similar to PCA, in which the sum of squares is maximized, the criterion maximized by DA is the Fisher ratio. This criterion, first described by Fisher, is the ratio of the variance between groups (i.e., separation between groups) divided by the variance within groups (i.e., experimental variability among spectra belonging to the same group).[32]

In DA, a new set of axes that best separates data into groups is created, placing members of the same group as close together as possible while moving the groups as far apart from one another as possible.[32,38] These discriminant axes, or canonical variates (CVs), are linear combinations of the original features as shown in Equation 1.7. The new axes are defined by f_i, while $\bar{x}_{A\ or\ B}$ represent the centroids of the classes, $x_{i'}$ is a row vector corresponding to sample i, and C_{AB} is the pooled variance-covariance matrix for groups A and B.

$$f_i = (\bar{x}_A - \bar{x}_B)C_{AB}^{-1}x_{i'} \qquad \text{Equation 1.7}$$

This matrix is calculated according to Equation 1.8 where N_A is the number of samples in group A and N_B is the number of samples in group B. The calculation of C_{AB} is for two groups but this can be expanded if more than two groups are present in the data.

$$C_{AB} = \frac{(N_A-1)C_A+(N_B-1)C_B}{(N_A+N_B-2)} \qquad \text{Equation 1.8}$$

The new axes can be plotted against one another to produce an observations plot as shown in Figure 1.4. The number of samples must exceed the number of variables so that the variance-covariance matrix can be inverted.[32,57] Therefore, PCA often precedes DA to reduce the number of variables. In addition, the Mahalanobis distance from the sample to the centroid of any given group is calculated. Samples are classified into groups based on the smallest distance value.

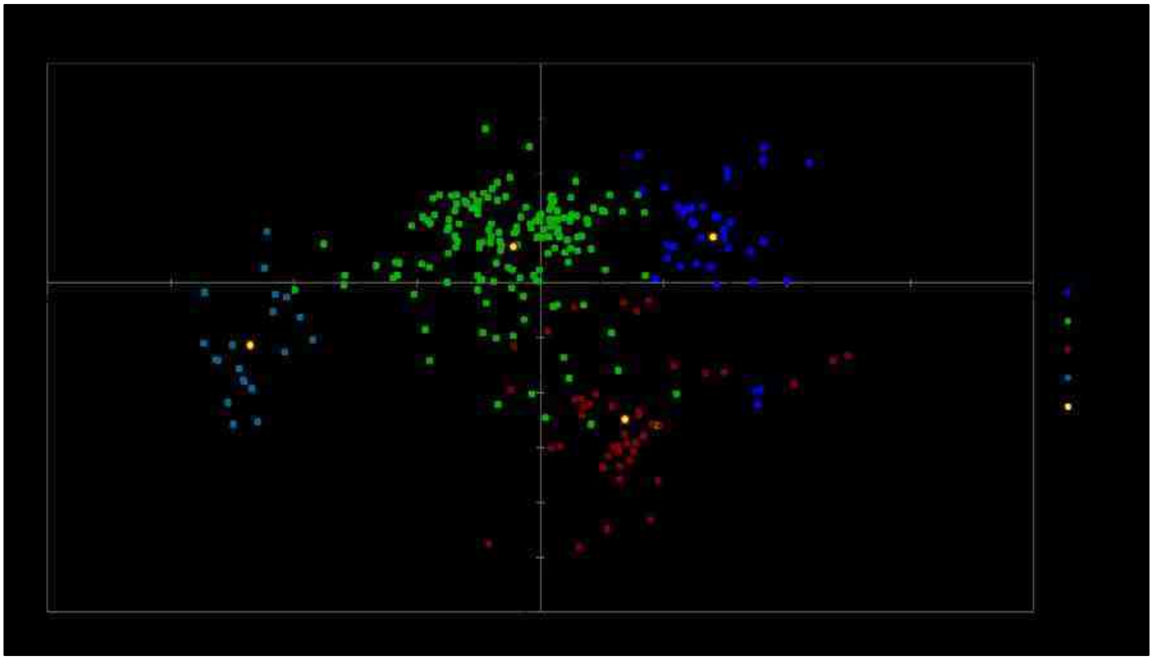


Figure 1.4. Example of an observations plot from DA.

Once these calculations have been performed and samples classified, the classification accuracy is tested by cross-validation. Cross-validation determines the probability of the sample belonging to each of the groups by removing each sample in turn from the data set, creating a classification rule, and predicting the classification of the sample.[32] There are several ways to carry out cross-validation: resubstitution, hold-out, and leave-one out method. In resubstitution, the complete data set is employed as a training set to develop a classification procedure based on the known class membership of each sample. The class membership of every sample in the data set is then predicted by the model and the accuracy of correct classification determined.[57] The hold-out method partitions the available data into two portions: a training set for development of the classification model, and a test set for prediction of classification. By separating the data used to build the classification model from the data used to evaluate its performance, the estimate of error is unbiased.[57] The leave-one-out method temporarily deletes a sample from the data set, a classifier is built from the training set of remaining observations, and the model is used to predict the group membership of the deleted sample. This is then repeated for each sample in the data set resulting in a nearly unbiased estimate of classification accuracy.[57] Each of these methods has its drawbacks which should be considered when selecting a method.

DA has been used for the analysis of electrical tape,[42,43] gasoline,[58] hair dye,[46] soil,[49] pen ink,[51] and tire rubber.[59]

CHAPTER 2. RAMAN SPECTROSCOPY

2.1. Review of Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopic technique based on inelastic scattering of monochromatic light, usually from a laser source. Inelastic scattering means that the frequency of photons in monochromatic light changes upon interaction with an analyte. In a Raman experiment, photons of the laser light are absorbed by the sample and then reemitted. The frequency of the reemitted photons is shifted up or down in comparison with the original monochromatic frequency, which is called the Raman effect. This shift provides information about vibrational, rotational and other low frequency transitions in molecules. Raman spectroscopy can be used to study solid, liquid and gaseous samples.[60]

The Raman effect is based on molecular deformations in the electric field determined by molecular polarizability. The polarizability measures the ease with which the electron cloud around a molecule can be distorted.[61] Scattered light arises from three different scattering phenomena: Rayleigh, Stokes, and anti-Stokes. Rayleigh scattering is a form of elastic scattering, while Stokes and anti-Stokes are forms of inelastic scattering. A molecule absorbs a photon and is excited from the ground state to a virtual excited state. Rayleigh scattering occurs when the excited molecule returns back to the same basic vibrational state and emits light with the same frequency as the excitation source. If the final vibrational state of the molecule is more energetic than the initial state, then the emitted photon will be shifted to a lower frequency in order for the total energy of the system to remain balanced. This shift in frequency is designated as a Stokes shift. If the final vibrational state is less energetic than the initial state, then the emitted photon will be shifted to a higher frequency, and this is designated as an anti-Stokes shift. This is illustrated in Figure 2.1.

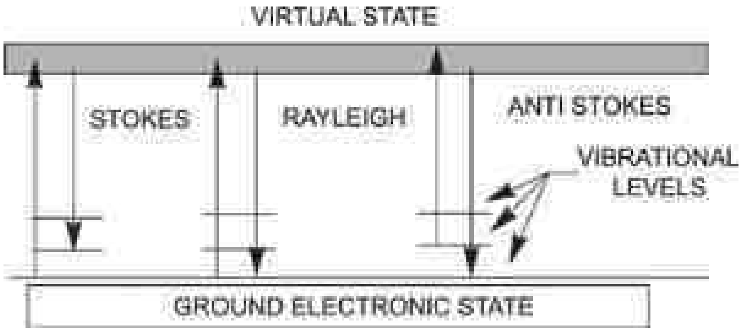


Figure 2.1. Schematic demonstrating formation of Stokes and anti-Stokes lines.[62]

Raman spectroscopy has the advantage of being a quick, non-destructive technique requiring only a small sample size. This makes Raman an ideal technique for forensic applications where preservation of evidence is crucial. Sample preparation is often minimal with highly reproducible spectra. Raman is only hindered by its low sensitivity and fluorescence interference. The fluorescence can sometimes be controlled by increasing the wavelength of the exciting laser. Caution should also be taken as the laser can be destructive to some samples at high power levels.

2.2. Blue Toners

2.2.1. Materials and Methods

2.2.1.1. Instrumental Analysis

A Foster and Freeman FORAM Raman Spectral Comparator (Foster and Freeman, Worcestershire UK) outfitted with a 30mW, 785nm laser and adjustable power source was used to acquire the data. The FORAM can be run at 100%, 25%, and 10% laser power and has approximately 8cm^{-1} resolution. One hundred blue toner samples affixed to non-standardized paper were obtained from the United States Secret Service library. This library was created with the majority of samples being obtained from Original Equipment Manufacturer (OEM) printers. The toner was sampled by obtaining printouts from various printers. The printouts were assigned to the printer make and model; however, the cartridge number was not recorded. Therefore, it cannot be guaranteed that a non-OEM cartridge was used in the OEM printer.

The samples were provided as two fifty-sample sets. The first set contained various sized cutouts ranging from test strips to pictures and was used as a training set. These samples were stored in plastic page protectors numbered accordingly within a cardboard envelope. The second set contained 4mm circular hole punches taken from either a test strip or a picture. The hole punches were affixed to individual glass microscope slides using double sided tape, numbered accordingly, and stored in a microscope box. Both sets were stored away from light at a room temperature fluctuating between 20-25°C.

A method was developed to sample the toner directly on the paper substrate without interference from the paper. The sample was exposed to the 785nm laser for 15 scans at 9 seconds each with the laser power at 10%. A range from 400-2000 cm^{-1} was sampled with an approximate resolution of 8 cm^{-1}. Three replicates were taken for each sample at locations as distant from one another as possible. Four of the original samples were deemed unsuitable for analysis resulting in a final total of ninety-six samples. Figure 2.2 illustrates the lack of interference from the paper in the toner samples. One minor peak from the paper can be seen in the toner spectrum near 1100 cm^{-1}. However, this peak appears in every spectrum in the same manor allowing it to be negated in further analysis. All spectra were found to be highly reproducible within a sample as seen in Appendix A.

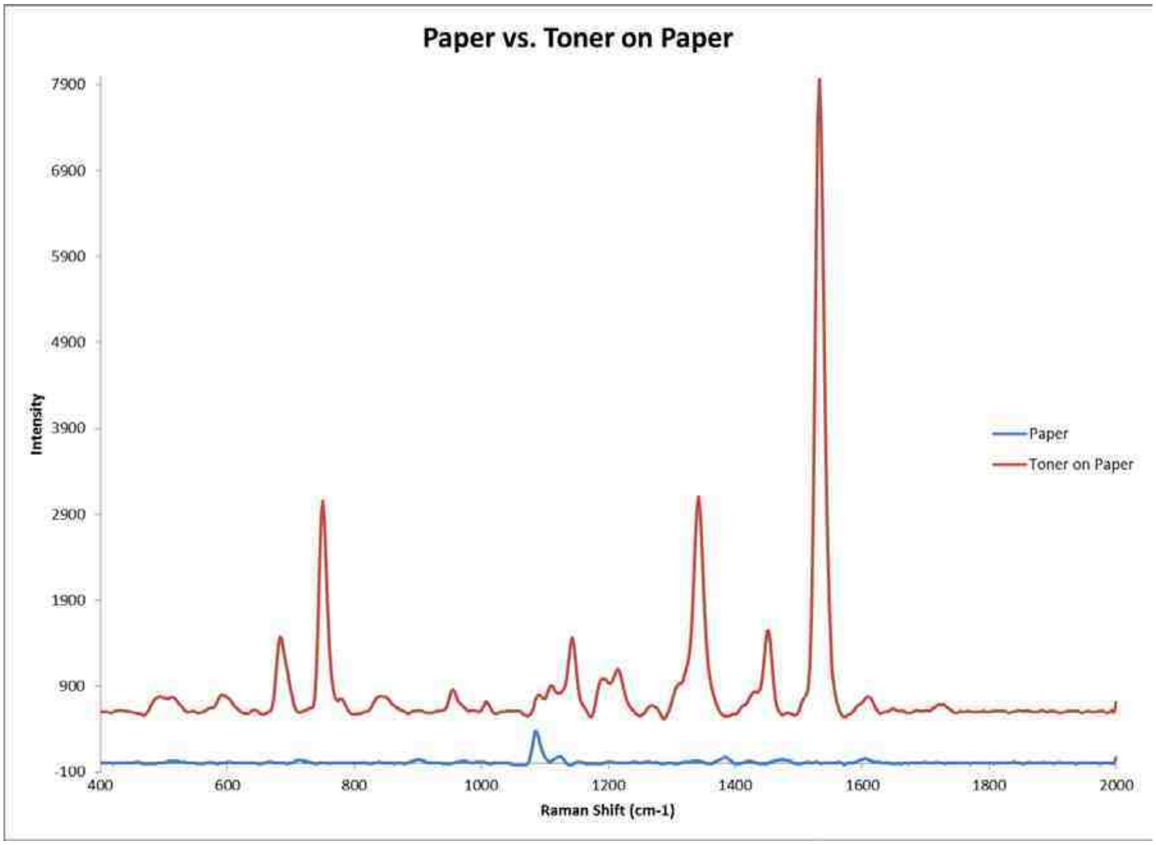


Figure 2.2. Spectral Comparison illustrating lack of paper interference in blue toner spectrum.

2.2.1.2. Data Analysis

Data Analysis was performed using FORAM and XLSTAT2010 (Addinsoft, Paris France) software. All spectra were background corrected and smoothed in the FORAM software, then exported to Excel 2010 (Microsoft Corporation, Redmond WA). Baseline subtraction was then performed using the average of the baseline between 1800 and 2000 cm^{-1}. The final step of preprocessing was normalization by the square root of the sum of squares. This data was then analyzed by PCA and AHC in XLSTAT2010. DA was then performed using the data from PCA. Seven principle components were used based on the scree plot. Classes were then assigned to each sample using the observations plot from PCA.

2.2.2. Results and Discussion

2.2.2.1. Statistical Results

The PCA observations plot for blue toners analyzed by Raman spectroscopy is shown in Figure 2.3 below. PCA analysis indicates three main classes based on plotting the observations according to the first two principal components, which account for 39.36% of the total variance in the sample set. Further separation could occur when more principal components are examined in additional dimensions. However, the use of additional dimensions does not provide any clear separation that would allow all three replicates to remain in the same class. This is necessary to account for intra-sample variability, therefore a conservative three class system was chosen.

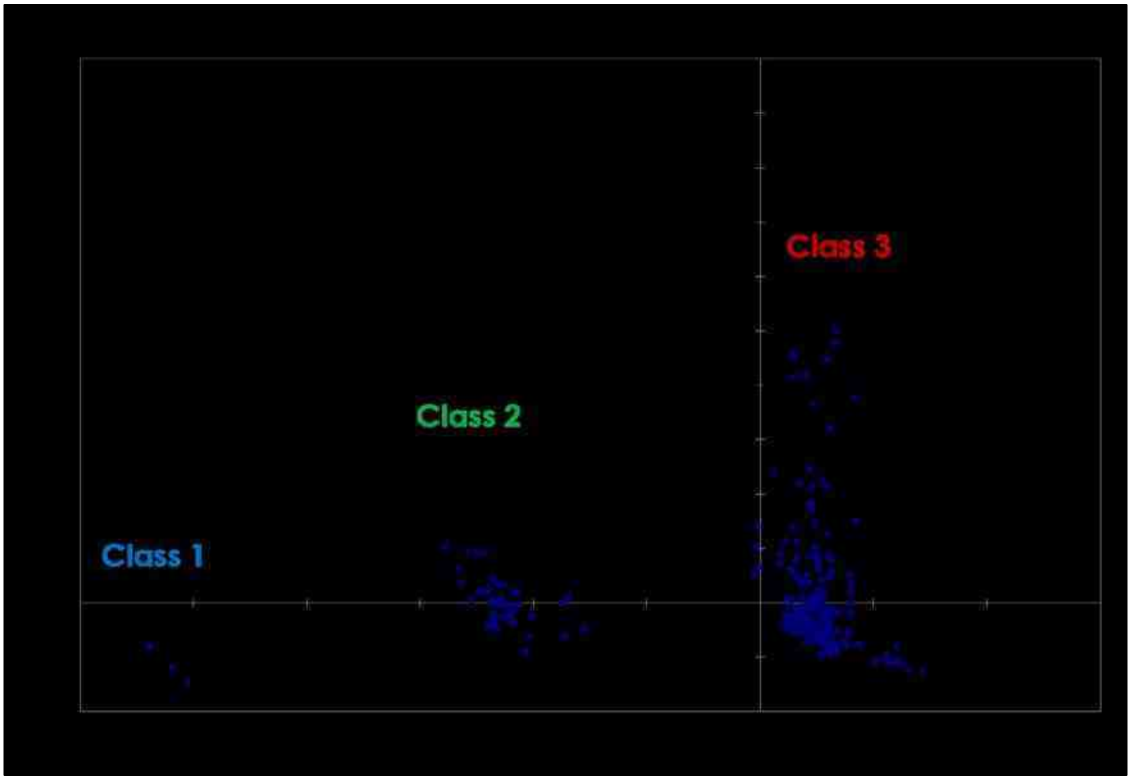


Figure 2.3. Observations plot from PCA for the blue toners with three groups/classes shown.

DA was performed using the data gained from PCA. Table 2.1 shows the eigenvalues relevant to this study. A number of principal components had to be selected to perform DA. To determine the appropriate number, the scree plot shown in Figure 2.4 was utilized. Seven principal components with an approximate variance of 70% were decided upon. To meet or exceed 95% cumulative variance, 52 principal components would have been required. The Kaiser criterion also resulted in the need for 52 principal components.

Table 2.1. Eigenvalues and variability associated with each principal component (PC).

Principal Component	Eigenvalue	Variability (%)	Cumulative (%)
PC1 (F1)	142.322	22.203	22.203
PC2 (F2)	109.991	17.159	39.362
PC3 (F3)	62.786	9.795	49.158
PC4 (F4)	50.829	7.930	57.087
PC5 (F5)	41.901	6.537	63.624
PC6 (F6)	24.093	3.759	67.383
PC7 (F7)	16.051	2.504	69.887
PC8 (F8)	14.529	2.267	72.153
PC9 (F9)	12.567	1.961	74.114
PC10 (F10)	11.450	1.786	75.900
PC11 (F11)	9.078	1.416	77.316
PC12 (F12)	8.381	1.308	78.624
PC13 (F13)	7.993	1.247	79.871
PC14 (F14)	6.778	1.057	80.928
PC15 (F15)	6.023	0.940	81.868
PC16 (F16)	5.304	0.827	82.695
PC17 (F17)	5.064	0.790	83.485
PC18 (F18)	4.831	0.754	84.239
PC19 (F19)	4.240	0.661	84.900
PC20 (F20)	3.792	0.592	85.492
PC21 (F21)	3.683	0.575	86.067

PC 22 (F22)	3.414	0.533	86.599
PC23 (F23)	3.251	0.507	87.106
PC24 (F24)	3.056	0.477	87.583
PC25 (F25)	2.906	0.453	88.036
PC26 (F26)	2.850	0.445	88.481
PC27 (F27)	2.679	0.418	88.899
PC28 (F28)	2.431	0.379	89.278
PC29 (F29)	2.349	0.366	89.645
PC30 (F30)	2.283	0.356	90.001
PC31 (F31)	2.128	0.332	90.333
PC32 (F32)	2.001	0.312	90.645
PC33 (F33)	1.959	0.306	90.951
PC34 (F34)	1.889	0.295	91.245
PC35 (F35)	1.851	0.289	91.534
PC36 (F36)	1.800	0.281	91.815
PC37 (F37)	1.689	0.264	92.078
PC38 (F38)	1.656	0.258	92.337
PC39 (F39)	1.563	0.244	92.581
PC40 (F40)	1.526	0.238	92.819
PC41 (F41)	1.485	0.232	93.050
PC42 (F42)	1.450	0.226	93.277
PC43 (F43)	1.373	0.214	93.491
PC44 (F44)	1.352	0.211	93.702

PC45 (F45)	1.251	0.195	93.897
PC46 (F46)	1.220	0.190	94.087
PC47 (F47)	1.175	0.183	94.271
PC48 (F48)	1.127	0.176	94.446
PC49 (F49)	1.101	0.172	94.618
PC50 (F50)	1.047	0.163	94.781
PC51 (F51)	1.014	0.158	94.940
PC52 (F52)	0.990	0.155	95.094

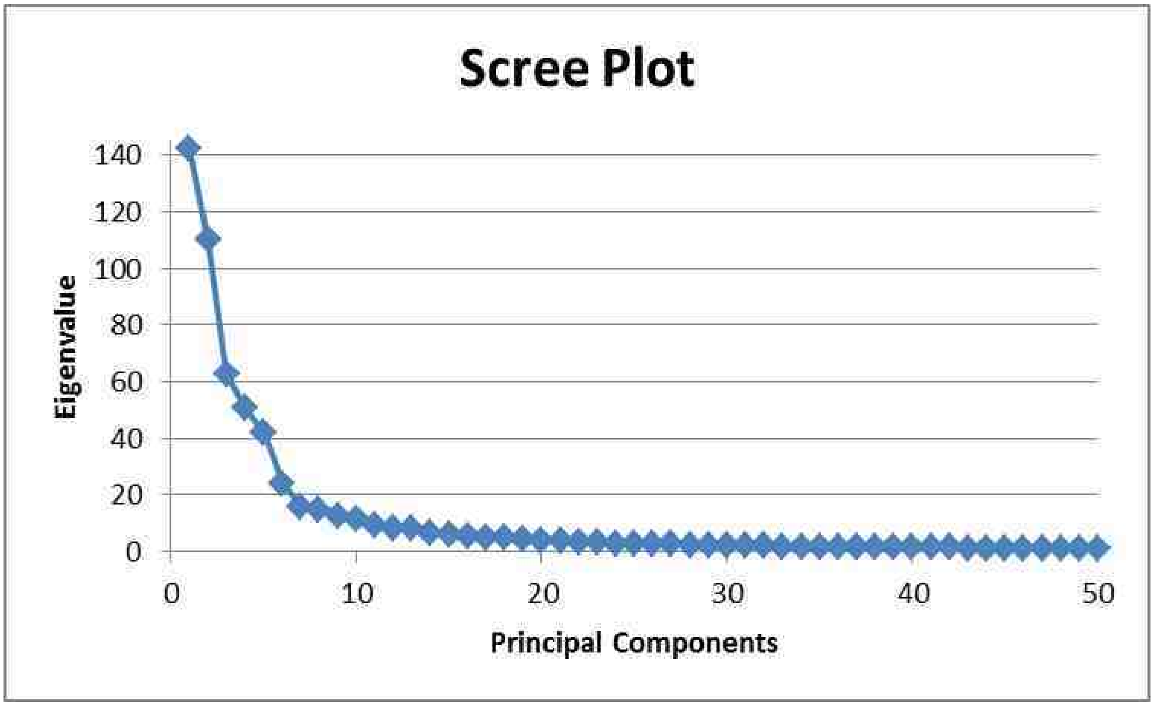


Figure 2.4. Scree plot of principal component factor scores F1-F50 for blue toners.

Figure 2.5 shows the results of DA using the first seven principal components and three classes based on the PCA results. DA resulted in 100% of the between class to within class variance accounted for in two dimensions. The three classes are completely separate with no overlap between the confidence ellipses. Due to this separation, the cross-validation results are exceptional. This is seen in the confusion matrix in Table 2.2. The samples located along the diagonal indicate those correctly classified, while samples outside the diagonal were incorrectly classified. Based on the three classes used, 100% of the samples were correctly classified. Some samples can be seen outside of the confidence ellipses due to additional variability that can be seen toward the lower end of the spectrum, resulting from minor paper fluorescence. These samples were not considered outliers, as this interference occurred in a region of the spectrum containing no pertinent information from the toner.

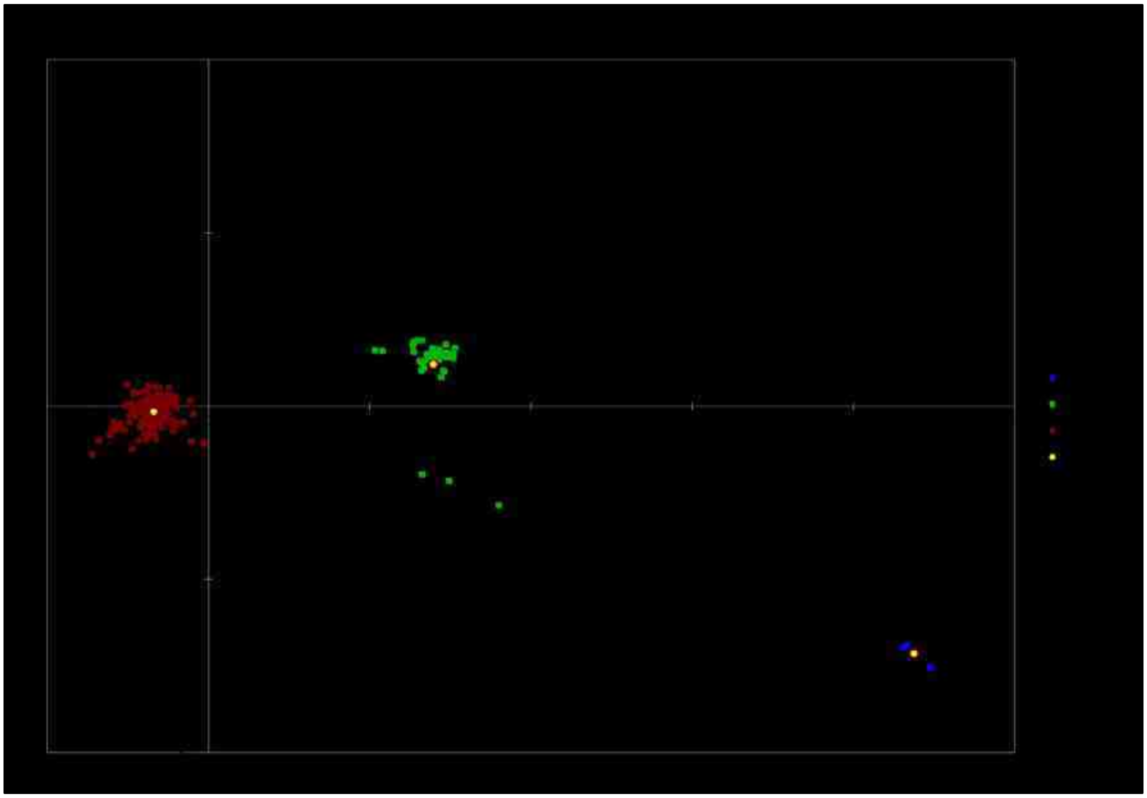


Figure 2.5. Observations plot from DA based on three classes of blue toners.

Table 2.2. Confusion matrix for cross-validation results from DA based on three classes.

From\To	1	2	3	Total	% correct
1	**3**	0	0	3	100.00%
2	0	**45**	0	45	100.00%
3	0	0	**239**	239	100.00%
Total	3	45	239	**287**	**100.00%**

2.2.2.2. External Validation

Twenty-five of the original samples were randomly selected and run a second time as a blind external validation. A random sampling tool within XLSTAT was used to select a representative sample of the original three classes. DA treated the external validation samples as supplemental data and predicted their class membership. The correct class was then determined by comparing this result to the original grouping of the sample. The results are shown in Table 2.3. 100% of the external validation samples were correctly predicted and classified. This is an exceptional result and is supported by the conservativeness of a three-class membership scheme.

Table 2.3. Confusion matrix for the external validation results of the supplemental data from DA.

From \ To	1	2	3	Total	% correct
1	**3**	0	0	3	100.00%
2	0	**12**	0	12	100.00%
3	0	0	**60**	60	100.00%
Total	3	12	60	**75**	**100.00%**

2.2.2.3. Formation of Classes

Chemometric analysis showed that there are three classes that contain reliably differentiable spectra. The next step was to determine if the manufacturer information was influencing the formation of these classes. Upon examining the printer make and model, no pattern between the classes emerged based solely on this information. Table

2.4 shows known manufacturer information by class. It is important to note that several manufacturers are interchangeable or compatible, (e.g. Canon and Kodak, Sharp and Ricoh, Konica and Minolta, and Okidata and Xerox). For example, a Xerox cartridge can be used in an Okidata printer, therefore obtaining cartridge information is important. Several manufacturers can be found in more than one class, such as Canon, Kodak, IBM, Ricoh, Lexmark, and Xerox. This could imply that manufacturer information may not be valuable for differentiating toners, or it may support the speculation that more than one composition is used within a manufacturer.

The composition of the toner cartridges was examined next. While most of the detailed information involving composition is proprietary, evidence from the MSDS sheets suggests that manufacturers use more than one composition depending on the printer. Based on what we know about Raman, the spectra were finally examined to determine if the formation of these three classes was influenced by a difference in pigment within the composition. Figure 2.6 shows the central object of each class determined by DA.

Table 2.4. Manufacturer information separated by class.

Class 1	Xerox
Class 2	Canon Kodak IBM Ricoh Lexmark Digital Colorwriter
Class 3	Canon Kodak IBM Ricoh Lexmark Sharp HP Konica Minolta Mita QMS Panasonic Xerox Toshiba Tektronix Okidata

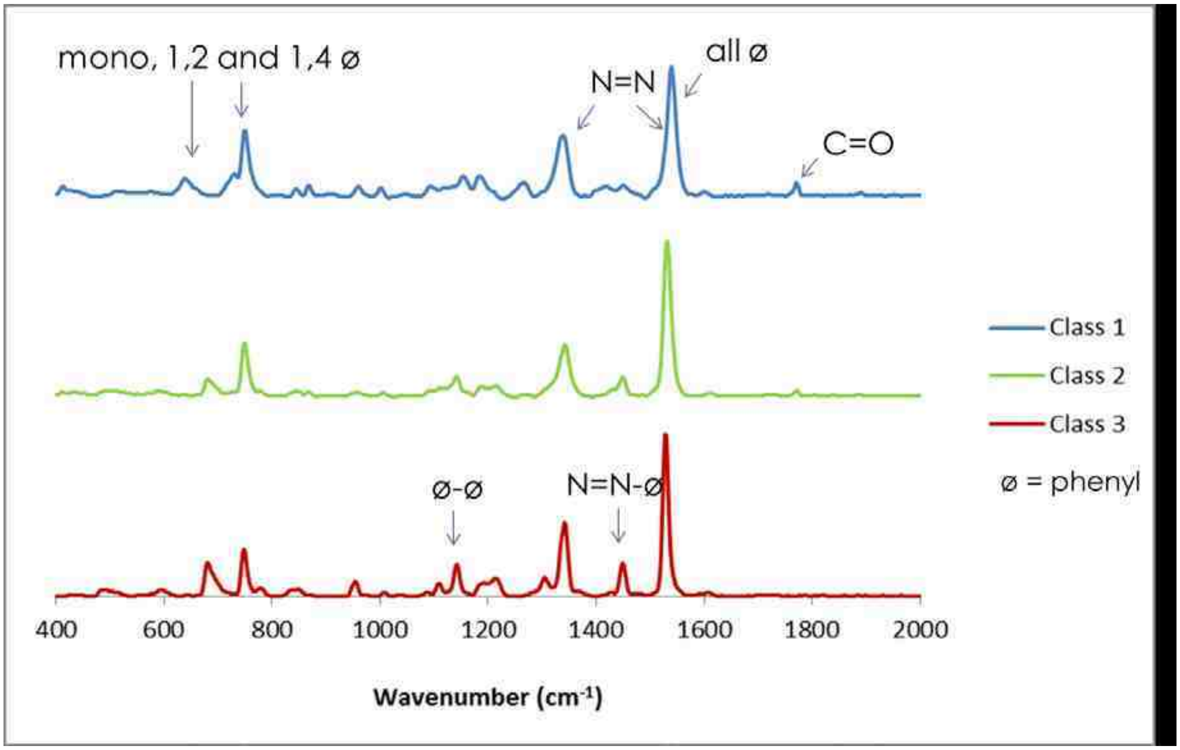


Figure 2.6. Blue toner central objects of each class determined by DA.

Copper phthalocyanine, Victoria blue, and azo pigments are all possible contributors to these three classes. All three classes contain a band that can be assigned to an azo bond (-N=N-) or phenyl group at 1560 cm^{-1}. Other characteristic bands from these three pigments can be seen in Figure 2.6 above.

2.2.3. Conclusions

Raman has proven to be an effective method of analyzing blue toner on paper. Sample preparation is minimal and nondestructive. Spectra are also highly reproducible between replicates making this a quick and effective method of analysis.

The chemometric analysis of the samples using PCA and DA resulted in significant results. Three classes formed within the sample set with high internal and external validation accuracy. Significant differences between the classes can be seen at approximately 700, 825, 950, 1100, 1300, 1450, and 1750 cm^{-1}. Analyzing the factor loadings plot provides information as to which spectral regions are contributing the most to each principal component and thus, which variables are being used to discriminate the classes.

2.3. Yellow Toners

2.3.1. Materials and Methods

2.3.1.1. Instrumental Analysis

Sampling and instrumental conditions were the same as described for the blue toners in section 2.2.1.1. A method was developed to sample the toner directly on the paper substrate. The sample was exposed to the 785nm laser for 15 scans at 10 seconds each with the laser power at 100%. A range from 400-2000 cm^{-1} was sampled with an approximate resolution of 8 cm^{-1}. Three replicates were taken for each sample at locations as distant from one another as possible. Due to interference from the paper, a blank area on the paper, front or back, was also sampled using the same parameters and method described above.

2.3.1.2. Data Analysis

Data Analysis was performed using FORAM software and XLSTAT2010 Addinsoft Software (Addinsoft, Paris France). All spectra were background corrected and smoothed in the FORAM software, then exported to Excel 2010 (Microsoft Corporation, Redmond WA). Baseline subtraction was then performed using the average of the baseline between 1830 and 1900 cm^{-1}. Spectral subtraction was then performed using $R=A-X*B$ where A is the yellow toner data point, B is the paper data point, and X is the ratio of the paper peak to toner peak at 515 cm^{-1}. This peak was determined to be solely contributed by the paper in both spectra. Some negative peaks were introduced into the spectrum as a result of this subtraction; therefore, the coefficient (X) was manually changed to the point of least negative contribution. Figure 2.7 illustrates the substantial paper interference making spectral subtraction necessary. An example of the spectrally subtracted data can be seen in Figure 2.8. The resultant data was normalized by the square root of the sum of squares to prepare for chemometric analysis. This data was then analyzed by PCA and AHC in XLSTAT. DA was then performed using the data from both PCA and AHC. Five principle components were used based on the scree plot. Classes were then assigned to each sample using the pre-assigned classes from AHC.

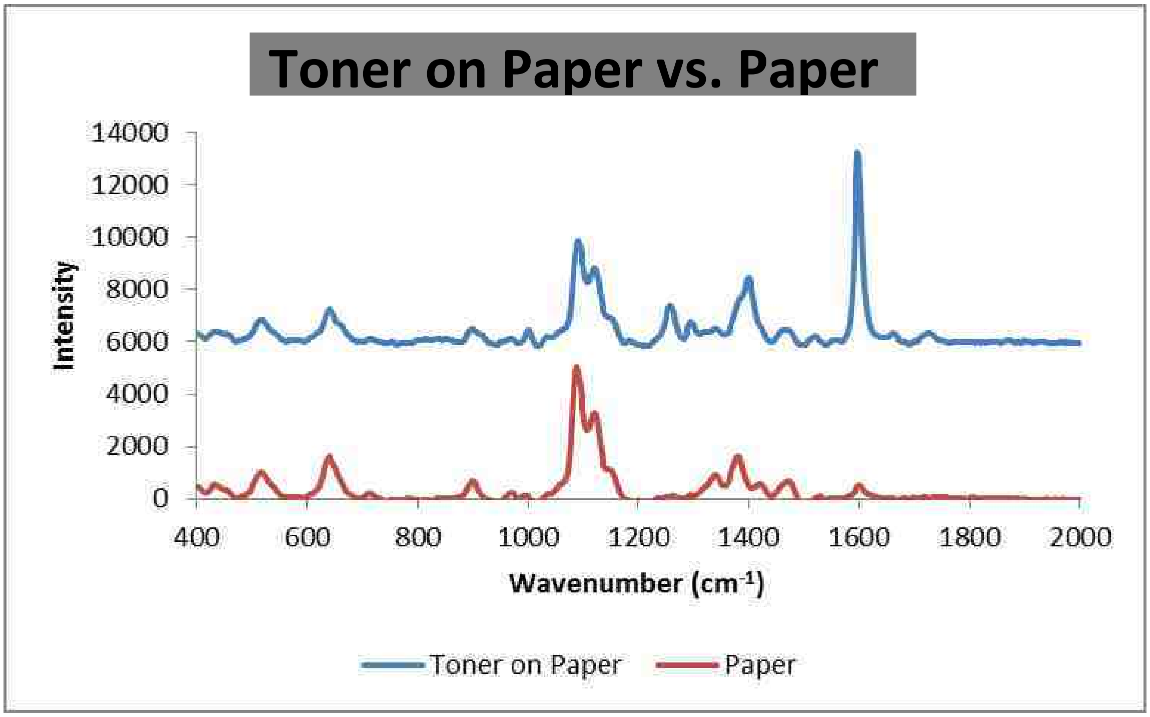


Figure 2.7. Spectral comparison illustrating the interference of paper in the yellow toner spectrum.

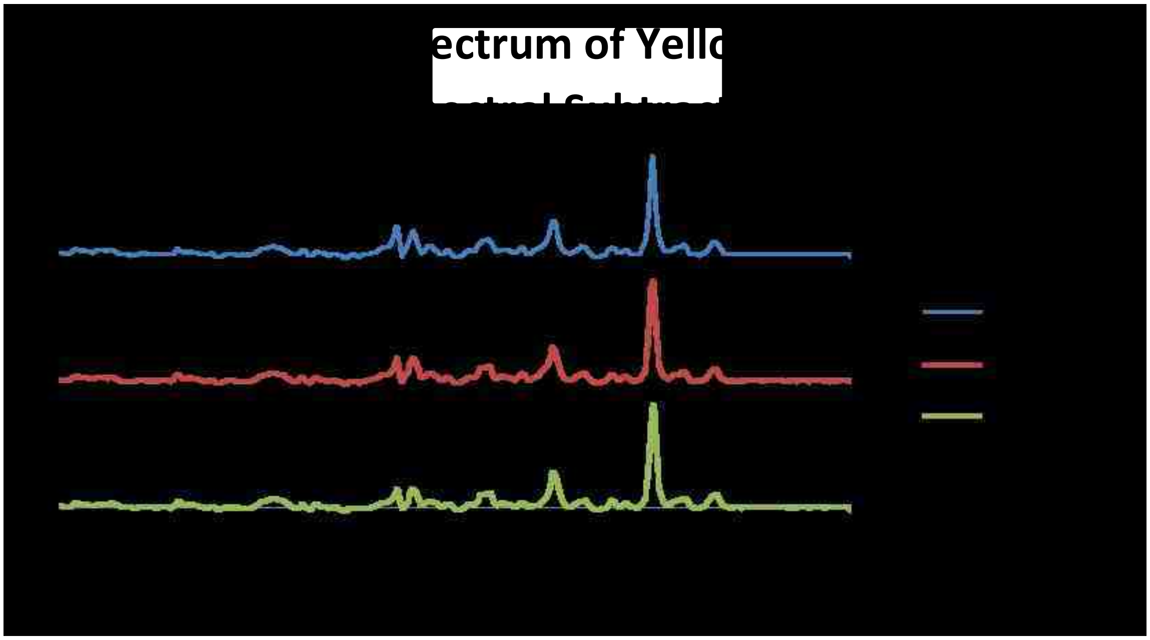


Figure 2.8. Example of a yellow toner resultant spectrum after spectral subtraction of paper.

2.3.2. Results and Discussion

2.3.2.1. Statistical Results

The AHC dendrogram for yellow toners on paper analyzed by Raman spectroscopy is shown in Figure 2.9 below. AHC analysis suggests that there are four classes based on the location of the truncation line, as determined by a histogram of node positions. Divisions at nodes to the right of the truncation line are most significant in establishing the number of classes. AHC was performed on the spectrally subtracted data for each yellow toner sample. The spectra of the central objects from AHC, shown in Figure 2.10, show several distinct differences between each class' spectra.

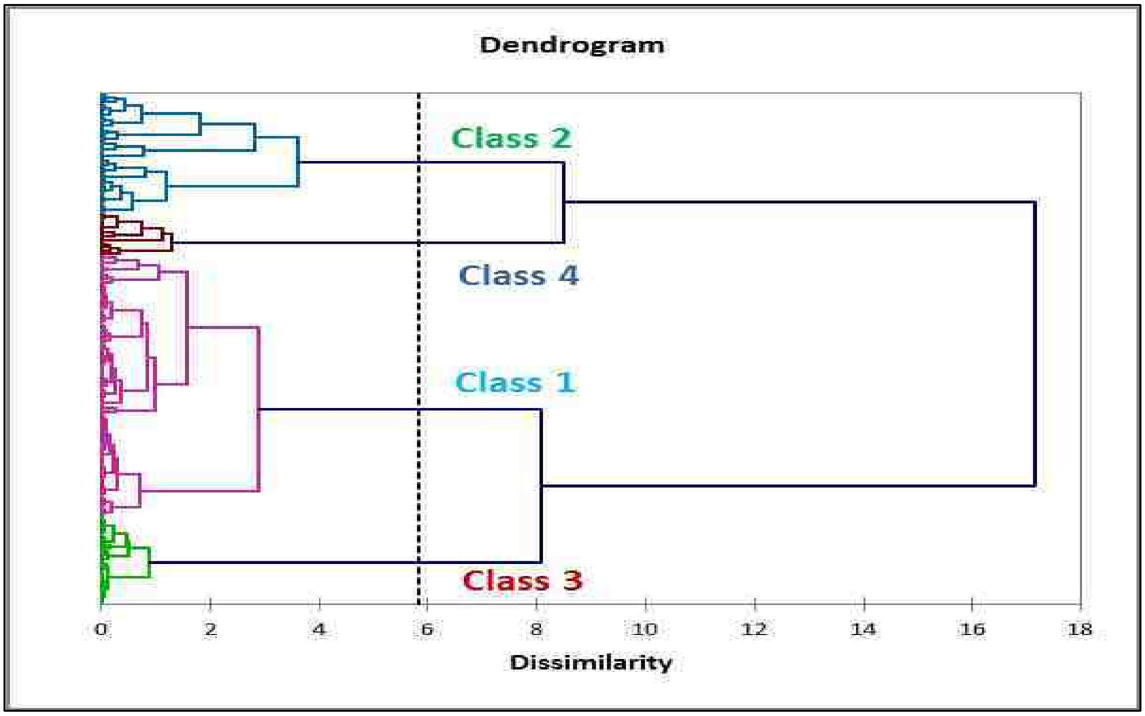


Figure 2.9. Dendrogram from AHC of yellow toner replicates forming four classes.

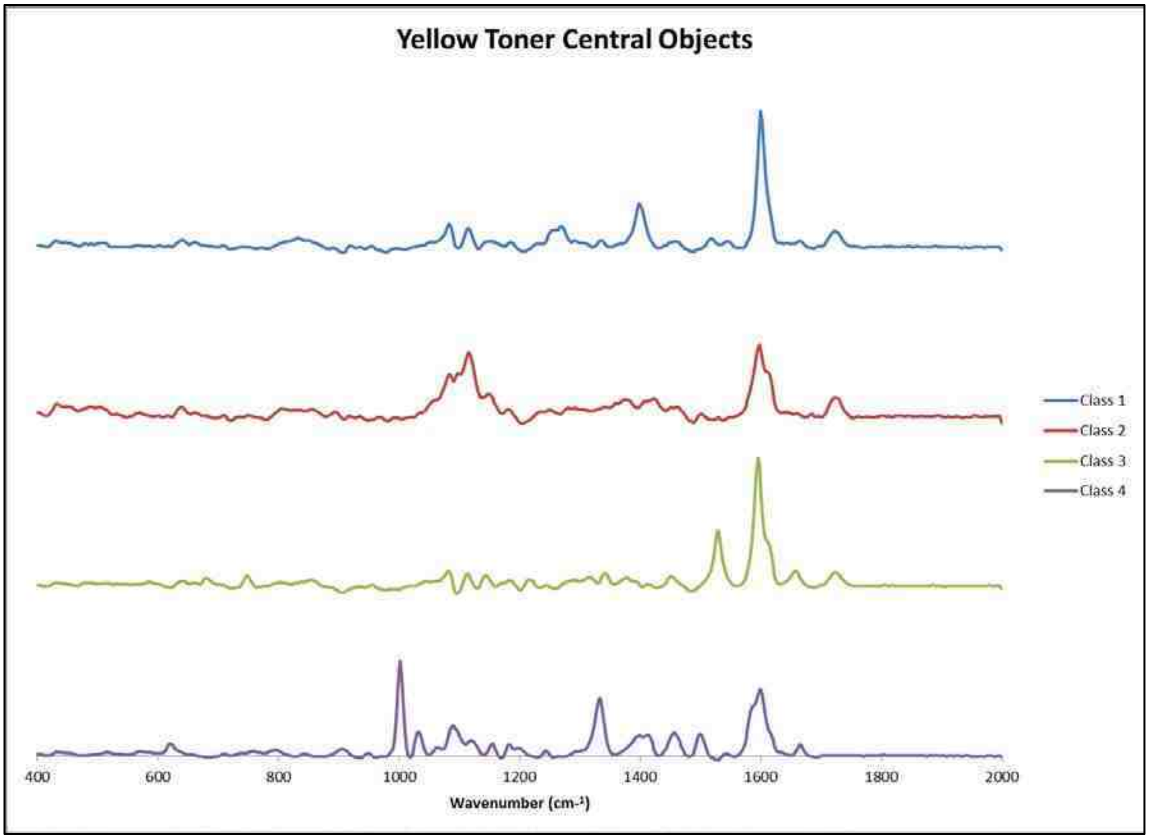


Figure 2.10. Central objects of the four yellow classes as determined by AHC.

The spectrally subtracted data was analyzed by PCA followed by DA. The observations plot from PCA yielded little useful information in determining classes; therefore, AHC was solely used for a classification scheme. A number of principal components had to be selected to perform DA. To determine the appropriate number, the scree plot shown in Figure 2.11 was utilized. Five principal components with a cumulative variance of 64% were selected. To meet or exceed 95% cumulative variance, 46 principal components would have been required. The Kaiser criterion would have resulted in the need for 52 principal components.

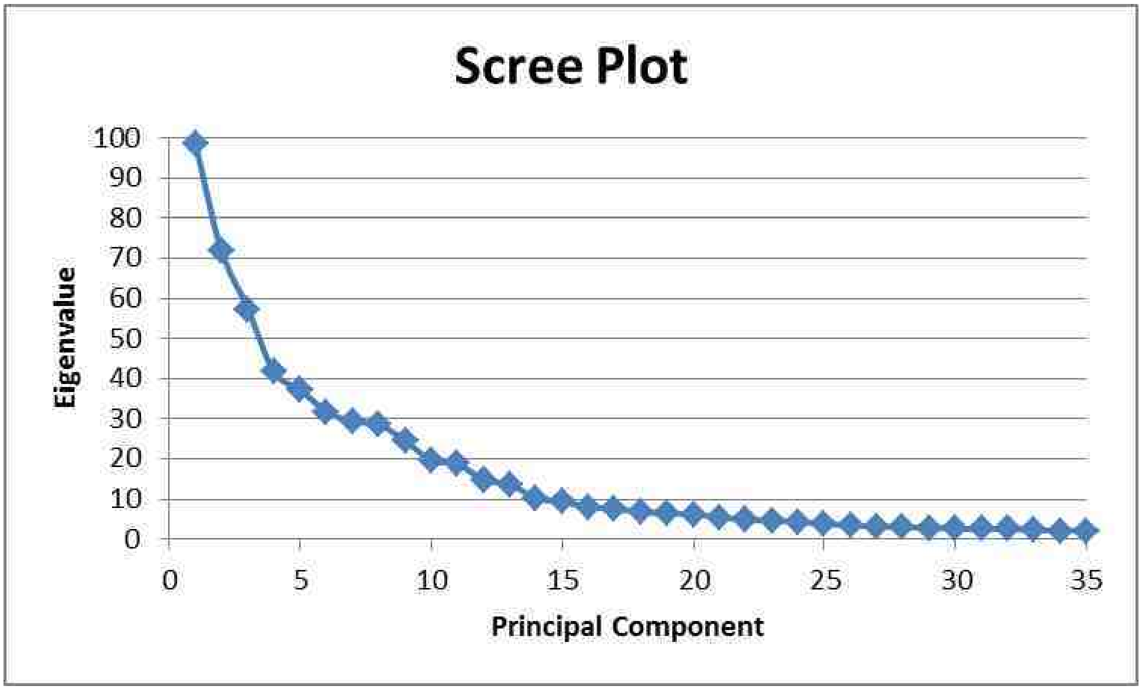


Figure 2.11. Scree plot of principal component factor scores F1-F35 for yellow toners.

Figure 2.12 shows the results of DA using the first four principal components and four classes based on the PCA and AHC results. DA resulted in 83.05% of the between class to within class variance accounted for in two dimensions. Class four is completely separated from the other three classes; however, the overlap of the confidence ellipses of class one, two, and three resulted in a slightly lower cross-validation result. This is seen in the confusion matrix in Table 2.5. The samples located along the diagonal indicate those correctly classified, while samples outside the diagonal were incorrectly classified. Based on the four classes used, 95.45% of the samples were correctly classified.

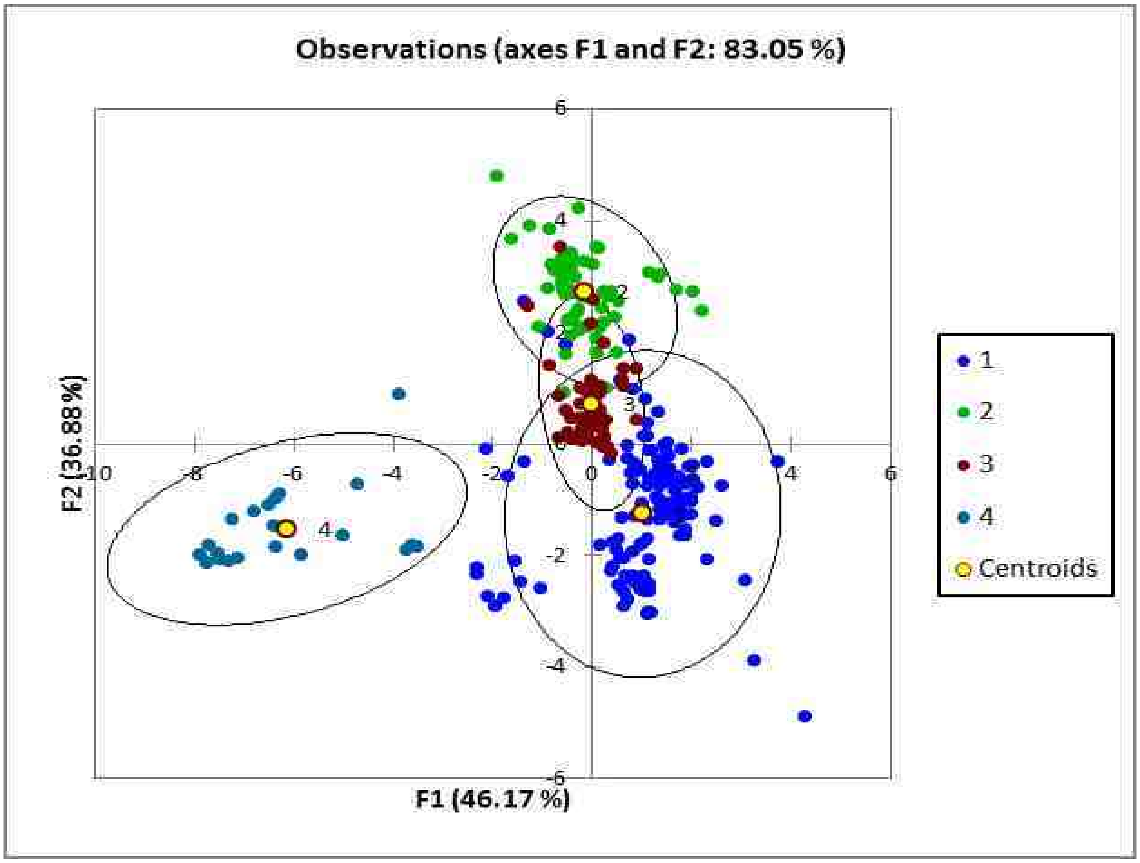


Figure 2.12. Observations plot from DA based on four classes of yellow toners.

Table 2.5. Confusion matrix for cross-validation results from DA based on four classes.

From \ To	1	2	3	4	Total	% correct
1	129	6	0	0	135	95.56%
2	0	59	3	0	62	95.16%
3	0	3	43	0	46	93.48%
4	0	0	0	21	21	100.00%
Total	129	68	46	21	**264**	**95.45%**

2.3.2.2. External Validation

Twenty-five of the original samples were randomly selected and run a second time as a blind external validation. A random sampling tool within XLSTAT was used to select a representative sample of the original three classes. DA treated the external validation samples as supplemental data and predicted which classes they would be grouped into. The correct class was then determined by the grouping of the original sample. The results are shown in Table 2.6. 72% of the external validation samples were correctly predicted and classified. The only samples incorrectly classified came from group two, skewing the overall prediction accuracy. Regardless, this is a poor result, making this model unsuitable for accurately discriminating yellow toners.

Table 2.6. Confusion matrix for the external validation results of the supplemental data from DA.

From \ To	1	2	3	4	Total	% correct
1	**42**	0	0	0	42	100.00%
2	**7**	**3**	**14**	0	24	12.50%
3	0	0	**6**	0	6	100.00%
4	0	0	0	**3**	3	100.00%
Total	49	3	20	**3**	**75**	**72.00%**

2.3.3. Conclusions

Raman analysis within the above parameters has proven to be an ineffective method of analyzing yellow toner on paper. While sample preparation is minimal and nondestructive, the spectra obtained were not highly differentiable when subjected to chemometric analysis. However, the spectra did prove to be highly reproducible between replicates making this a quick and effective method of analysis. An adjustment to the method's laser wavelength could limit the paper interference and increase the signal response providing a more accurate statistical analysis. This optimization is discussed in further detail in Chapter 4.

The chemometric analysis of the samples using PCA, AHC, and DA resulted in non-significant results. Four classes formed within the sample set with high internal validation accuracy but poor external validation accuracy. The paper proved to have too much interference in the spectrum producing confusing and uninterpretable results upon spectral subtraction and subsequent statistical analysis.

2.4. Magenta Toners

2.4.1. Materials and Methods

Sampling and instrumental conditions were the same as described for the blue toners in section 2.2.1.1. A method was developed and optimized to sample the toner directly on the paper substrate. The sample was exposed to the 785nm laser for 35 scans at 10 seconds each with the laser power at 25%. A range from 400-2000 cm^{-1} was sampled with an approximate resolution of 8 cm^{-1}. Three replicates were taken for each sample at locations as distant from one another as possible. Due to a weak signal and interference from the paper, no usable data could be acquired with these parameters.

2.4.2. Conclusions

Raman analysis within the above parameters is ineffective in sampling and discriminating magenta toners on paper. The weak signal from the toner is overwhelmed by the paper signal in the spectrum. This is not surprising based on the wavelength of the laser. Resonance Raman could prove to be a more effective technique for these samples. Raman, in general, is a quick, non-destructive sampling techniques requiring little to no sample preparation. A minor adjustment in laser wavelength with optimization of the method described above could result in a stronger toner signal and subsequent discrimination using chemometric analysis.

CHAPTER 3. OVERALL CONCLUSIONS OF THE STUDY

Developing a method using Raman Spectroscopy that limited the interference of paper was successful for the blue toners; however, it was less successful for the yellow toners and completely unsuccessful for the magenta toners. Minor adjustments to the parameters may yield more successful results and will be discussed in further detail in the following chapter. Chemometric analysis also yielded valuable data for the blue toners suggesting a three class schematic. After all the data was collected and analyzed in this study, it was important to determine what the individual classes represent. Previous research has already suggested that the pigments within the toner are responsible for the Raman signal, as opposed to the polymer resins which are seen in infrared spectroscopy.[1] As previously stated, Victoria blue, copper phthalocyanine, and azo pigments and dyes are all possible contributors to these three classes. Figure 3.1 shows the chemical structures of these pigments and dyes. As for the yellow toners, while the paper interference was much more substantial, making accurate class determination difficult, possible contributors to these classes are acetamide and azo pigments. Figure 3.2 shows the chemical structure of some of these pigments. Further, more destructive, studies are needed to determine which pigments and dyes are truly present, including others that are not mentioned here.

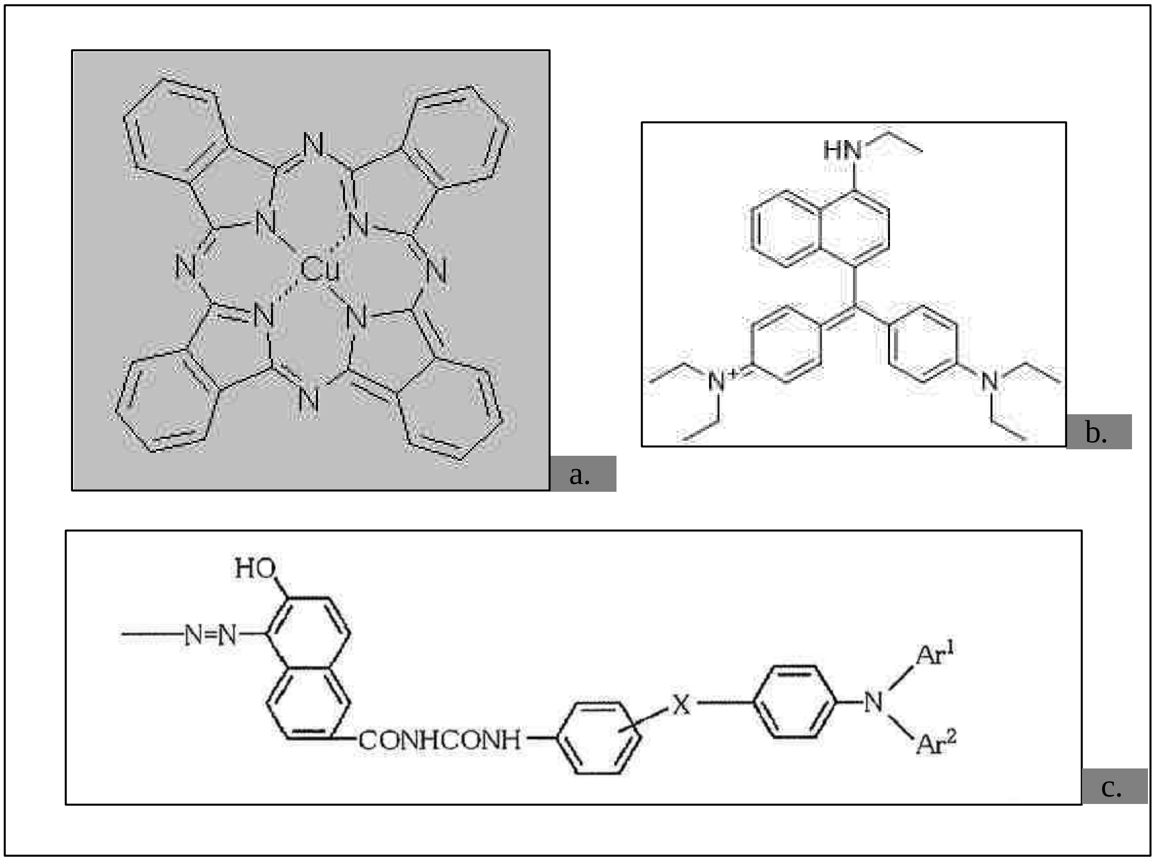
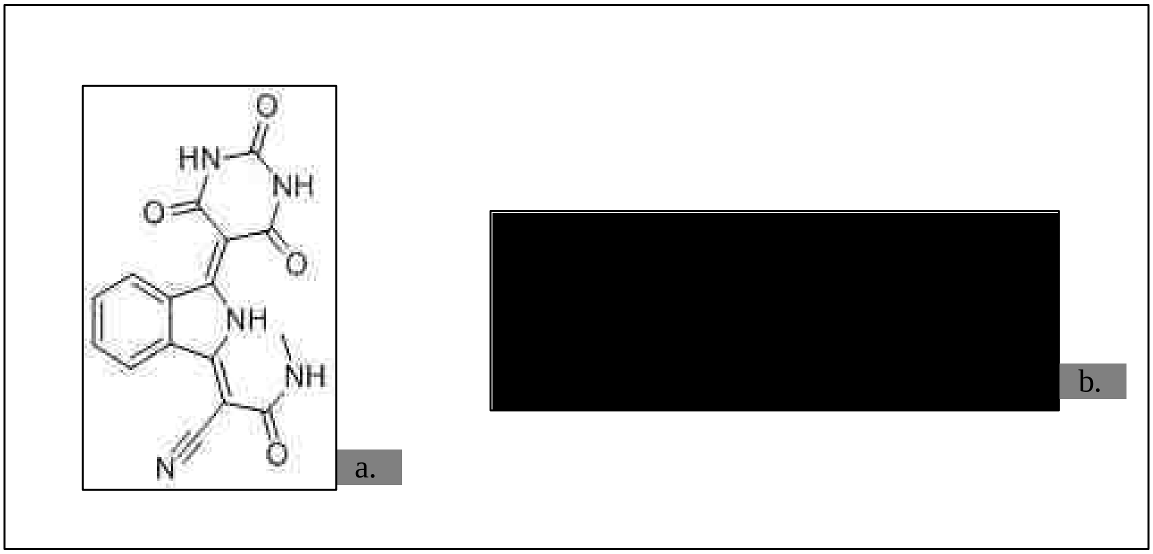

N
Cu
HN
N+
N
HO
N=N
CONHCONH
X
Ar1
Ar2
a.
b.
c.

Figure 3.1. Blue pigments (a). copper phthalocyanine, (b). victoria blue, (c). azo-pigment.


O
HN
NH
NH
NH
N
a.
b.

[redacted]

Figure 3.2. Yellow pigments (a). acetamide, (b). azo-pigment.

The research conducted in this work demonstrates the discrimination ability chemometrics can provide to a large varied data set. While no correlation between printer make, model, and cartridge number and toner could be determined, this instrumental technique can still be proven valuable in forensic case work. An examiner may not be able to trace a particular toner to its source, but this analysis could assist in question versus known examinations allowing exclusions, especially when combined with a complementary technique such as IR.

In a real world setting, a forensic document examiner would receive a questioned document and list of suspected printers. The examiner would then collect a test strip from each suspected printer. Each document would then be examined under a microscope to determine the presence of colored portions of the document. Figure 3.3 illustrates the separation of colors that can be seen from the individual color toner particles as they were melted to the paper. Raman spectra can be collected from each of these colored areas on each document. The samples can then be treated as a prediction set and analyzed using the already established chemometric models to determine class. The classification information gained for the unknown and knowns can then allow elimination or inclusion of each sample. If more colors are present in the document, it will provide more discriminating information to allow exclusion of a known. Table 3.1 illustrates the various combinations based on the blue and yellow classification information. A total of twelve combinations can be seen using only two toners. The addition of the third toner (i.e. magenta) would only increase the number of possible combinations, thus increasing the discriminating power of this technique. This method provides a quick, non-destructive way to narrow the number of samples during an analysis before continuing onto other more time consuming and destructive techniques.

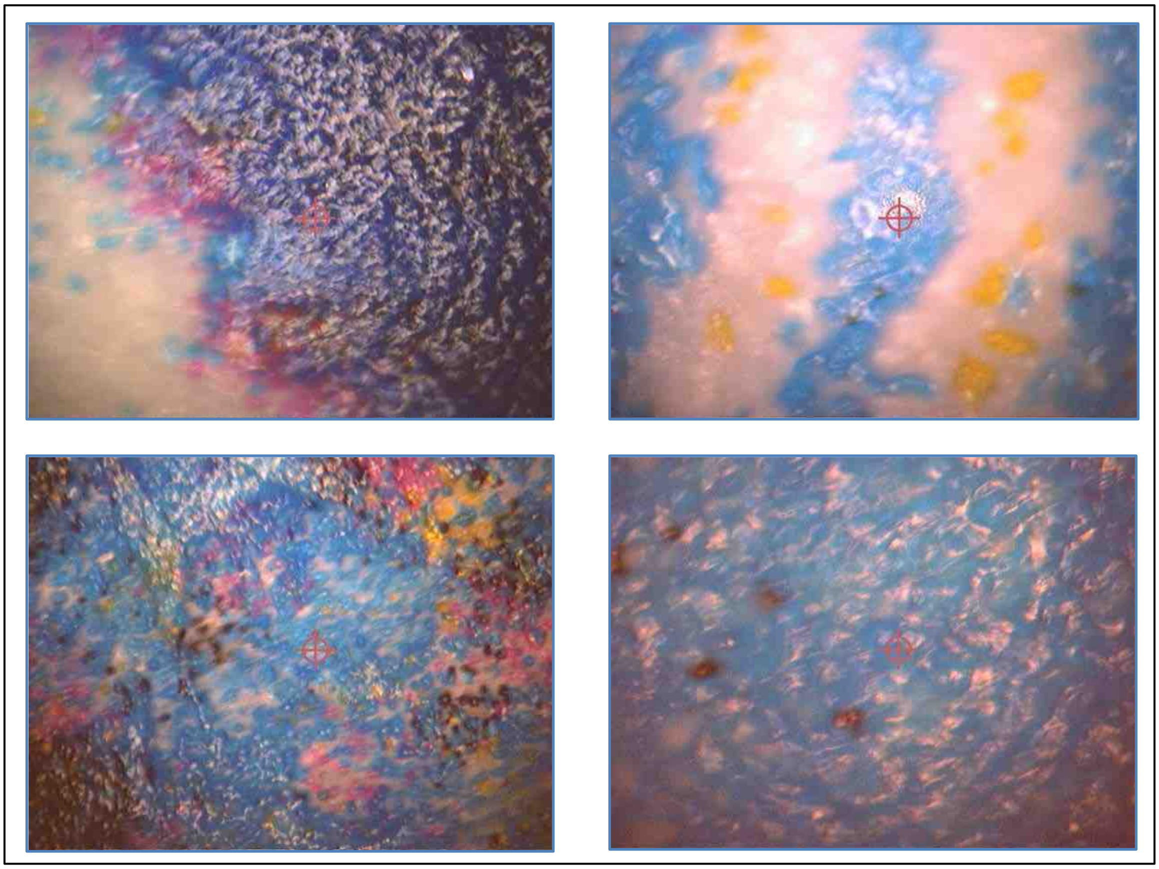
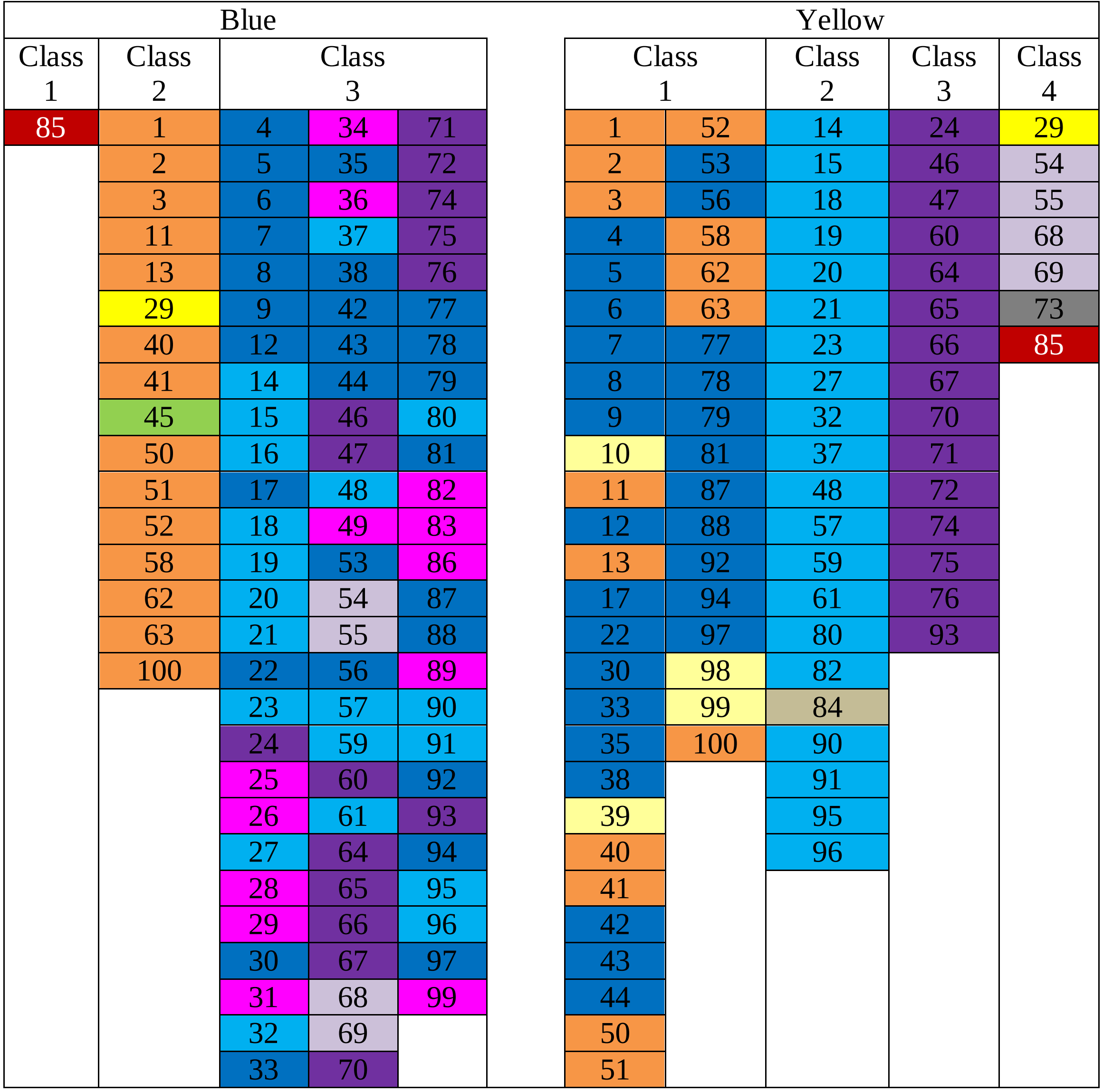


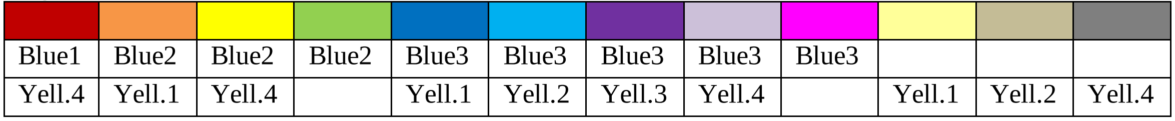
Figure 3.3. Micro pictograph of colored documents under 20X magnification.

Table 3.1. Discrimination of samples using combined class information for each sample’s results utilizing two colors.



Blue					Yellow				
Class 1	Class 2	Class 3			Class 1		Class 2	Class 3	Class 4
85	1	4	34	71	1	52	14	24	29
	2	5	35	72	2	53	15	46	54
	3	6	36	74	3	56	18	47	55
	11	7	37	75	4	58	19	60	68
	13	8	38	76	5	62	20	64	69
	29	9	42	77	6	63	21	65	73
	40	12	43	78	7	77	23	66	85
	41	14	44	79	8	78	27	67	
	45	15	46	80	9	79	32	70	
	50	16	47	81	10	81	37	71	
	51	17	48	82	11	87	48	72	
	52	18	49	83	12	88	57	74	
	58	19	53	86	13	92	59	75	
	62	20	54	87	17	94	61	76	
	63	21	55	88	22	97	80	93	
	100	22	56	89	30	98	82		
		23	57	90	33	99	84		
		24	59	91	35	100	90		
		25	60	92	38		91		
		26	61	93	39		95		
		27	64	94	40		96		
		28	65	95	41				
		29	66	96	42				
		30	67	97	43				
		31	68	99	44				
		32	69		50				
		33	70		51				

Legend:



🟥	🟧	🟨	🟩	🟦	🟦	🟪	🟪	🟪	🟨	🟫	⬛
Blue1	Blue2	Blue2	Blue2	Blue3	Blue3	Blue3	Blue3	Blue3			
Yell.4	Yell.1	Yell.4		Yell.1	Yell.2	Yell.3	Yell.4		Yell.1	Yell.2	Yell.4

CHAPTER 4. FUTURE RESEARCH AND RECOMMENDATIONS

This work has laid the groundwork for future research in this area. Due to some of the shortcomings (i.e., lack of cartridge information, unavailability of various laser wavelengths, and lack of proportional representation of samples within each manufacturer), this work can be treated as a preliminary characterization. The first recommendation is to acquire a data set with known history and parameters. Make, model, and cartridge number should be recorded for each sample. In addition to a printed test strip, the pure toner should be collected for comparison purposes. Next, each sample should be collected on a variety of common printer and copier papers to account for any differences that may occur between paper types. Multiple papers from the same batch and from varying batches for each paper type should also be used. The laser printers chosen should contain a variety of manufacturers, cover a range of products within each manufacturer, and include the most common printer types. Care should be taken to include an appropriate and consistent amount of samples from each manufacturer. Information should also be gathered on the fusing temperature used in each laser printer and how this affects compatible toner cartridges used on that respective printer, which are known to vary in composition. Changing each of these parameters allows one to analyze and take into account real world variation.

The second recommendation would be to use a different laser to collect spectra from the yellow and magenta toners. Since we are evaluating the different pigments used in toners, resonance Raman should be employed. Resonance Raman can amplify the signal intensity by 10^6 and allows for selectivity of the pigments and dyes by selecting a laser wavelength at or approaching the absorption line of the analyte. For yellow toners, a 457nm or 473nm blue laser should be used. For magenta toners, a 488nm or 514nm blue/green laser should be used. A tunable dye laser that allows the analyst to change the specific wavelength would be more practical and beneficial in this type of analysis. In addition, collecting reflectance spectra of each sample would assist in determining the true maximum absorbance and thus allow a more accurate laser wavelength selection to optimize results. This also provides an additional variable for discrimination. The main priority is to use a method that is non-destructive to the document, preserving its original state. However, fluorescence can be a problem when using lasers approaching the lower end of the visible range. Minor fluorescence was observed in some of our spectra in this study, this was most likely contributed by the paper in these samples. If major fluorescence proves to be a problem with the toner samples, surfaced-enhanced resonance Raman (SERRS) should be considered. Also, a 1064nm laser could be considered to avoid any fluorescence problems; however, the paper interference could possibly increase as the toner signal would be weaker than if resonance Raman was employed.

The third recommendation would involve extending the project to create a database of various currencies around the world. Counterfeiting currency is a major issue with new techniques constantly surfacing. Some of the latest counterfeiting techniques have involved using laser printers. The ultimate goal would be to compile a database of all the current printer toners on the market and be able to trace them back to a source printer. However, this may be unpractical and difficult to maintain as new products are constantly being created. Rather, composing a database of all the existing currencies from various countries would allow for an easy detection method of counterfeit bills. This would not require obtaining the specific inks, as this information is tightly controlled by each individual country. The database would be composed of solely the spectra obtained from a range of banknotes or dollar bills to allow for a quick screening method of counterfeit currency. Also, including the spectra of counterfeit currency that has already been recovered would assist in screening for new and different counterfeit bills on the market. However, in building this database, any variation from bill to bill and variation that may occur over time should be studied.

Any future analysis on laser printer toners should take into consideration the degradation of samples that may occur from a variety of sources. As most real world samples do not occur in perfect conditions, studying the effects degradation has on both the toner and paper is necessary.

LIST OF REFERENCES

LIST OF REFERENCES

1. Foster + Freeman, Inc. Discriminating Toners with the FORAM (Application Note 02). Foster + Freeman: Sterling, VA, 2008. WEB: http://www.fosterfreeman.com/.

2. Bopp, L. C.; Horn, R.; Hopkins, G.; McCaulley, C.; Nagata, S.; Consulting, T. W. S. S., Life-Cycle Impact of Toner and Ink for CU-Boulder.

3. Totty, R. N., The examination of photocopy documents. *Forensic Science International* 1990, *46* (1), 121-126.

4. Totty, R., Analysis and differentiation of photocopy toners. *Forensic Sci. Rev* 1990, *2* (1), 1-23.

5. Mazzella, W.; Roux, C.; Lennard, C., The computer-assisted identification of colour photocopiers. *Science and Justice* 1995, *35* (2), 117-126.

6. Shiver, F.; Nelson, L., Nondestructive differentiation of full-color photocopies. *Journal of Forensic Sciences* 1991, *36* (1), 145-152.

7. Egan, W. J.; Morgan, S. L.; Bartick, E. G.; Merrill, R. A.; Taylor, H. J., 3rd, Forensic discrimination of photocopy and printer toners II. Discriminant analysis applied to infrared reflection-absorption spectroscopy. *Analytical and Bioanalytical Chemistry* 2003, *376* (8), 1279-85.

8. Suzuki, S.; Suzuki, Y.; Higashikawa, Y.; Kishi, T.; Mammo, Y., Analysis and identification of photocopied toners. Proceedings of the international symposium on the forensic aspects of trace evidence, FBI Academy, Quantico, VA, June 24--28, 1991:219-20.

9. Heuser, H. G., Methods concerning classification and identification of photocopies. Proceedings of the international symposium on questioned documents, FBI Academy, Quantico, VA, July 30-August 1, 1985:145-57.

10. Levy, E. J.; Wampler, T. P., Pyrolysis GC/MS of inks and paper in document identification. Proceedings of the international symposium on questioned documents, FBI Academy, Quantico, VA, July 30-Aug 1, 1985:141-43.

11. Levy, E.; Wampler, T., Applications of pyrolysis gas chromatography–mass spectrometry to toner materials from photocopiers. *Journal of Forensic Sciences* 1986, *31* (1), 258-71.

12. Munson, T. O., A simple method for sampling photocopy toners for examination by pyrolysis gas chromatography. *International Journal of Forensic Document Examiners* 1999, *5*, 3.

13. Munson, T., The classification of photocopies by pyrolysis gas chromatography-mass spectrometry. *Journal of Forensic Sciences* 1989, *34* (2), 352-65.

14. Chang, W. T.; Huang, C. W.; Giang, Y., An improvement on pyrolysis gas chromatography for the differentiation of photocopy toners. *Journal of Forensic Sciences* 1993, *38*, 843-843.

15. Zimmerman, J.; Mooney, D.; Kimmett, M., Preliminary examination of machine copier toners by infrared spectrophotometry and pyrolysis gas chromatography. *Journal of Forensic Sciences* 1986, *31* (2), 489-493.

16. Mizrachi, N.; Aizenshtat, Z.; Levy, S.; Elkayam, R., Classification and identification of color photocopiers by FT-IR and GC/MS. *Journal of Forensic Sciences* 1998, *43*, 353-361.

17. Armitage, S.; Saywell, S.; Roux, C.; Lennard, C.; Greenwood, P., The analysis of forensic samples using laser micro-pyrolysis gas chromatography mass spectrometry. *Journal of Forensic Sciences* 2001, *46* (5), 1043-1052.

18. Szynkowska, M. I.; Czerski, K.; Paryjczak, T.; Parczewski, A., Ablative analysis of black and colored toners using LA-ICP-TOF-MS for the forensic discrimination of photocopy and printer toners. *Surface and Interface Analysis* 2010, *42* (5), 429-437.

19. Kemp, G. S.; Totty, R., The differentiation of toners used in photocopy processes by infrared spectroscopy. *Forensic Science International* 1983, *22* (1), 75-83.

20. Totty, R. N.; Kemp, G. S.; German, B.; Metcalfe, G. S.; Hubbleday, E., Analysis of duplicator (photocopy) toners. Proceedings of the international symposium on questioned documents, FBI Academy, Quantico, VA, July 30-Aug. 1, 1985:49-69.

21. Tănase, I. G.; Udrştioiu, E. G.; Bunaciu, A. A.; Aboul-Enein, H. Y., Infrared spectroscopy in qualitative analysis of laser printer and photocopy toner on questioned documents. *Instrumentation Science and Technology* 2009, *37* (1), 30-39.

22. Bartick, E. G.; Merrill, R. A.; Egan, W. J.; Kochanowski, B. K.; Morgan, S. L. In *Forensic discrimination of photocopy toners by FT-infrared reflectance spectroscopy*, AIP Conference Proceedings, 1998; p 257.

23. Merrill, R. A.; Bartick, E. G.; Mazzella, W. D., Studies of techniques for analysis of photocopy toners by IR. *Journal of Forensic Sciences* 1996, *41*, 264-271.

24. Vikman, K.; Sipi, K., Applicability of FTIR and Raman spectroscopic methods to the study of paper-ink interactions in digital prints. *Journal of Imaging Science and Technology* 2003, *47* (2), 139-148.

25. Lennard, C.; Mazzella, W., A simple combined technique for the analysis of toners and adhesives. *Journal of the Forensic Science Society* 1991, *31* (3), 365-371.

26. Mazzella, W. D.; Lennard, C. J.; Margot, P., Classification and identification of photocopying toners by diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS): I. Preliminary results. *Journal of Forensic Sciences, JFSCA* 1991, *36* (3), 820-837.

27. Mazzella, W. D.; Lennard, C. J.; Margot, P., Classification and identification of photocopying toners by diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS): II. Final report. *Journal of Forensic Sciences, JFSCA* 1991, *36* (3), 820-837.

28. Froment, F.; Tournié, A.; Colomban, P., Raman identification of natural red to yellow pigments: ochre and iron-containing ores. *Journal of Raman Spectroscopy* 2008, *39* (5), 560-568.

29. Colombini, A.; Kaifas, D., Characterization of some orange and yellow organic and fluorescent pigments by Raman spectroscopy. *Preservation Science* 2010, *7*, 14-21.

30. Udrştioiu, E. G.; Bunaciu, A. A.; Aboul-Enein, H. Y.; Tănase, I. G., Infrared spectrometry in discriminant analysis of laser printer and photocopy toner on questioned documents. *Instrumentation Science and Technology* 2009, *37* (2), 230-240.

31. Skenderovic Bozicevic, M.; Gajovic, A.; Zjakic, I., Identifying a common origin of toner printed counterfeit banknotes by micro-Raman spectroscopy. *Forensic Science International* 2012, *223* (1-3), 314-20.

32. Morgan, S. L.; Bartick, E. G., Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis, Blackledge, R. D., Ed. John Wiley & Sons, Inc.: New York, 2007; pp 333-374.

33. William Daubert, et ux., etc., et al., petitioners v. Merrell Dow Pharmaceuticals, Inc. U.S. Supreme Court: 1993; Vol. 509 US 579, 589.

34. National Resource Council, Strengthening Forensic Science in the United States: a path forward. National Academies Press: Washington, DC, 2009.

35. Johnson, R.; Wichern, D., Applied Multivariate Statistical Analysis. Prentice Hall: Englewood Cliffs, NJ, 1982.

36. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., Chemometrics: A Practical Guide. John Wiley & Sons, Inc.: New York, 1998.

37. Brereton, R. G., Chemometrics for Pattern Recognition. John Wiley & Sons, Inc.: Chichester, England, 2009.

38. Brereton, R. G., Chemometrics: Data Analysis for the Laboratory and Chemical Plant. John Wiley & Sons, Ltd.: Chichester, England, 2003.

39. Brereton, R. G., Chemomentrics: Applications of Mathematics and Statistics to Laboratory Systems. Ellis Horwood Limited: Chichester, England, 1990.

40. StatSoft, I., Electronic Statistics Textbook. StatSoft: Tulsa, OK, 2011. WEB: http://www.statsoft.com/textbook/ (accessed January 2013).

41. Hida, M.; Sato, H.; Sugawara, H.; Mitsui, T., Classification of counterfiet coins using multivariate analysis with X-ray diffraction and X-ray fluroescence methods. *Forensic Science International* 2001, *115*, 129-134.

42. Goodpaster, J. V.; Sturdevant, A. B.; Kristen L. Andrews; Brun-Conti, L., Identification and comparison of electrical tapes using instrumental and statistical techniques: I. Microscopic surface texture and elemental composition. *Journal of Forensic Sciences* 2007, *52* (3), 610-629.

43. Goodpaster, J. V.; Sturdevant, A. B.; Andrews, K. L.; Briley, E. M.; Brun-Conti, L., identification and comparison of electrical tapes using instrumental and statistical techniques: II. Organic composition of the tape backing and adhesives. *Journal of Forensic Sciences* 2009, *54* (2), 328-338.

44. Sandercock, P. M. L.; Pasquier, E. D., Chemical fingerprinting of unevaporated automotive gasoline samples. *Forensic Science International* 2003, *134* (1), 1-10.

45. Sandercock, P. M. L.; Pasquier, E. D., Chemical fingerprinting of gasoline: II. Comparison of unevaporated and evaporated automotive gasoline samples. *Forensic Science International* 2004, *140* (1), 43-59.

46. Barrett, J. A.; Siegel, J. A.; Goodpaster, J. V., Forensic discrimination of dyed hair color: II. Multivariate statistical analysis. *Journal of Forensic Sciences* 2011, *56* (1), 95-101.

47. Turner, D. A.; Goodpaster, J. V., Comparing the effects of weathering and microbial degradation on gasoline using principal components analysis. *Journal of Forensic Sciences* 2012, *57* (1), 64-69.

48. Klemenc, S., In common batch searching of illicit heroin samples - evaluation of data by chemometrics methods. *Forensic Science International* 2000, *115* (1-2), 43-52.

49. Thanasoulias, N. C.; Piliouris, E. T.; Kotti, M.-S. E.; Evmiridis, N. P., Application of multivariate chemometrics in forensic soil discrimination based on the UV-Vis spectrum of the acid fraction of humus. *Forensic Science International* 2002, *130* (2-3), 73-82.

50. Muehlethaler, C.; Massonnet, G.; Esseiva, P., The application of chemometrics on Infrared and Raman spectra as a tool for the forensic analysis of paints. *Forensic Science International* 2011, *209* (1-3), 173-182.

51. Thanasoulias, N. C.; Parisis, N. A.; Evmiridis, N. P., Multivariate chemomentrics for the forensic determination of blue ball-point pen pinks based on their Vis spectra. *Forensic Science International* 2003, *138* (1-3), 75-84.

52. Manly, B. F. J., Multivariate Statistical Methods: A primer. 3rd ed.; Chapman & Hall/CRC: 2005; p 214.

53. Sharaf, M. A.; Illman, D. L.; Kowalski, B. R., Chemometrics. John Wiley & Sons, Inc.: 1986.

54. Mat Desa, W. N. S.; Daeid, N. N.; Ismail, D.; Savage, K., Application of unsupervised chemometric analysis and self-organizing feature map (SOFM) for the classification of lighter fuels. *Analytical Chemistry* 2010, *82* (15), 6395-6400.

55. Chazottes, V.; Brocard, C.; Peyrot, B., Particle size of soils under simulated scene of crime conditions: the interest of multivariate analyses. *Forensic Science International* 2004, *140*, 159-166.

56. Chalmers, J. M.; Everall, N. J., FTIR, FT-Raman and chemometrics: applications to the analysis and characterisation of polymers. *Analytical Chemistry* 1996, *15* (1), 18-25.

57. Varmuza, K.; Filzmoser, P., *Introduction to multivariate statistical analysis in chemometrics*. CRC: 2009.

58. Monfreda, M.; Gregori, A., Differentiation of unevaporated gasoline samples according to their brands, by SPME–GC–MS and multivariate statistical analysis. *Journal of Forensic Sciences* 2011, *56* (2), 372-380.

59. Sarkissian, G., The analysis of tire rubber traces collected after braking incidents using pyrolysis-gas chromatography/mass spectrometry. *Journal of Forensic Sciences* 2007, *52* (5), 1050-1056.

60. Princeton Instruments. Raman Spectroscopy Basics. WEB: http://content.piacton.com/Uploads/Princeton/Documents/Library/UpdatedLibrary/Raman_Spectroscopy_Basics.pdf (Accessed January 2013)

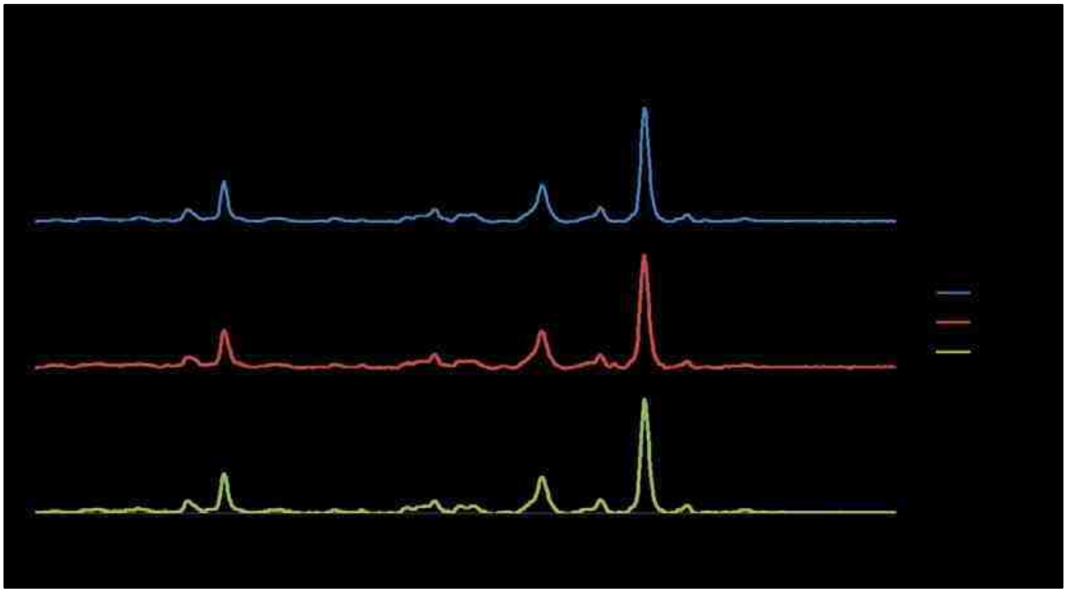
61. Harris, D. C.; Bertolucci, M. D., Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy. Dover Publications: 1989.

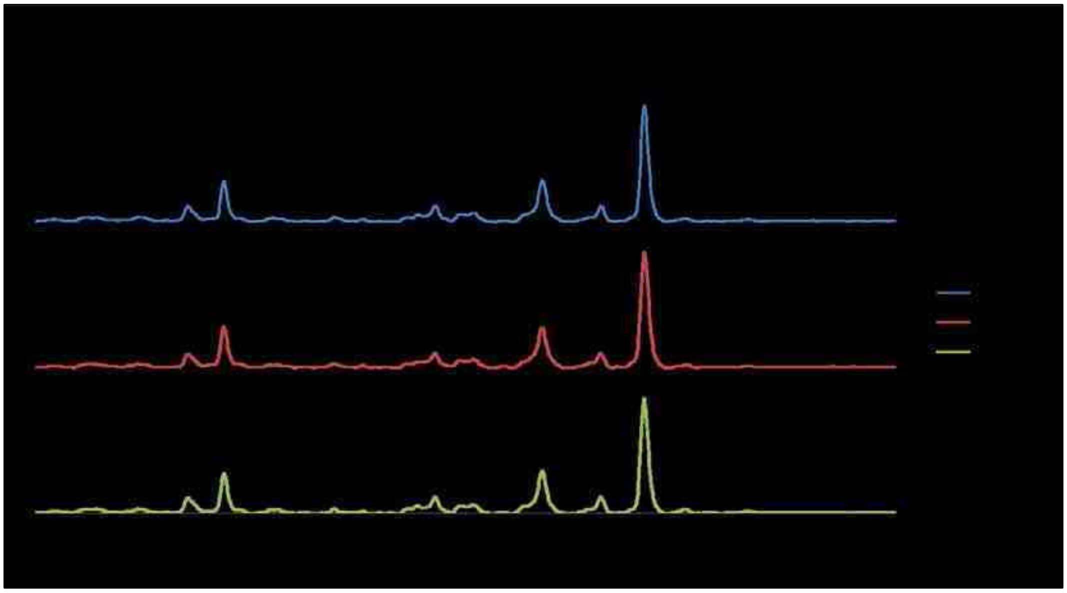
62. Andor Technology. Raman Spectroscopy: An Introduction to Raman Spectroscopy. WEB: http://www.andor.com/learning-academy/raman-spectroscopy-an-introduction-to-raman-spectroscopy (Accessed January 2013)

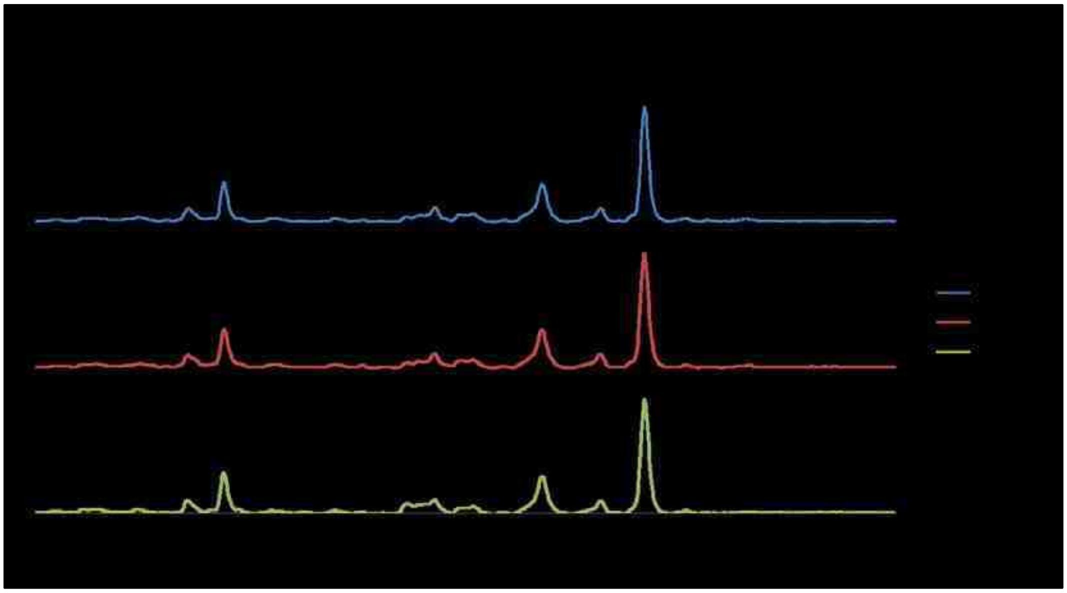
APPENDICES

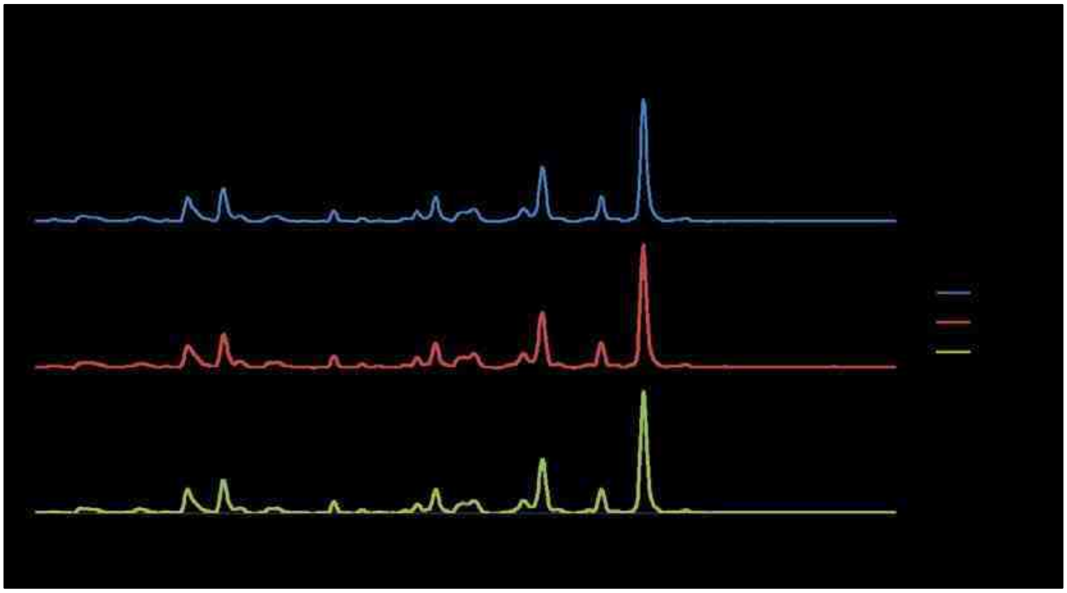
Appendix A. Blue Toner Spectra by Raman Spectroscopy

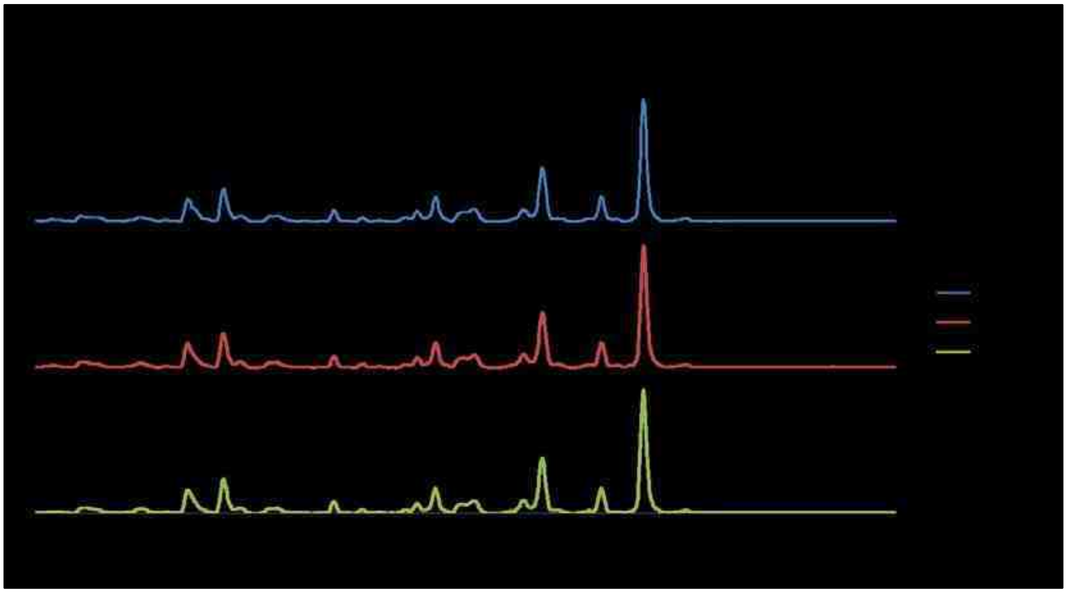
A.1 Training Samples

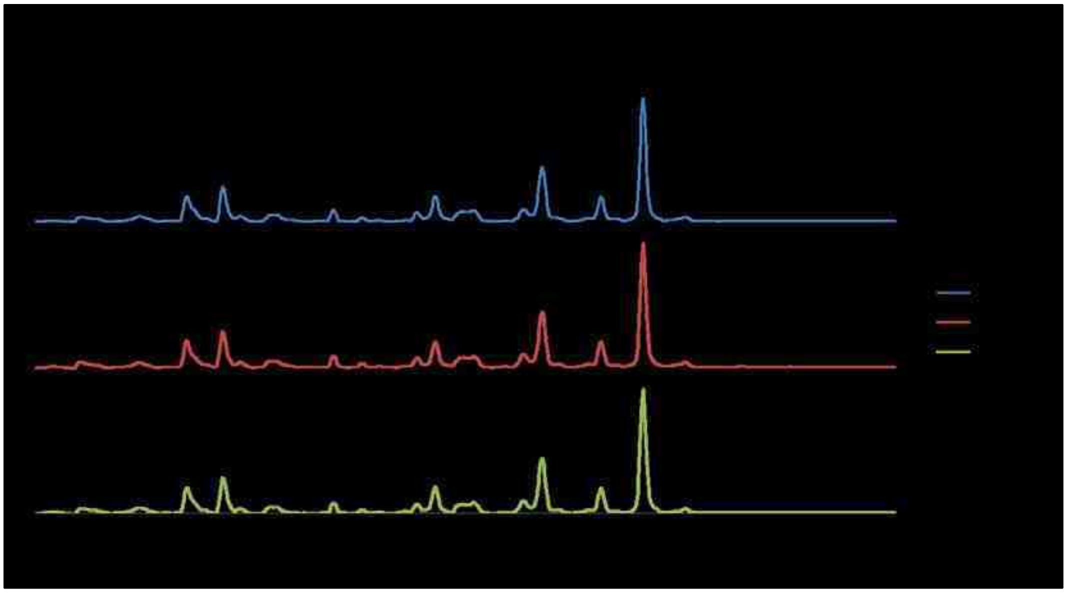


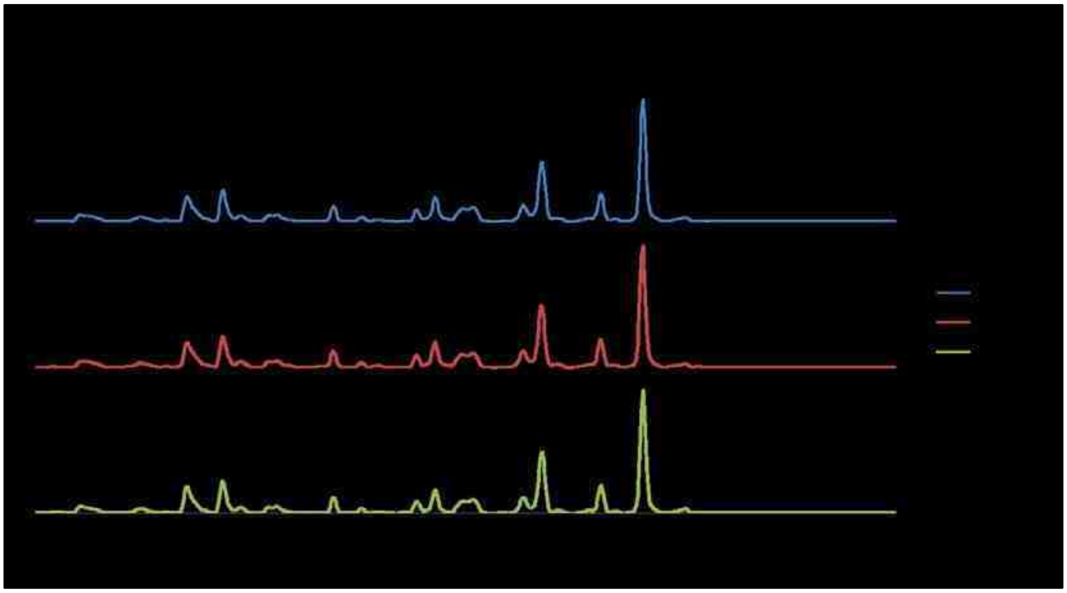


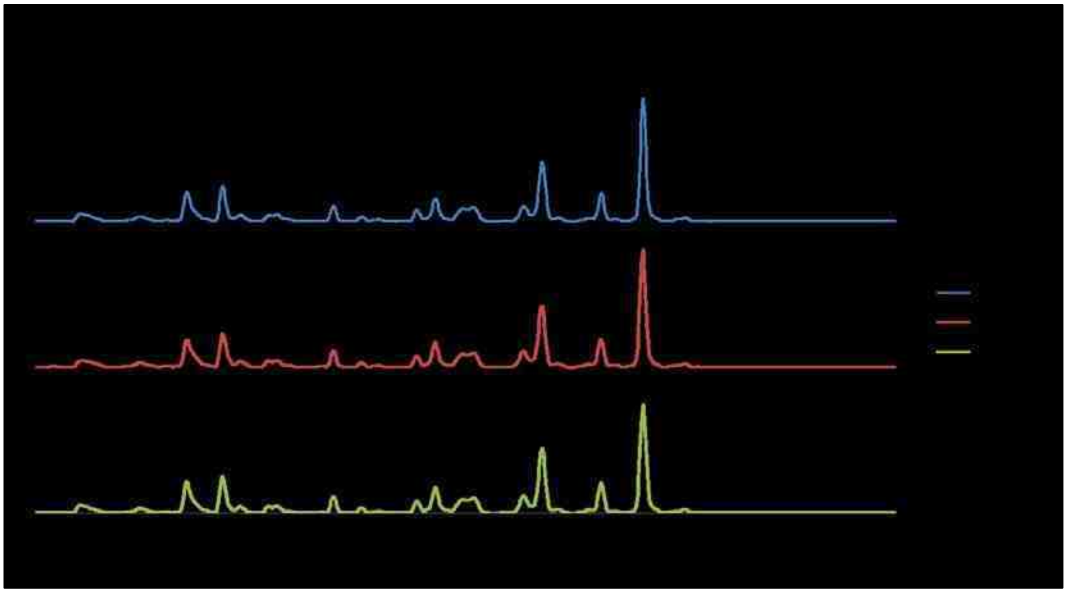


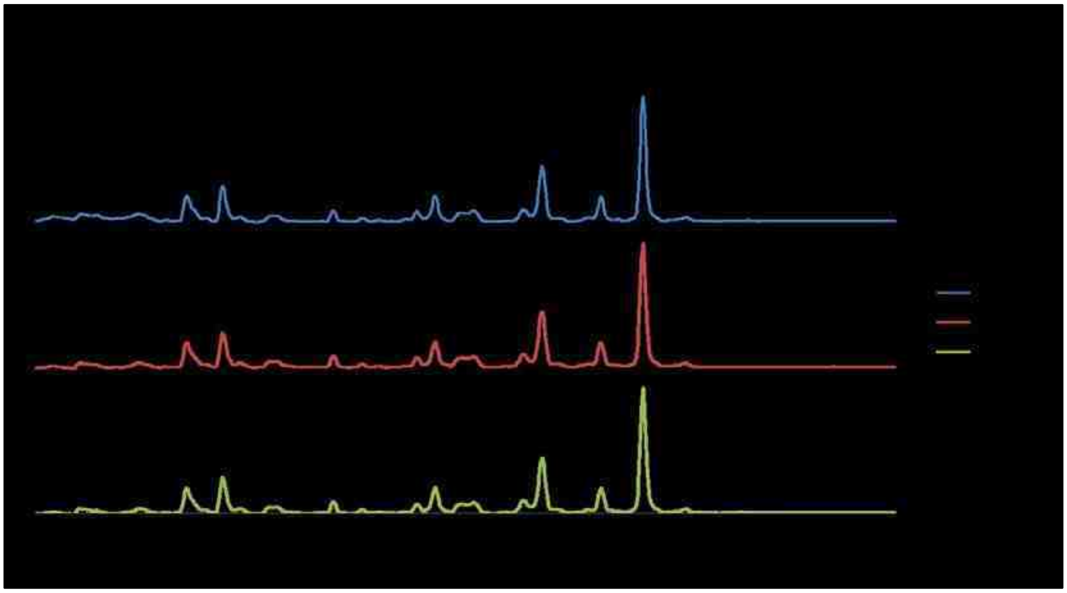


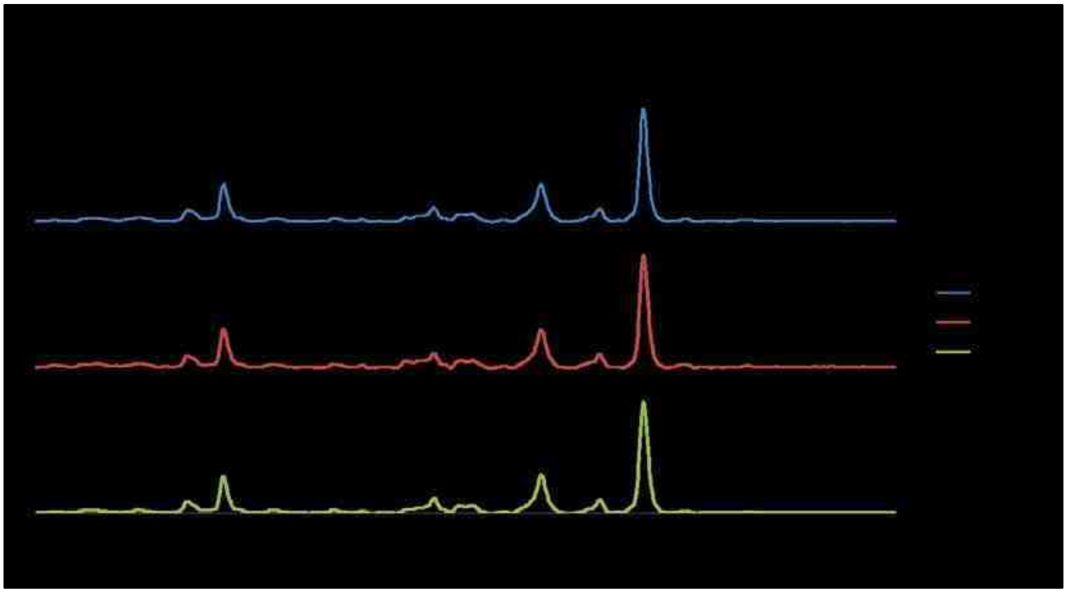


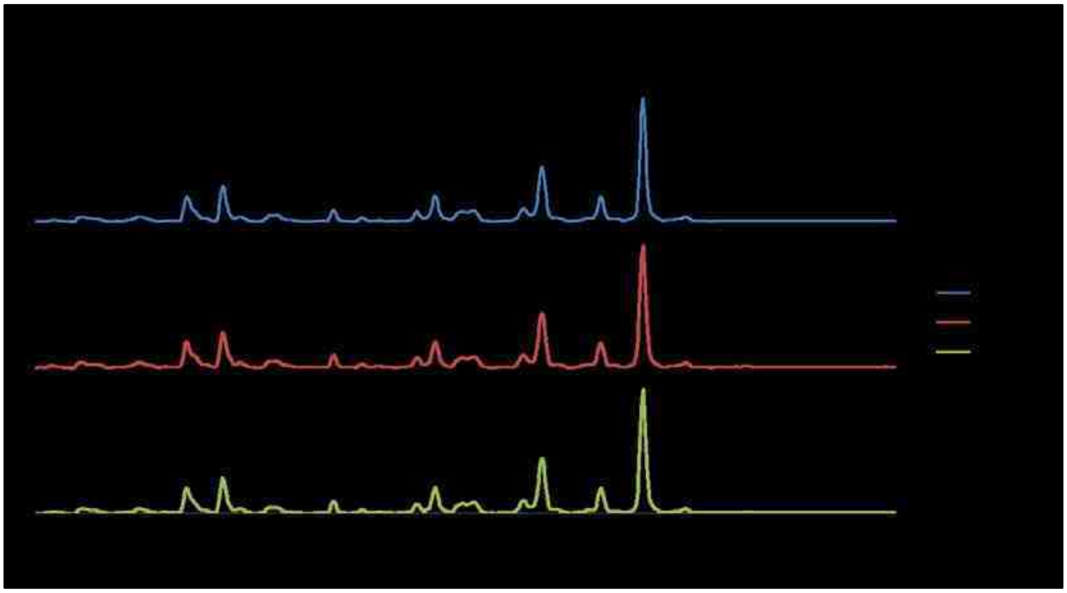


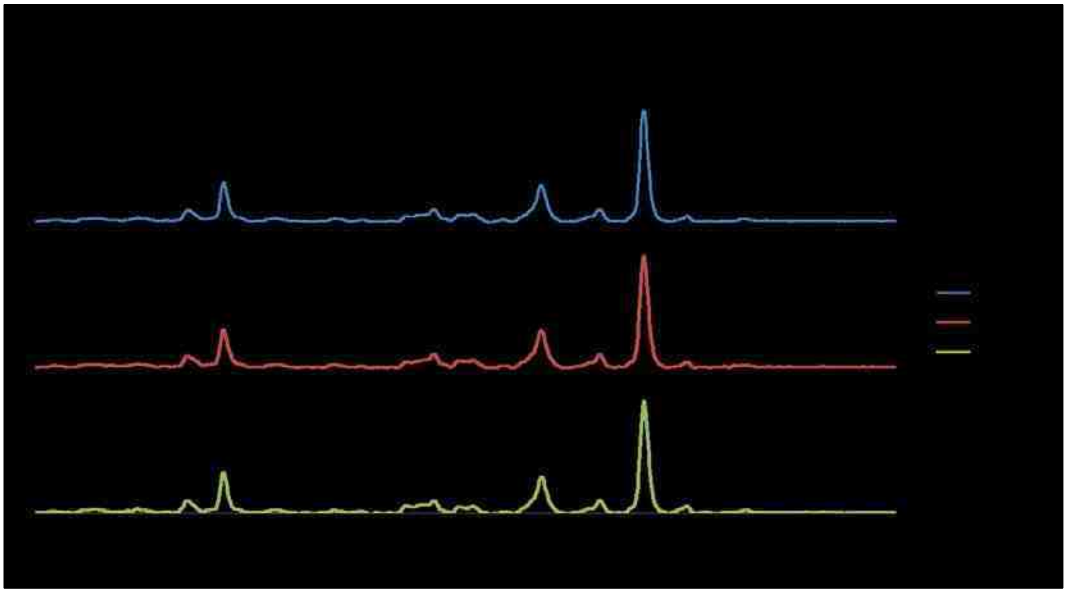


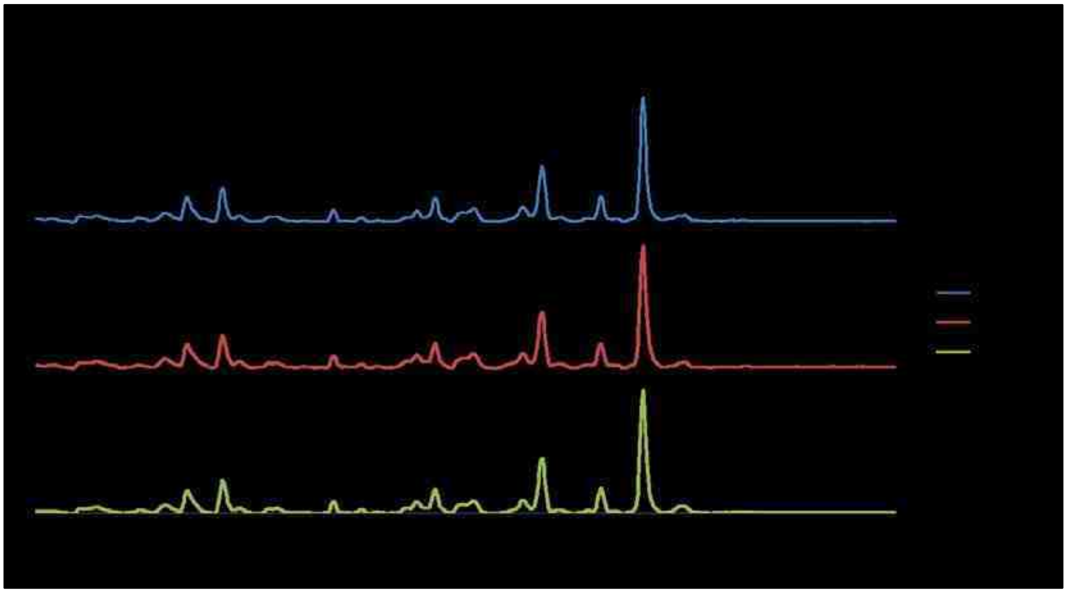


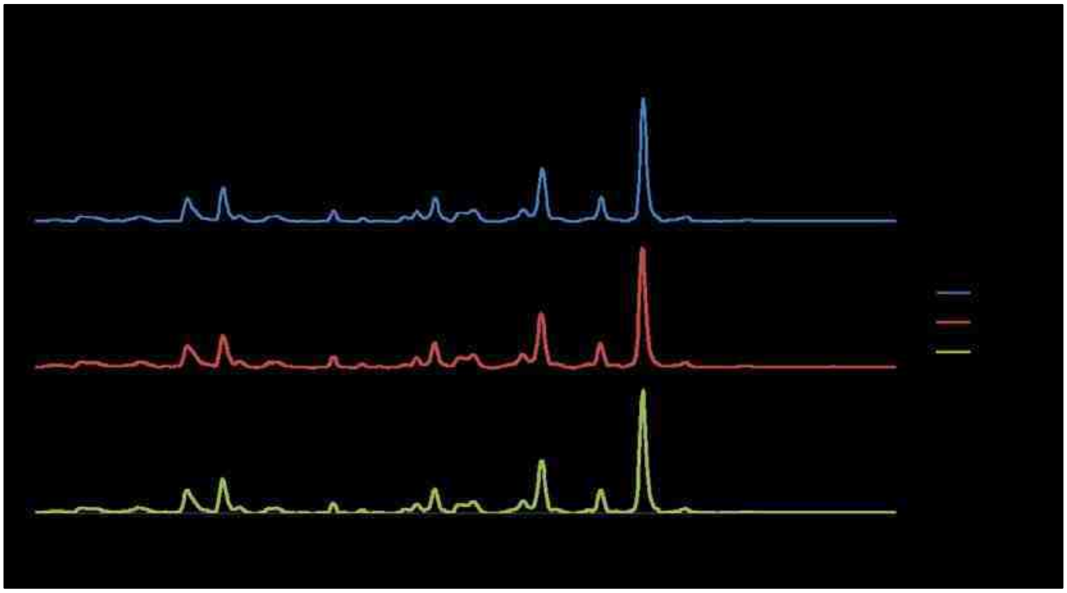


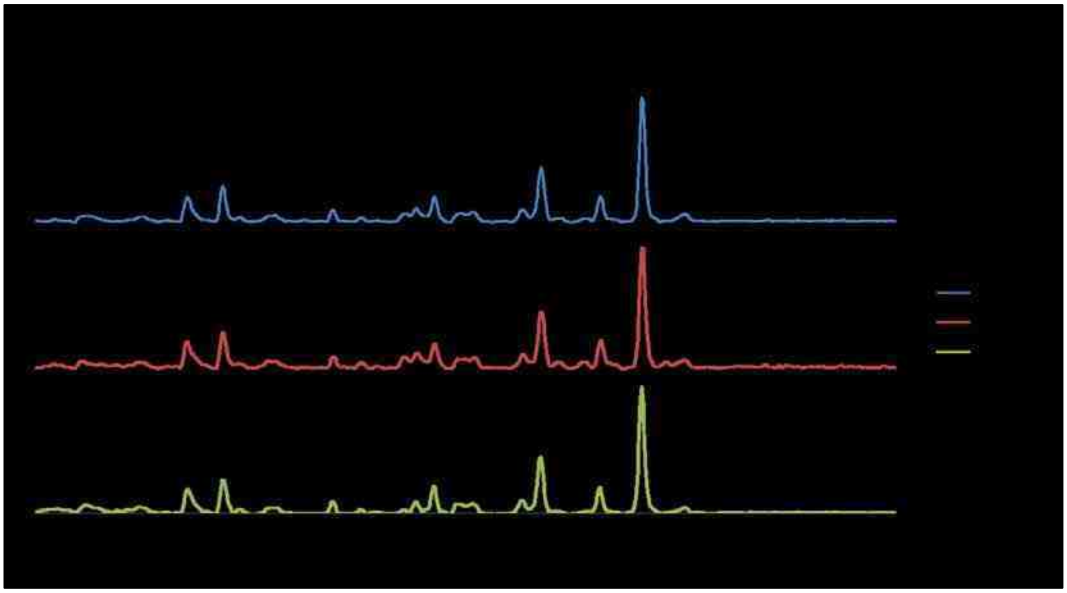


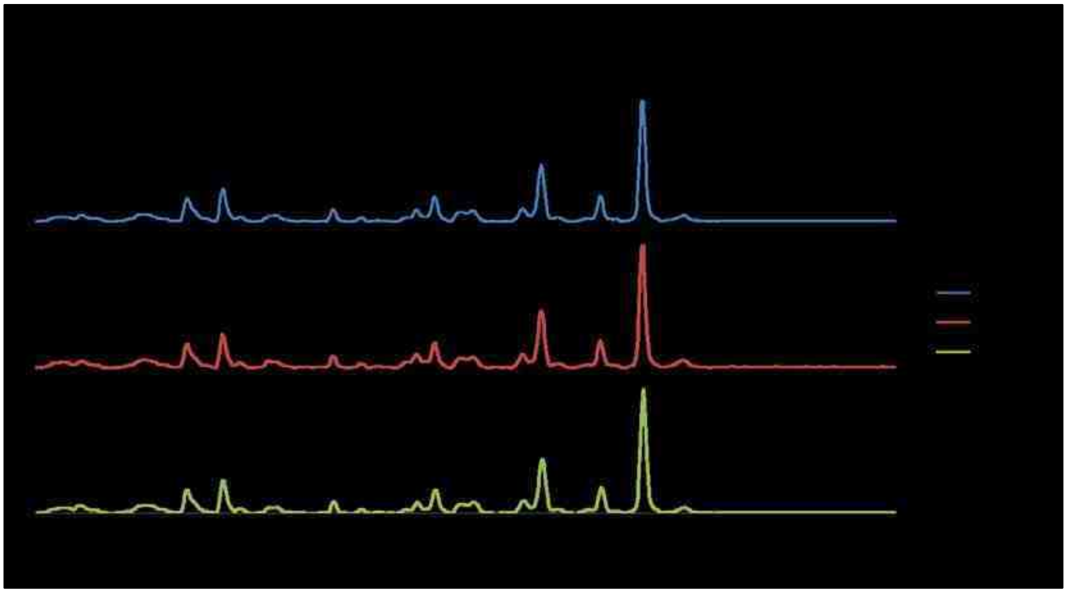


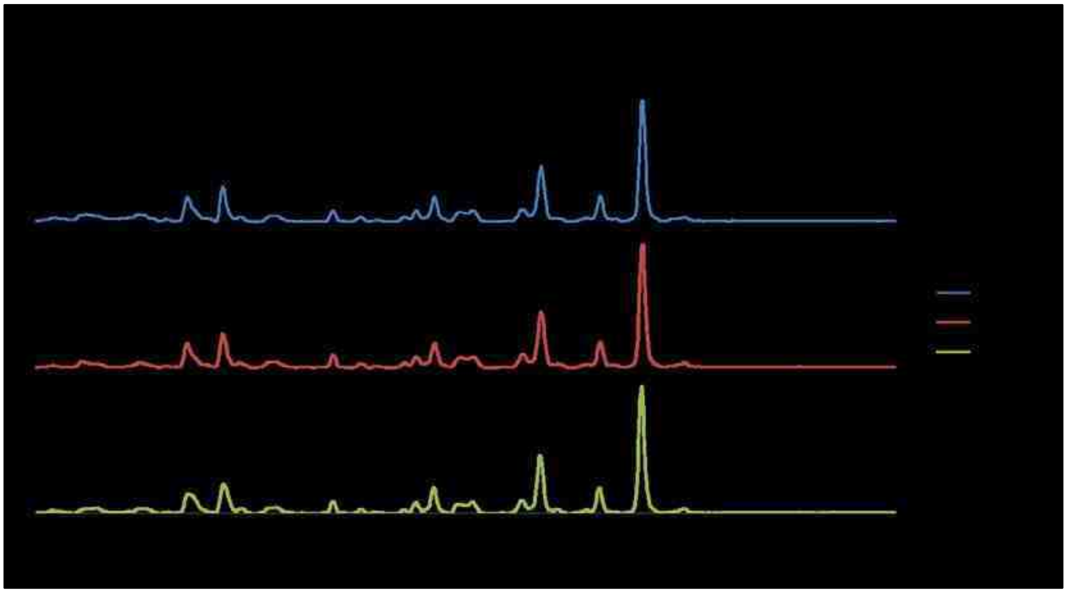


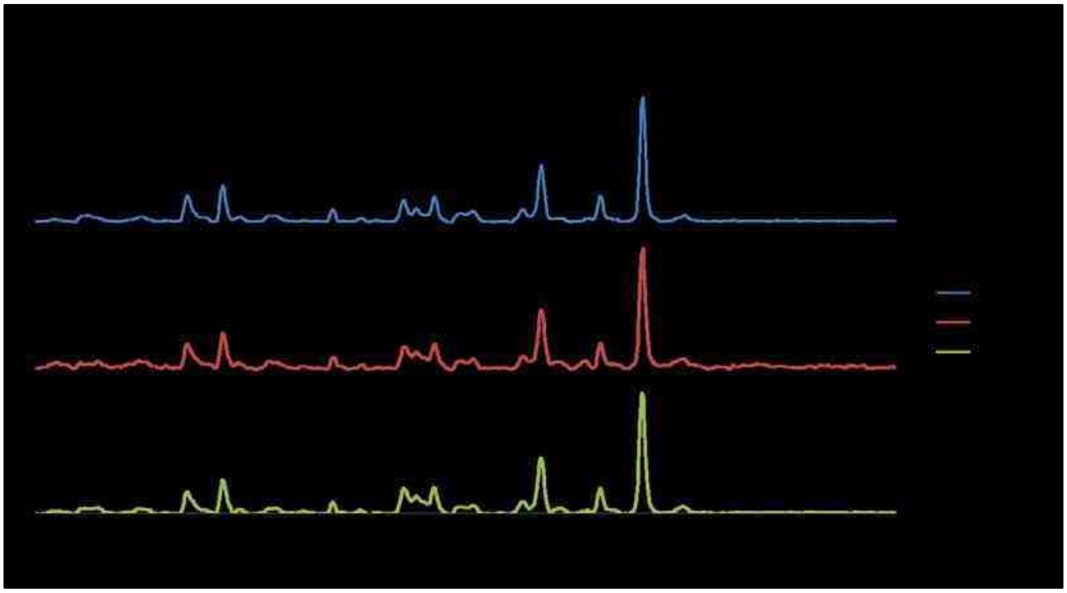


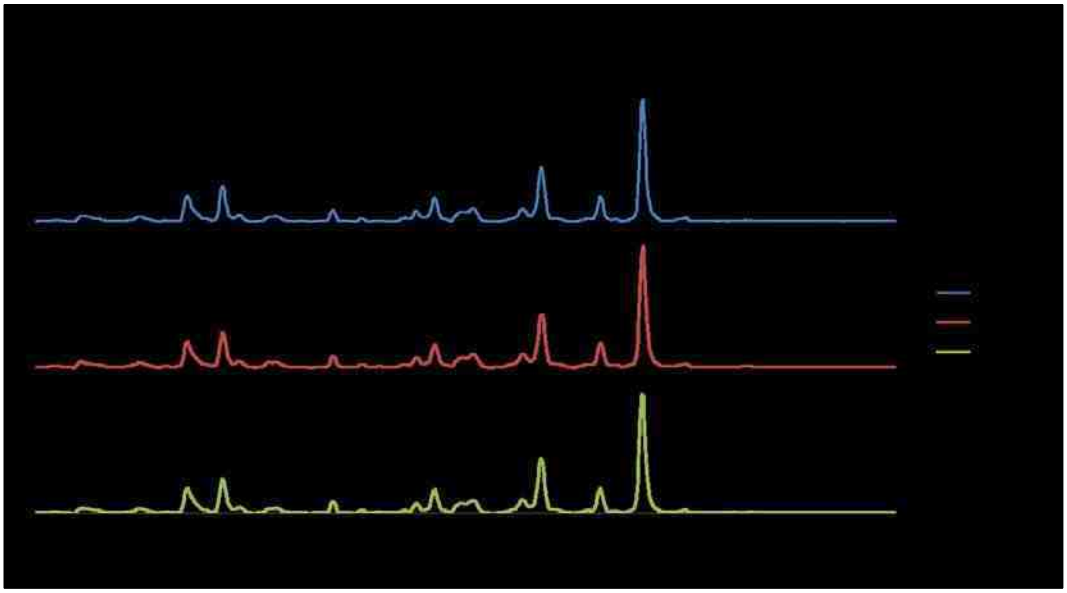


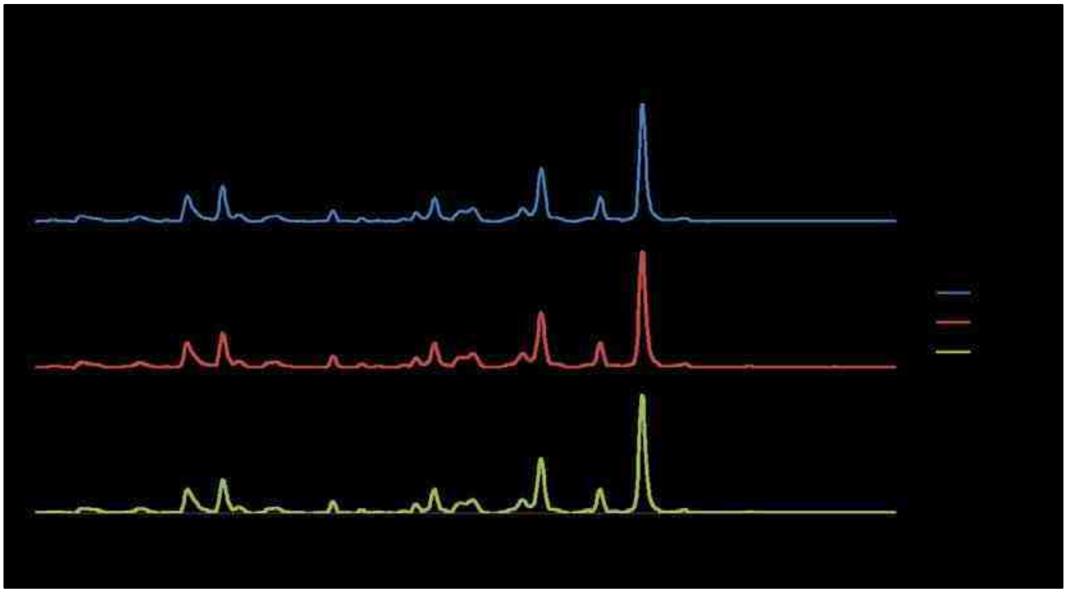


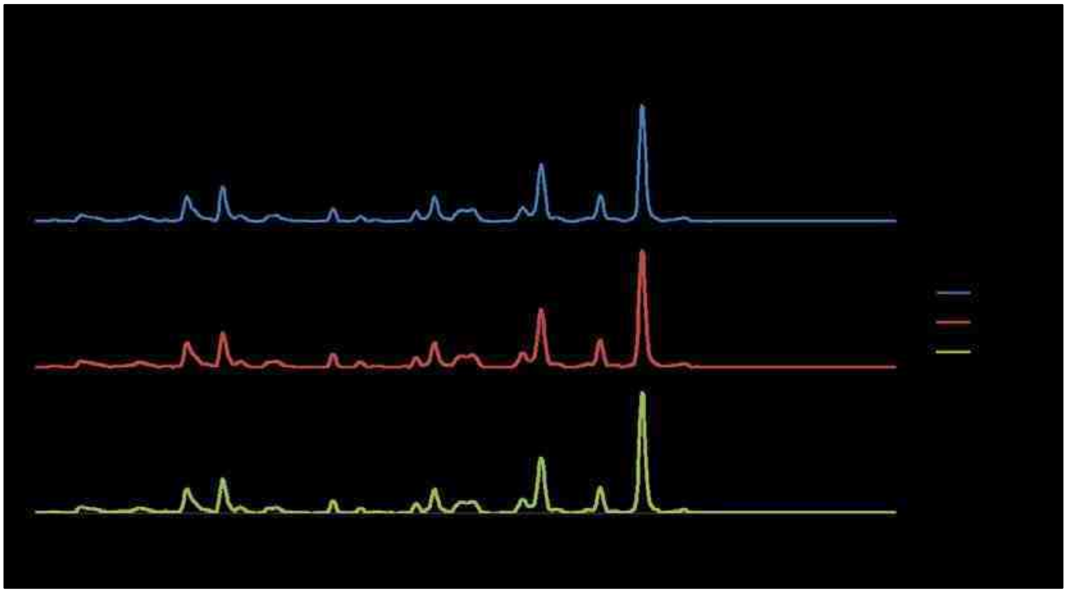


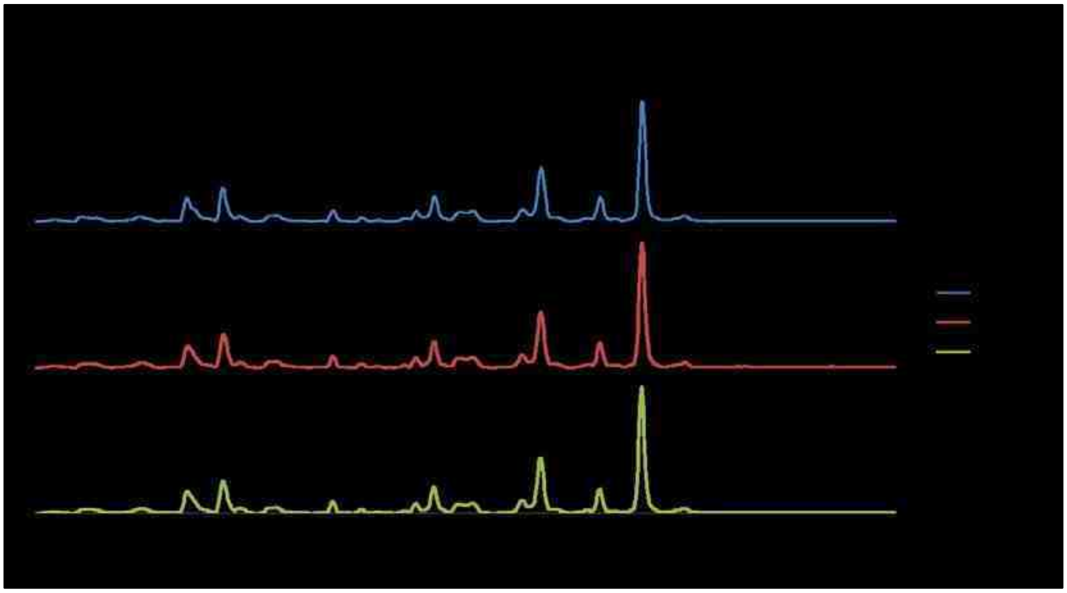


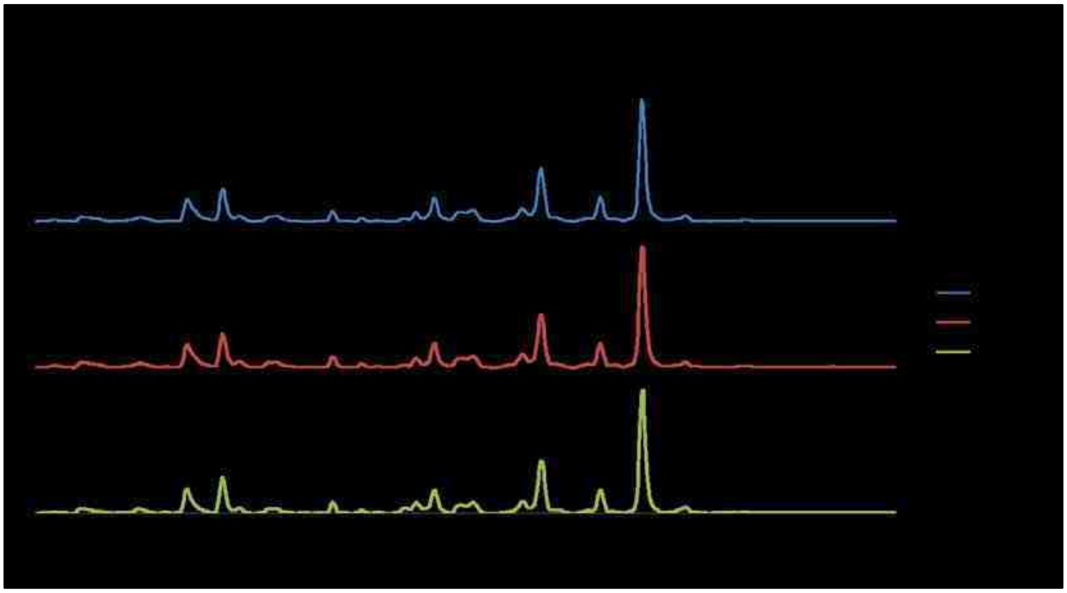


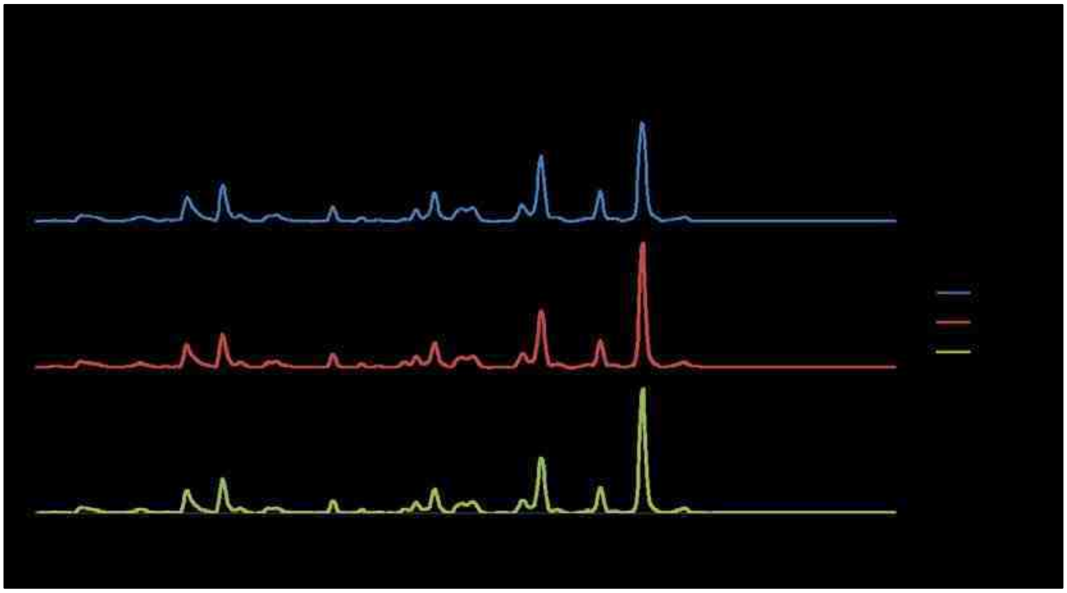


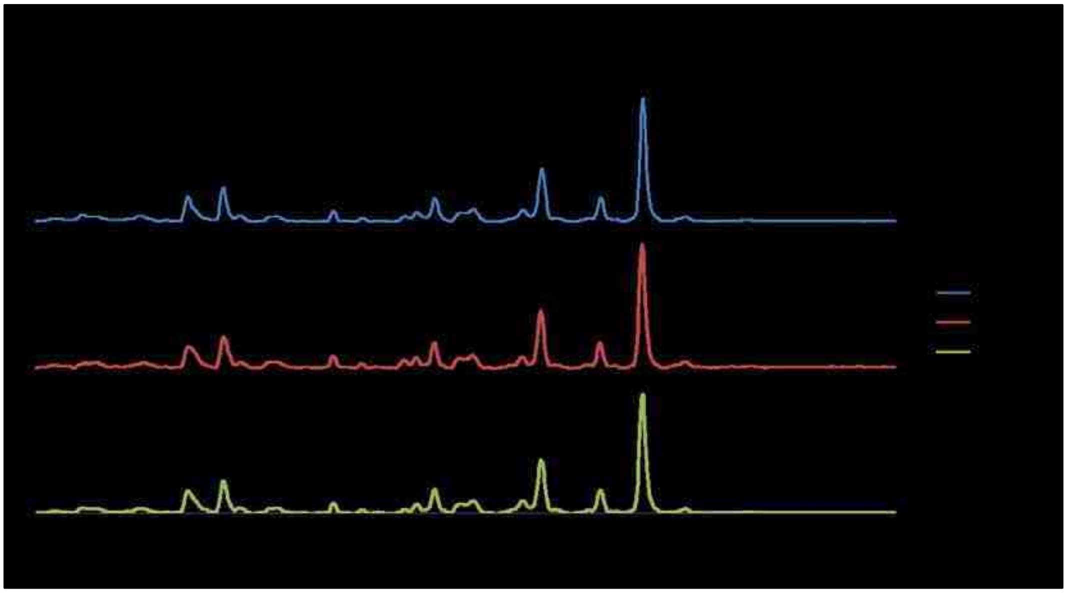


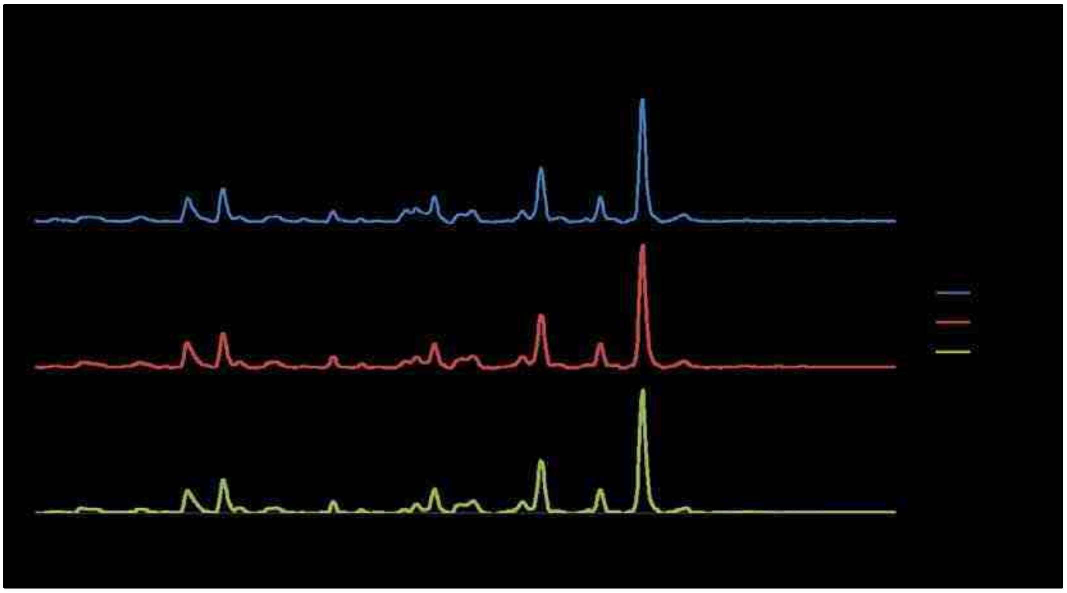


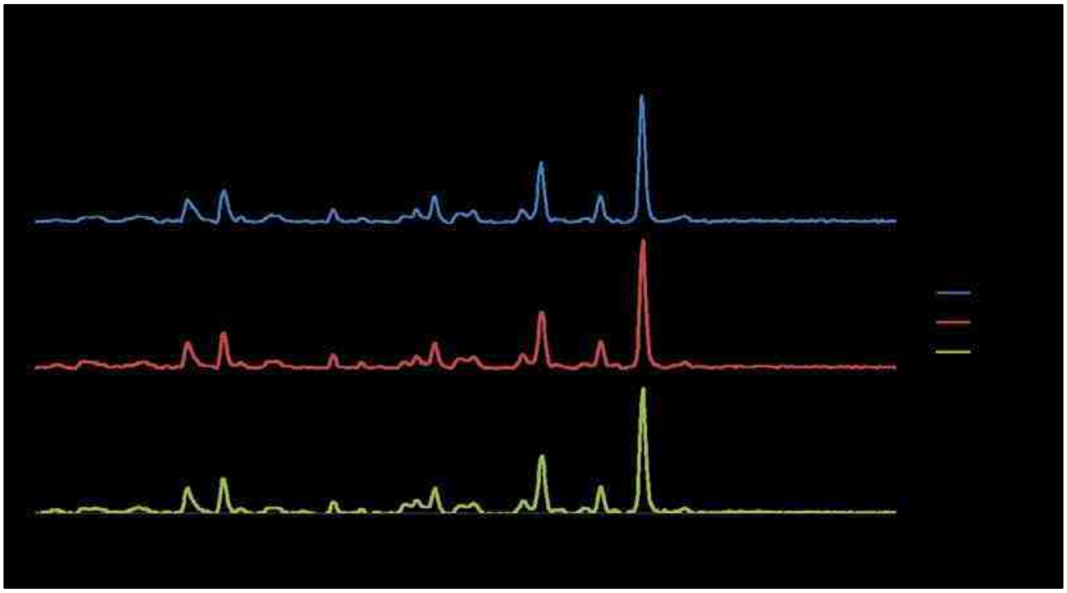


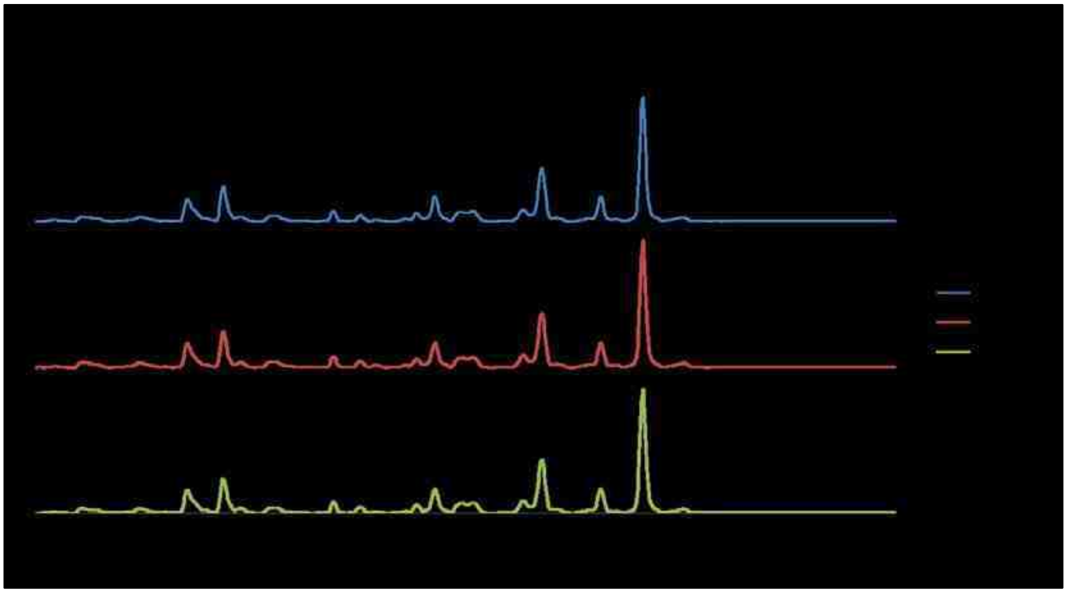


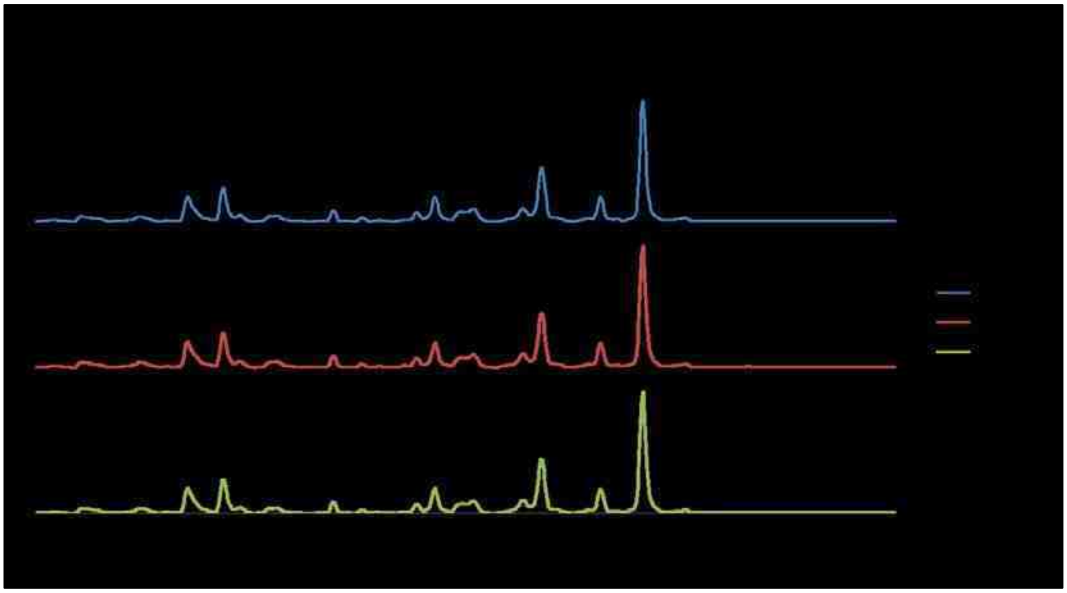


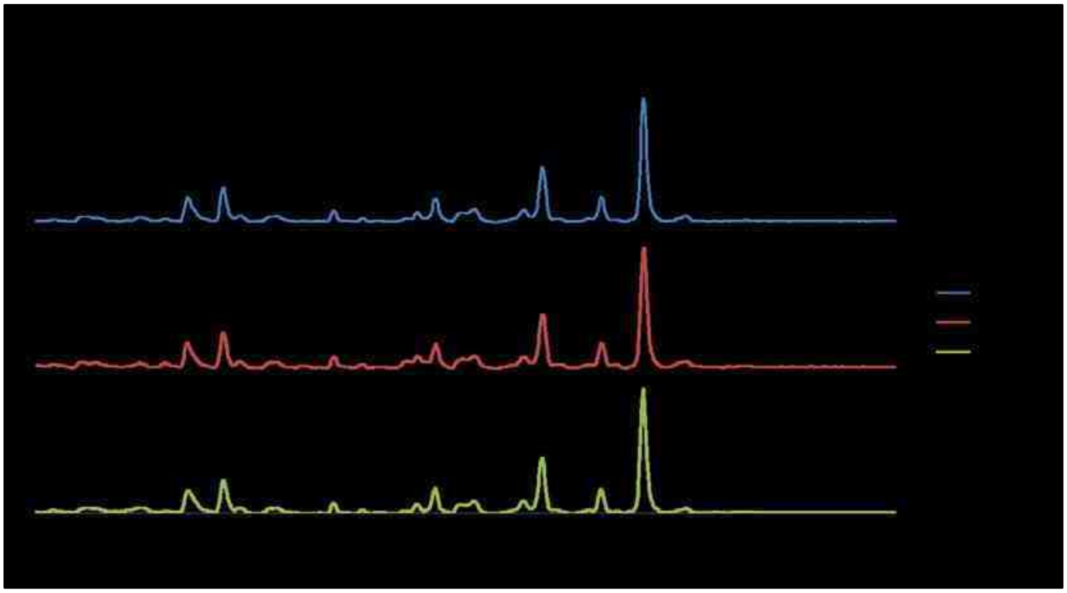


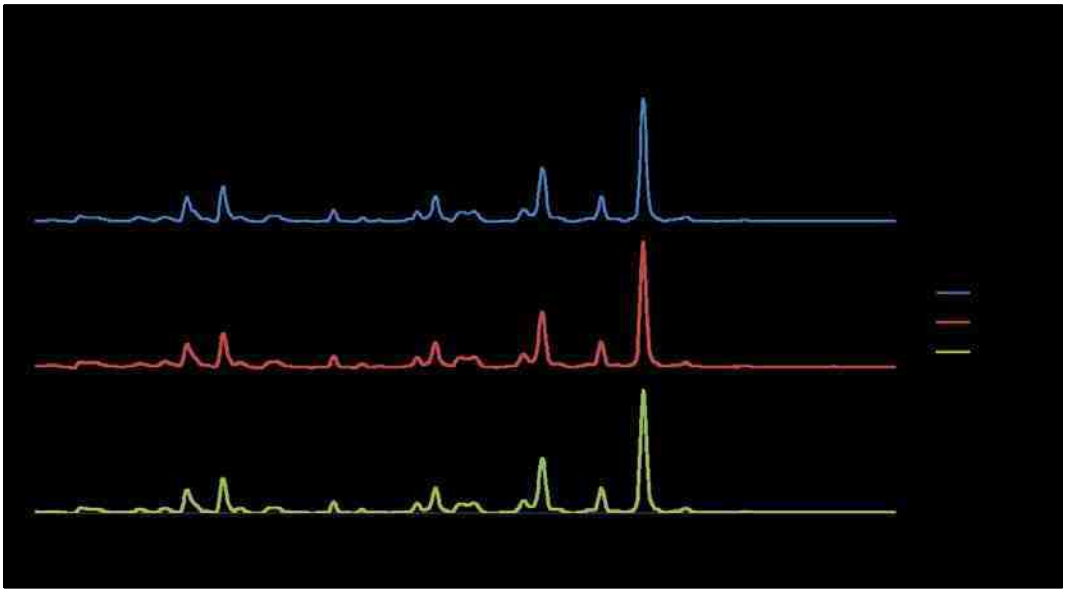


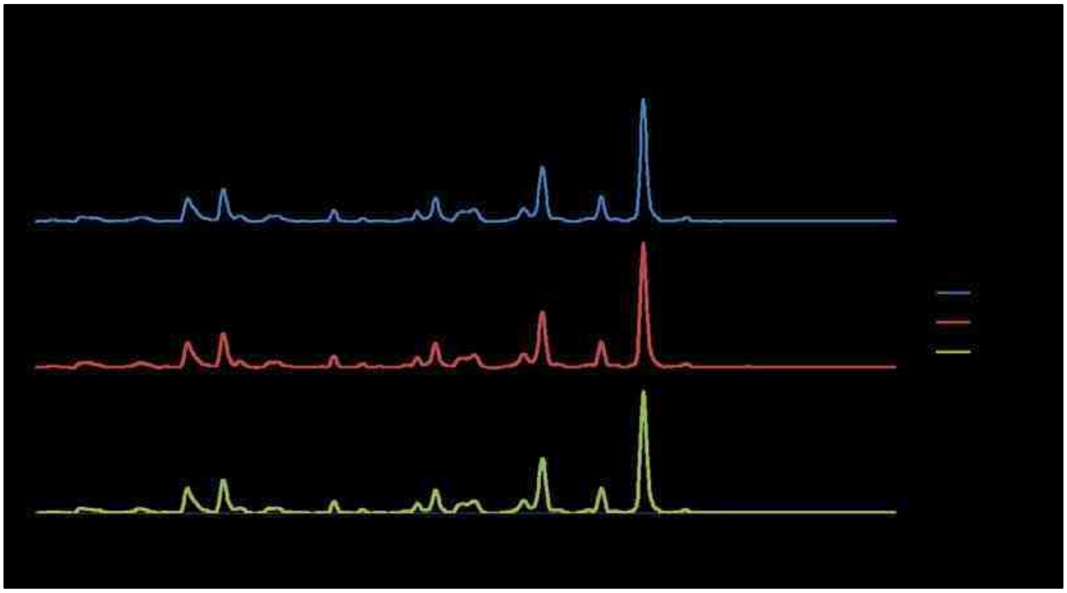


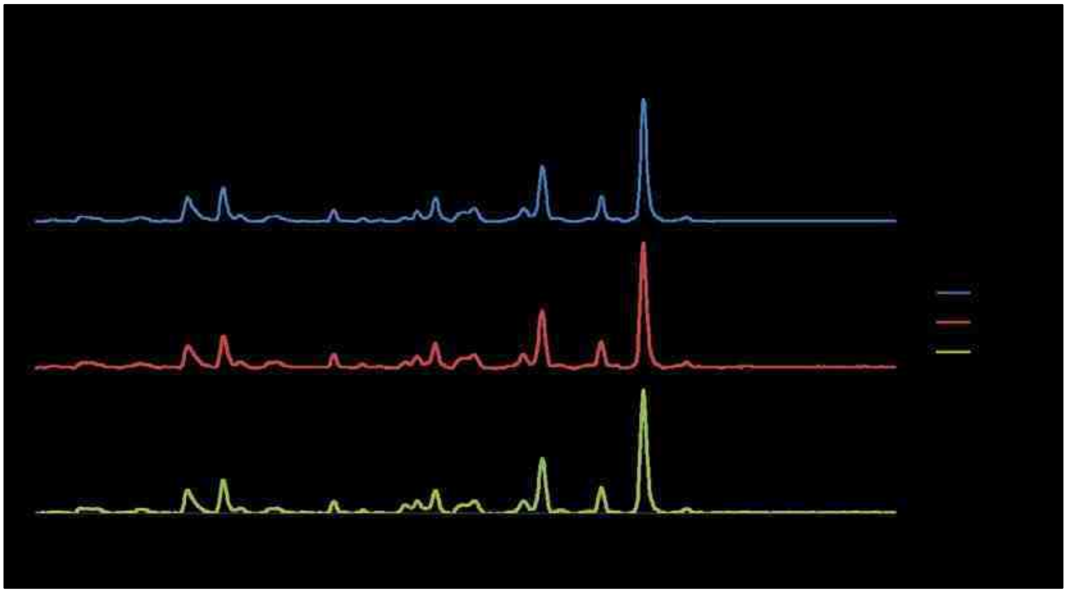


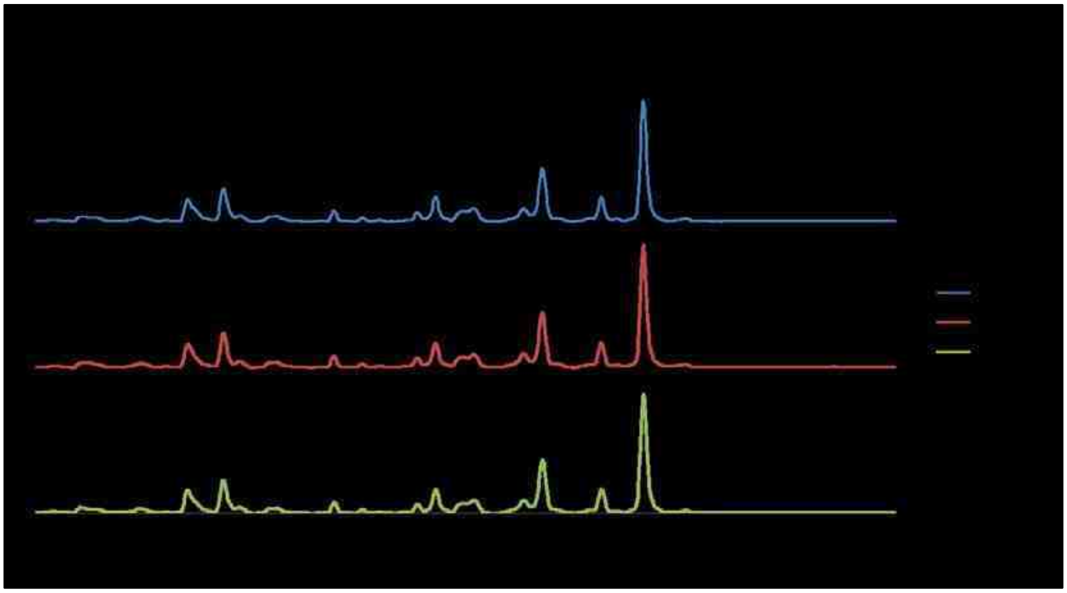


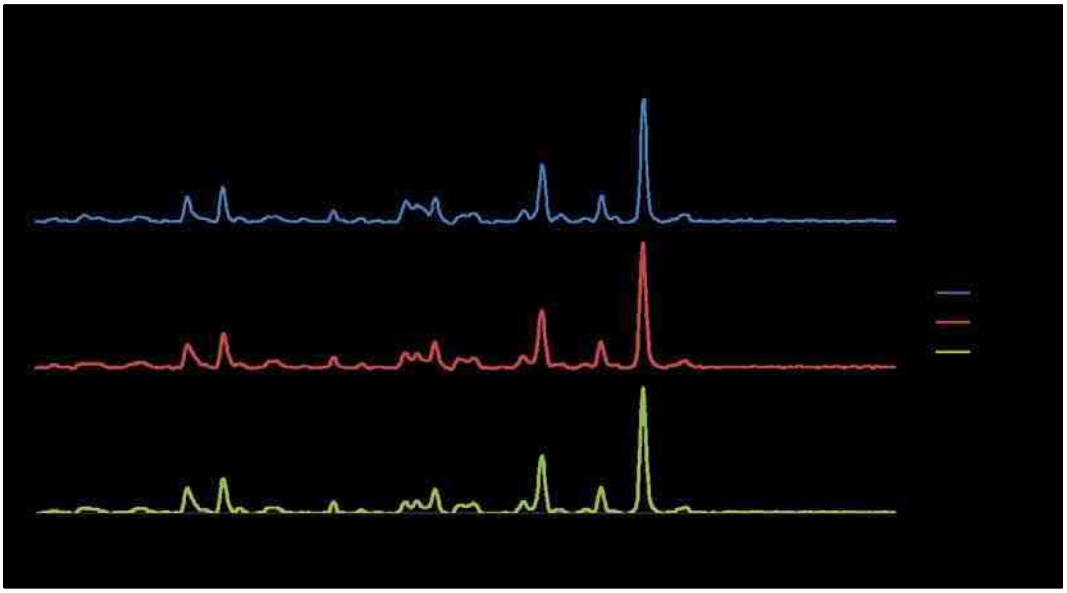


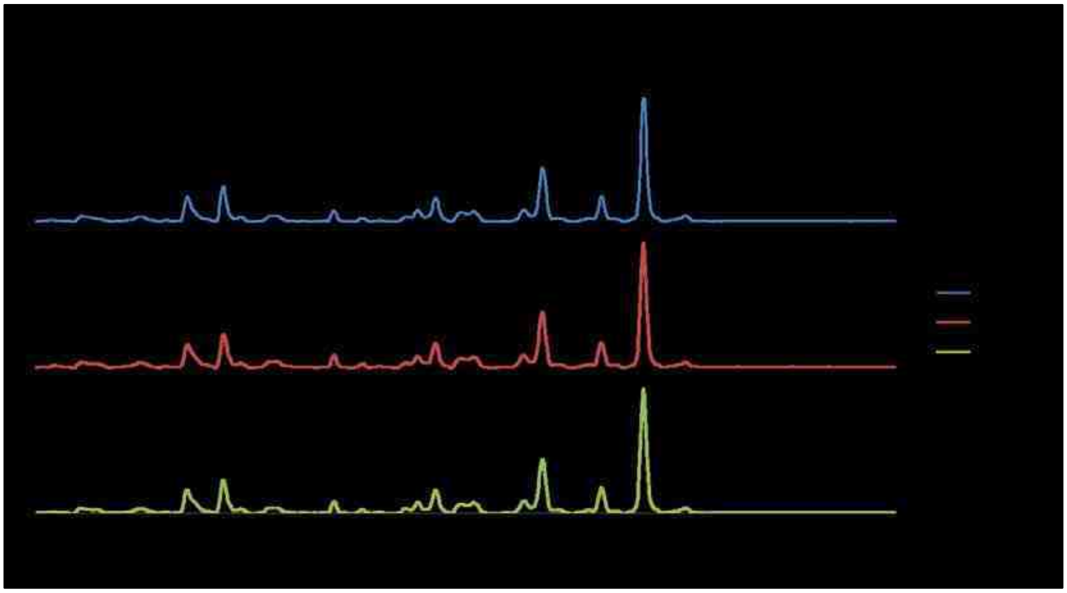


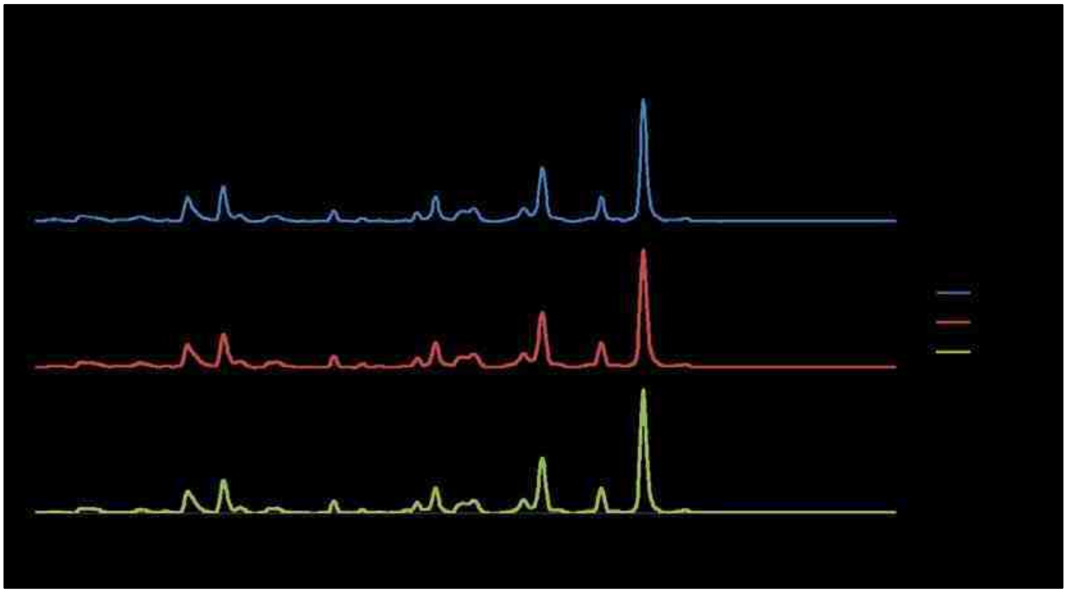


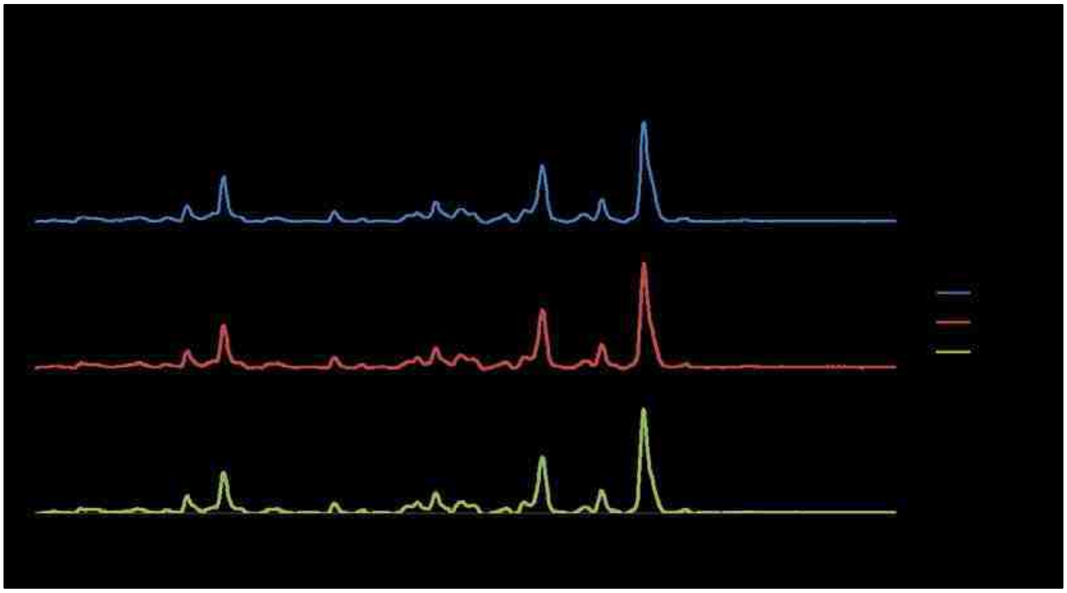


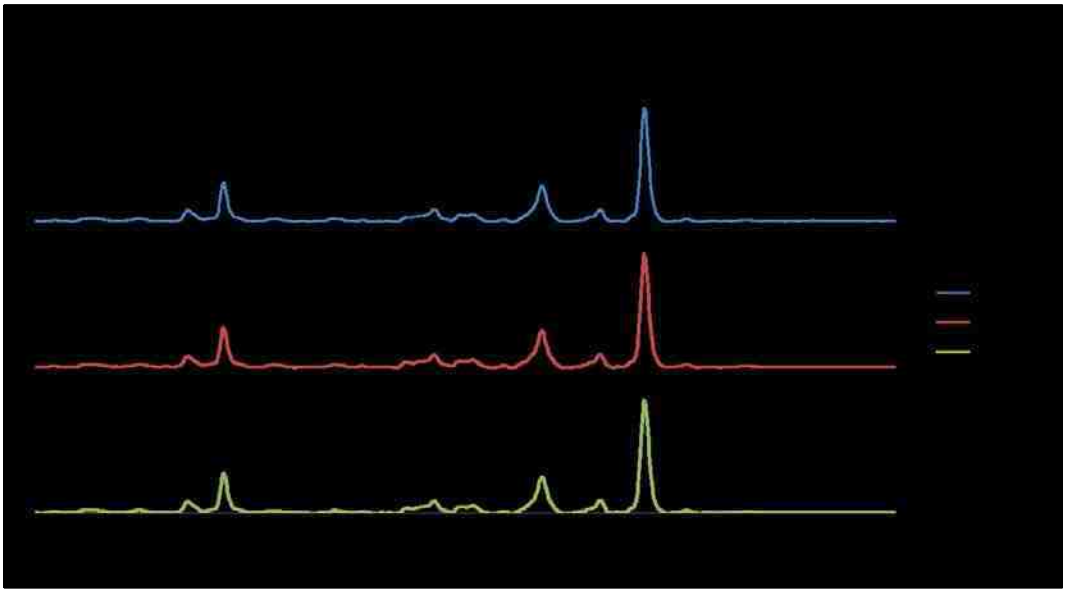


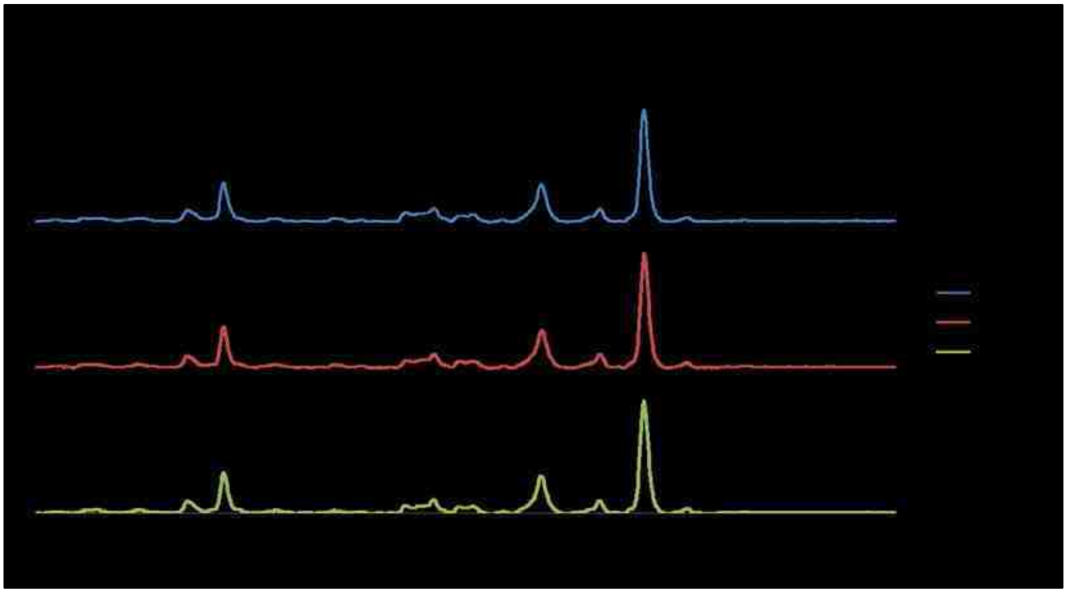


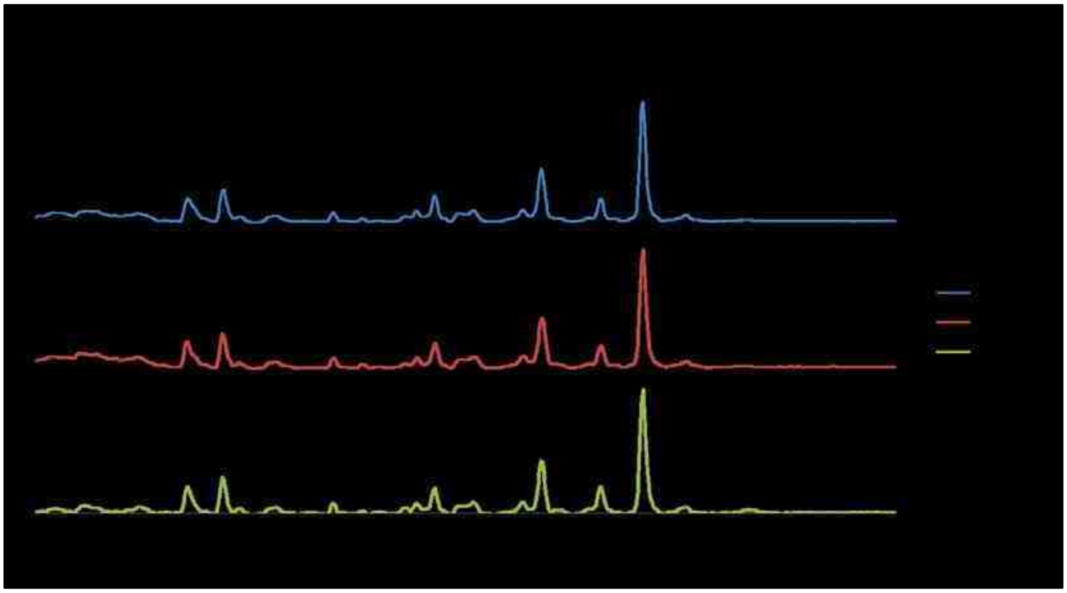


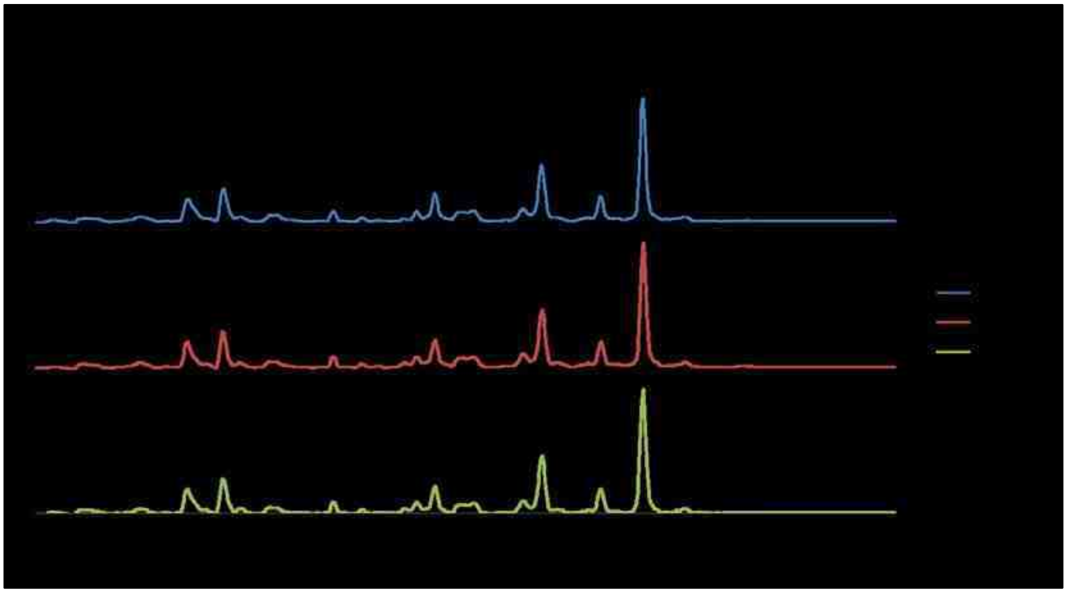


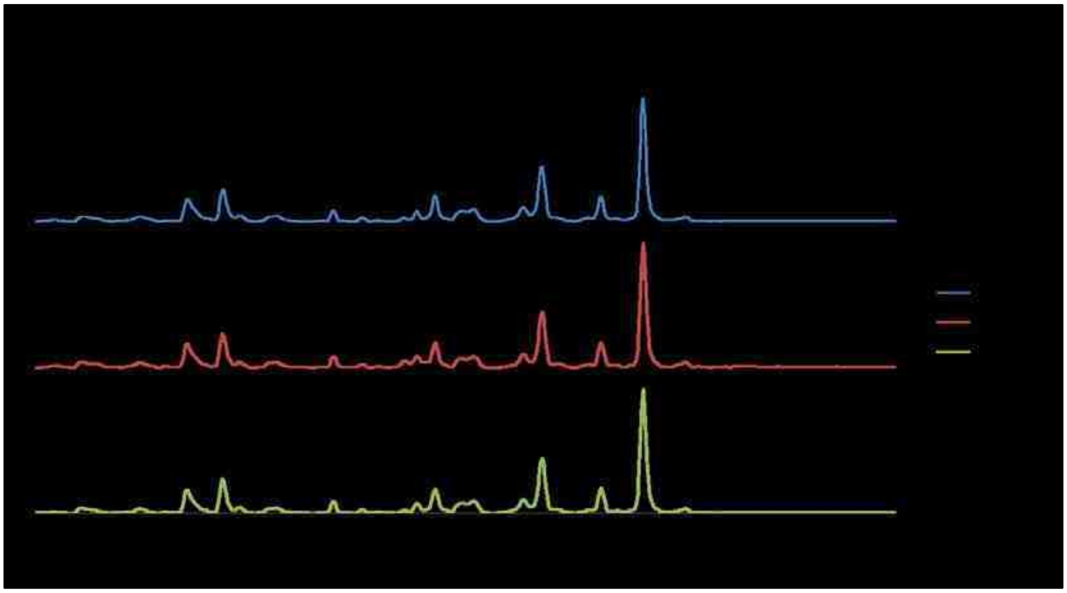


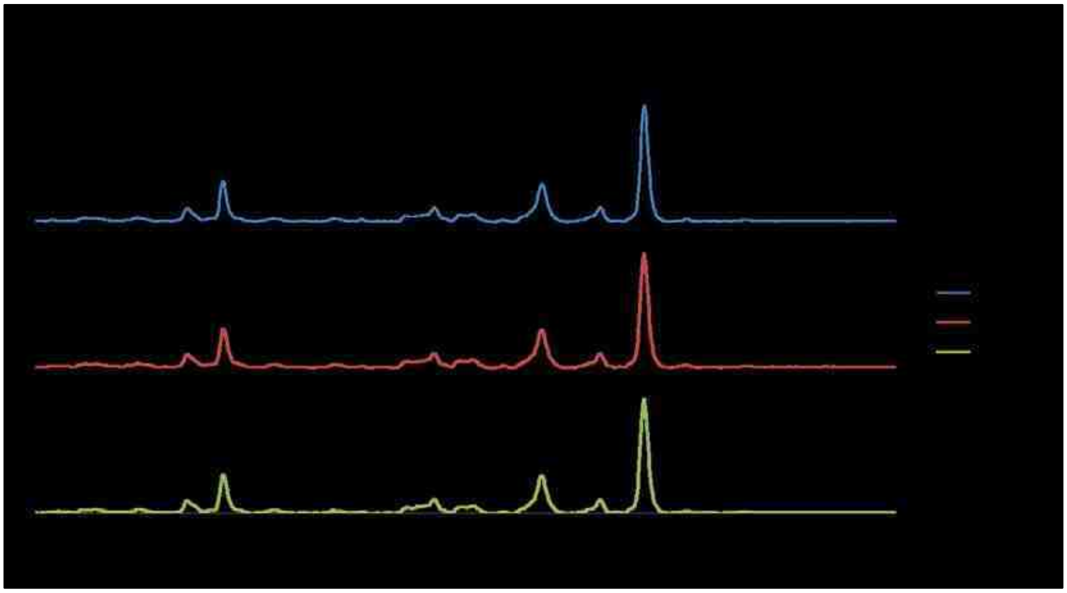


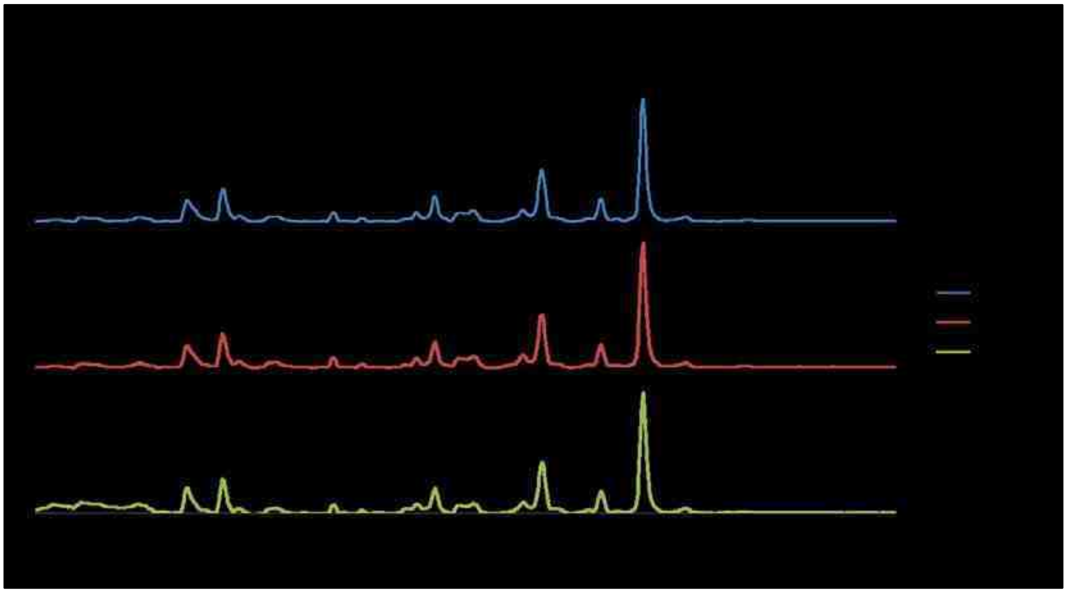


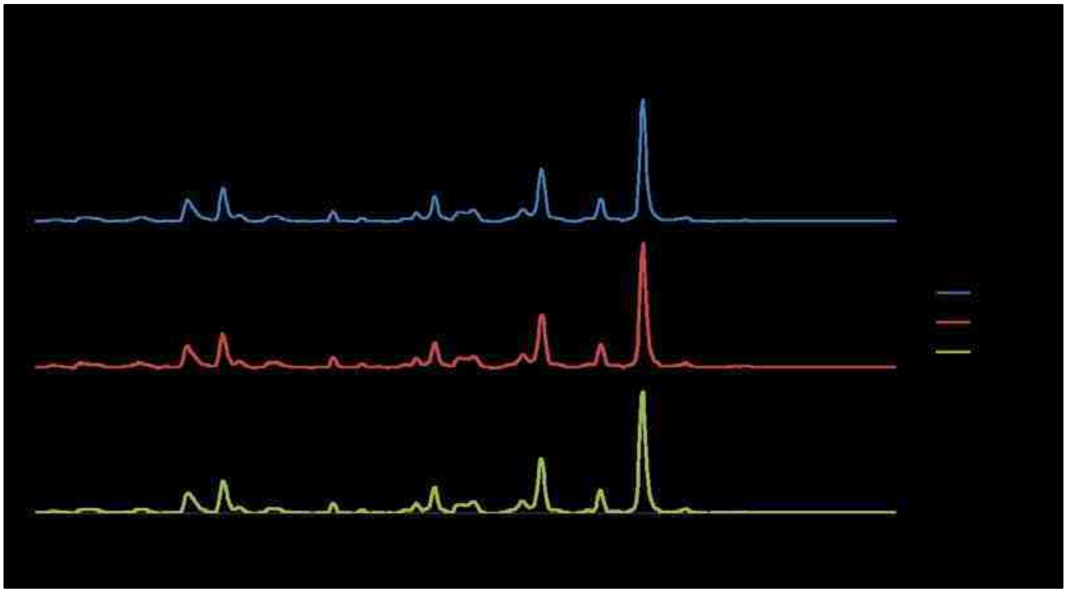


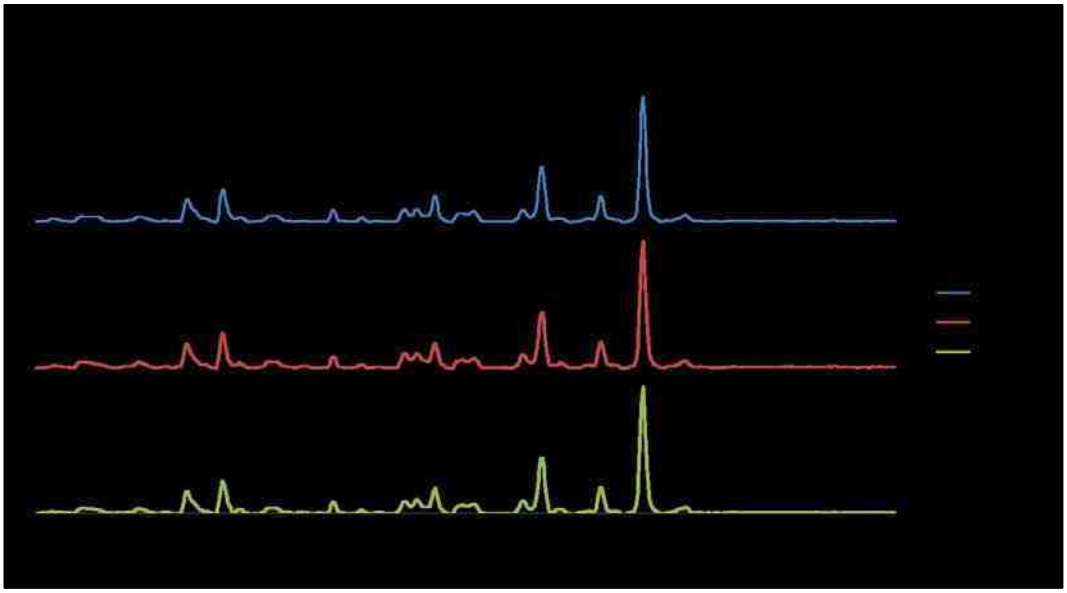


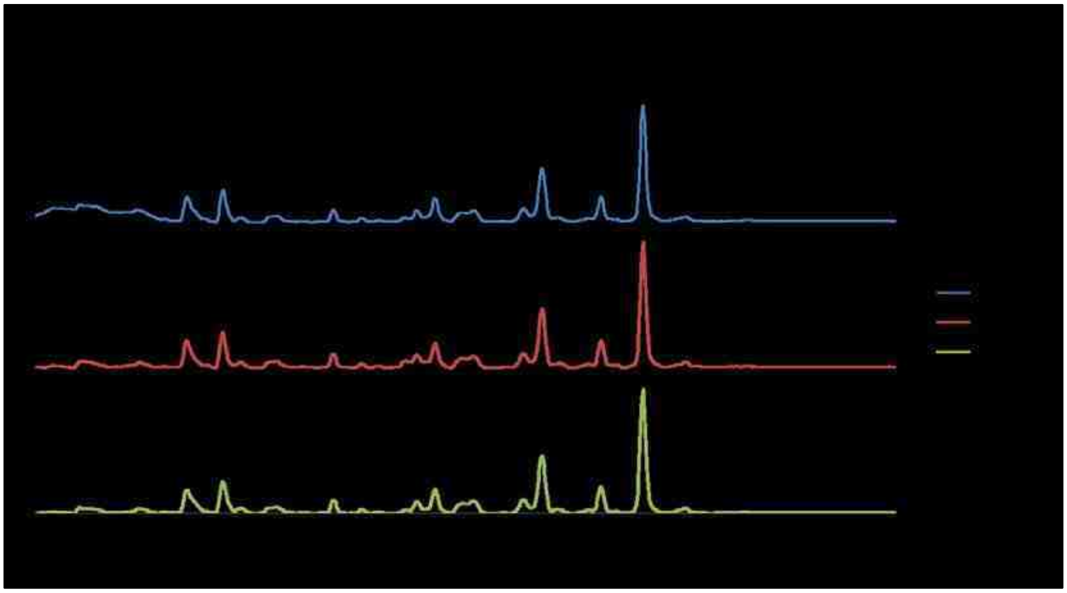


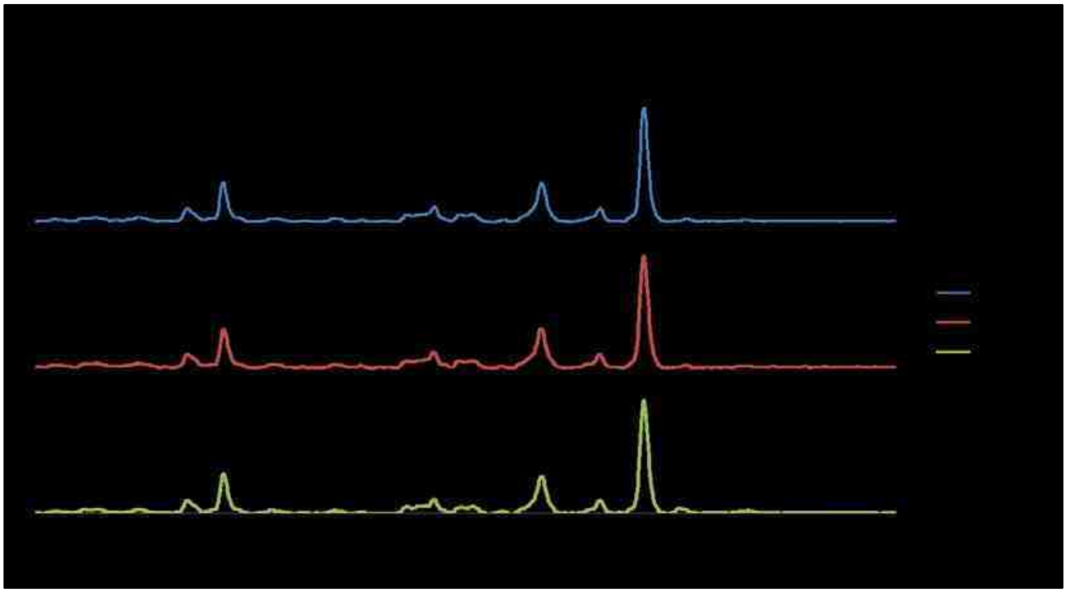


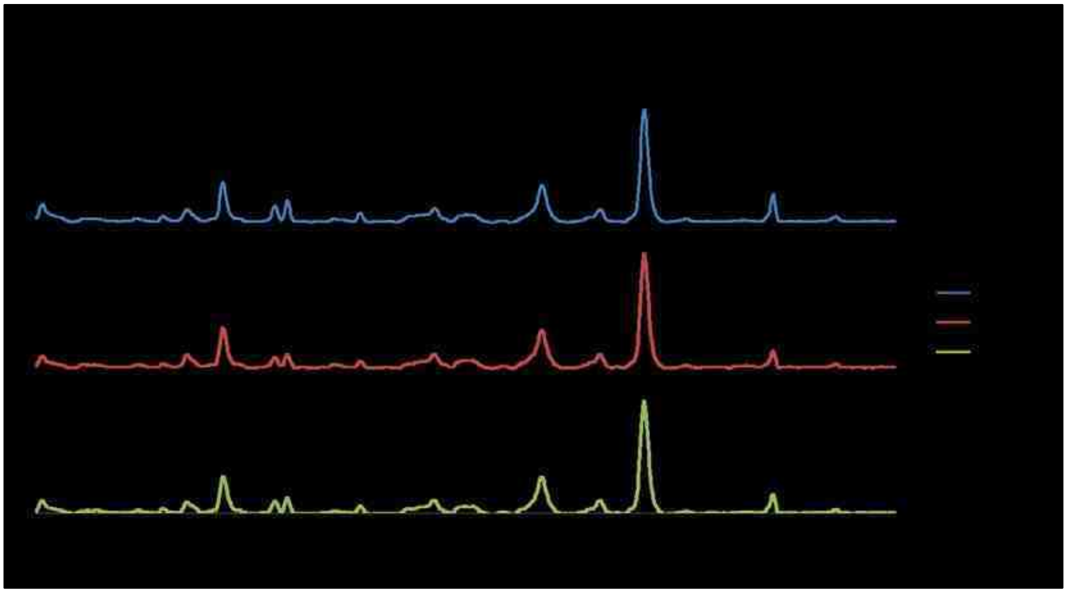


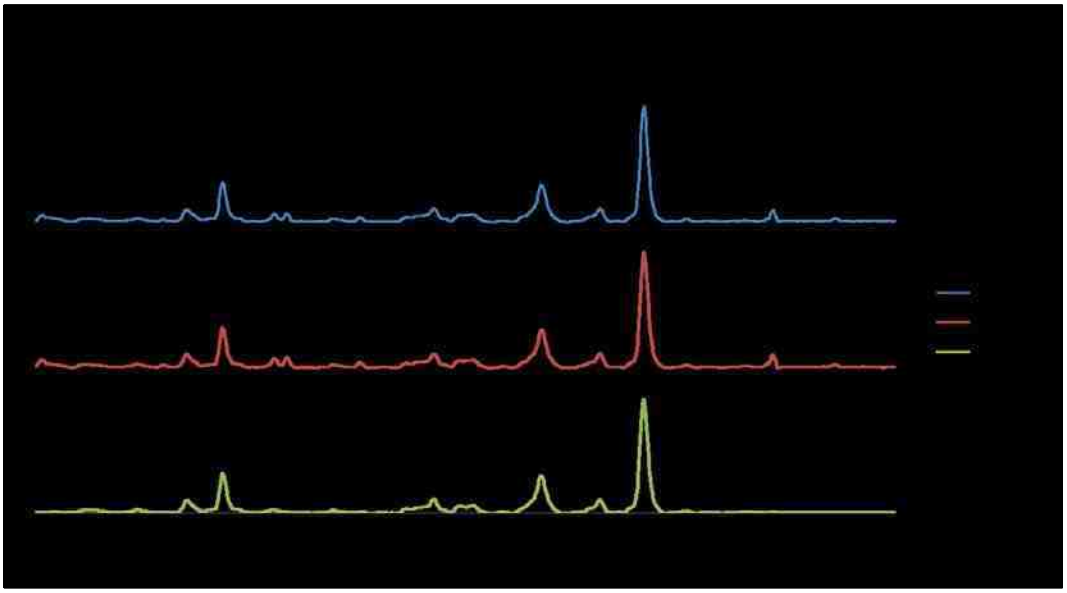


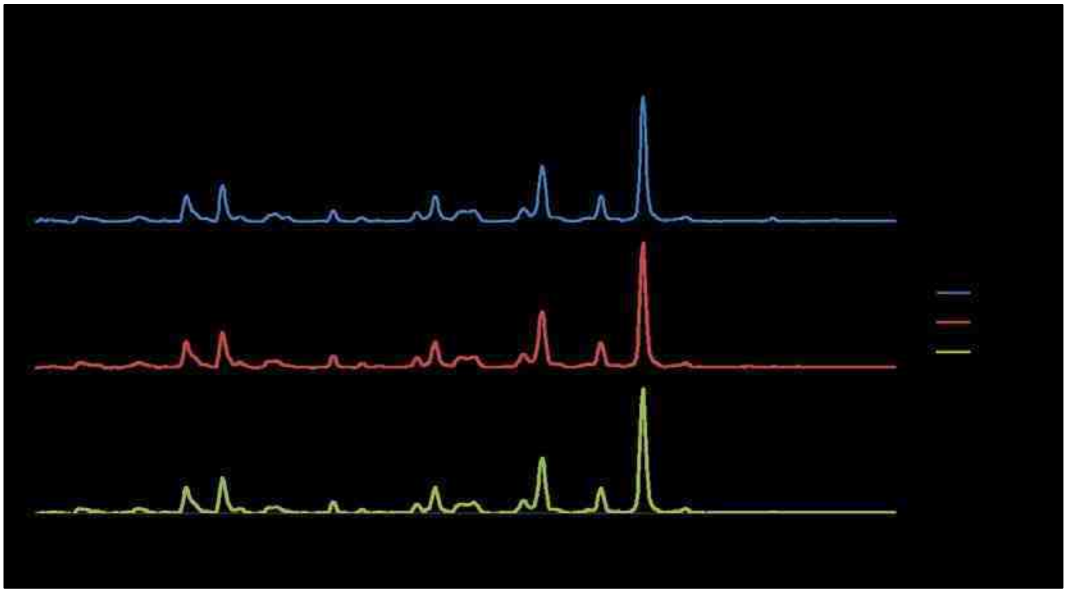


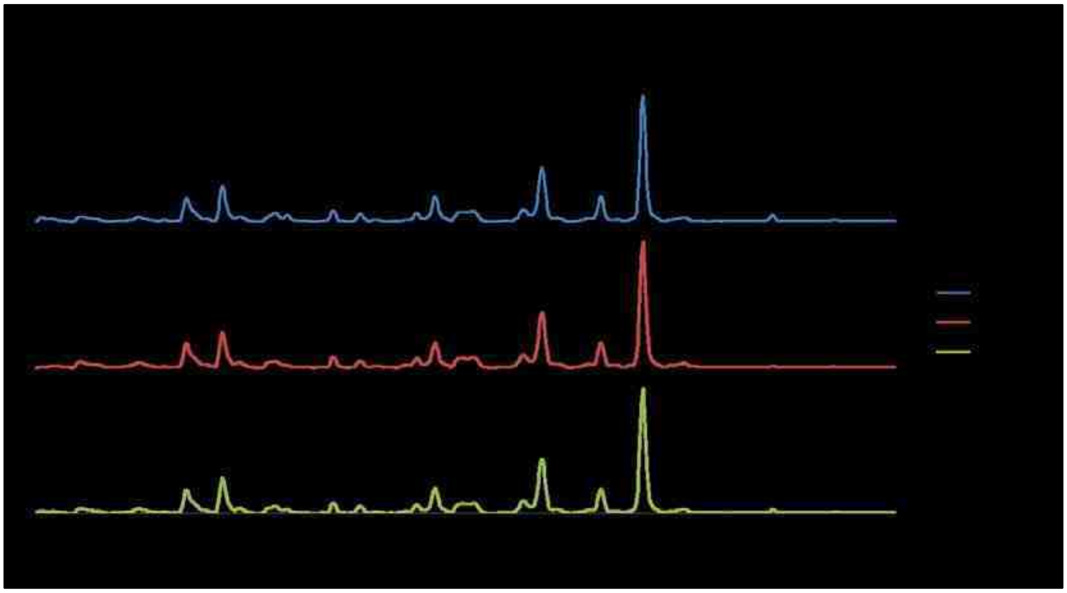


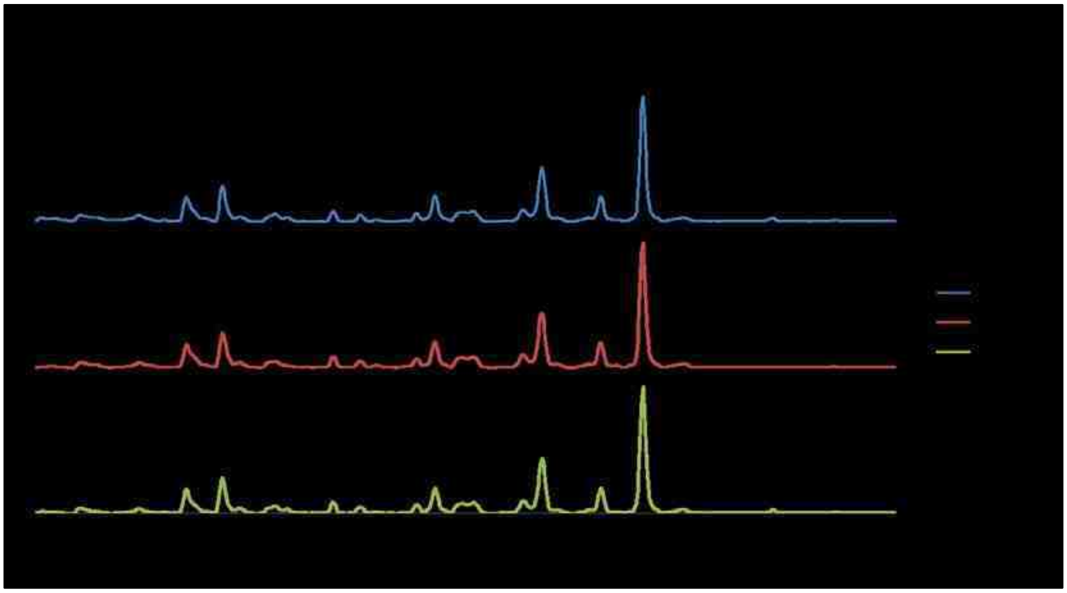


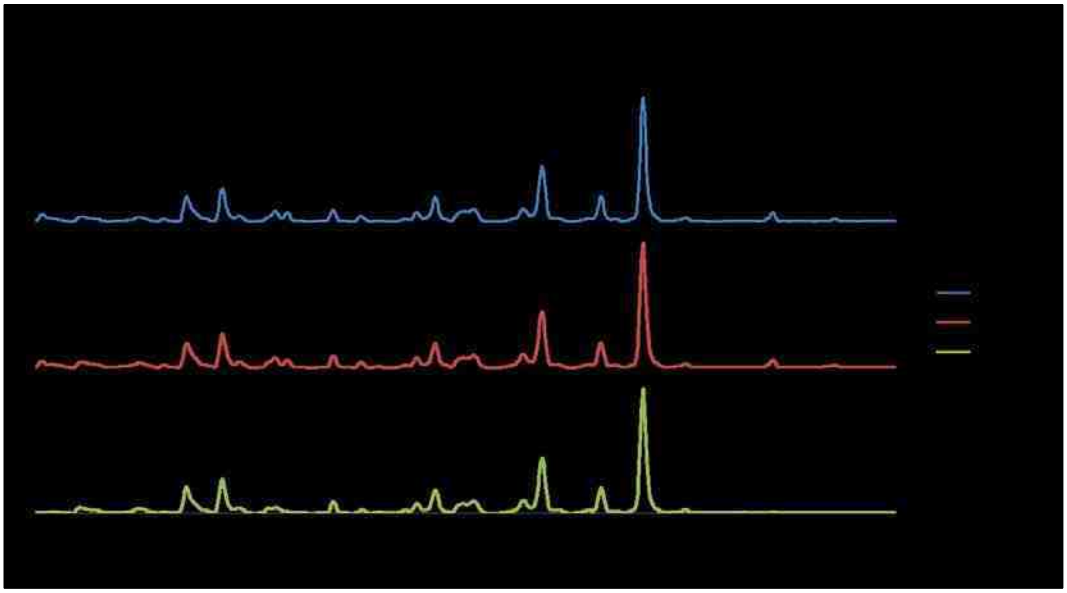


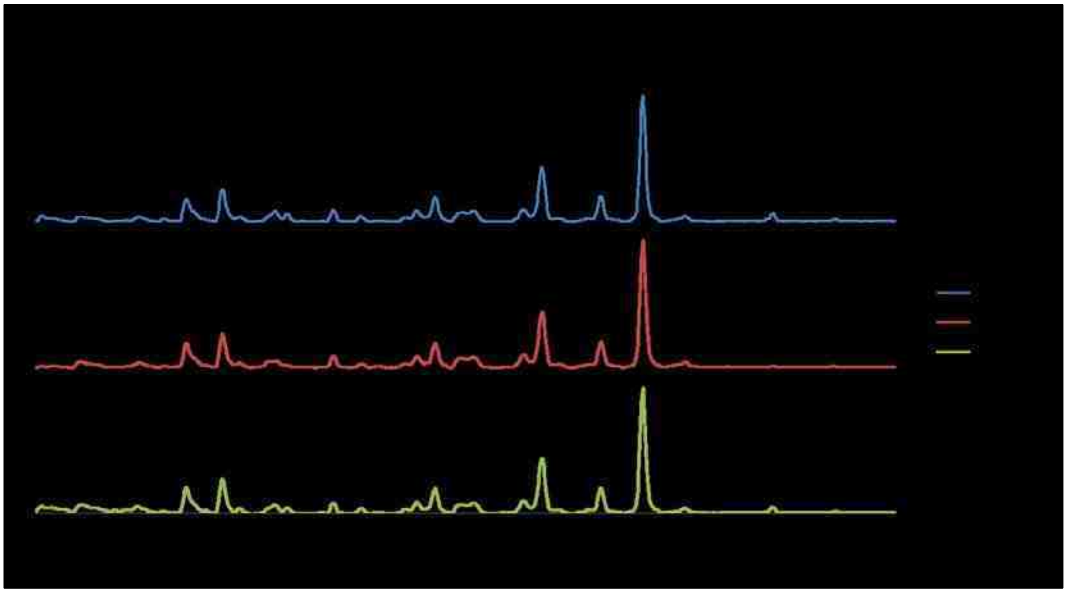


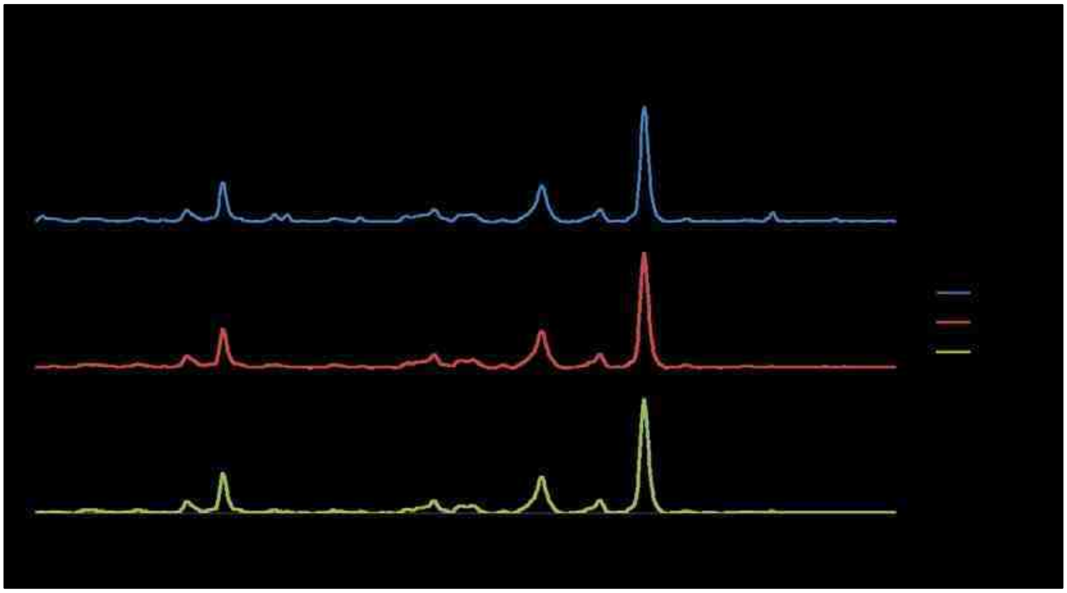


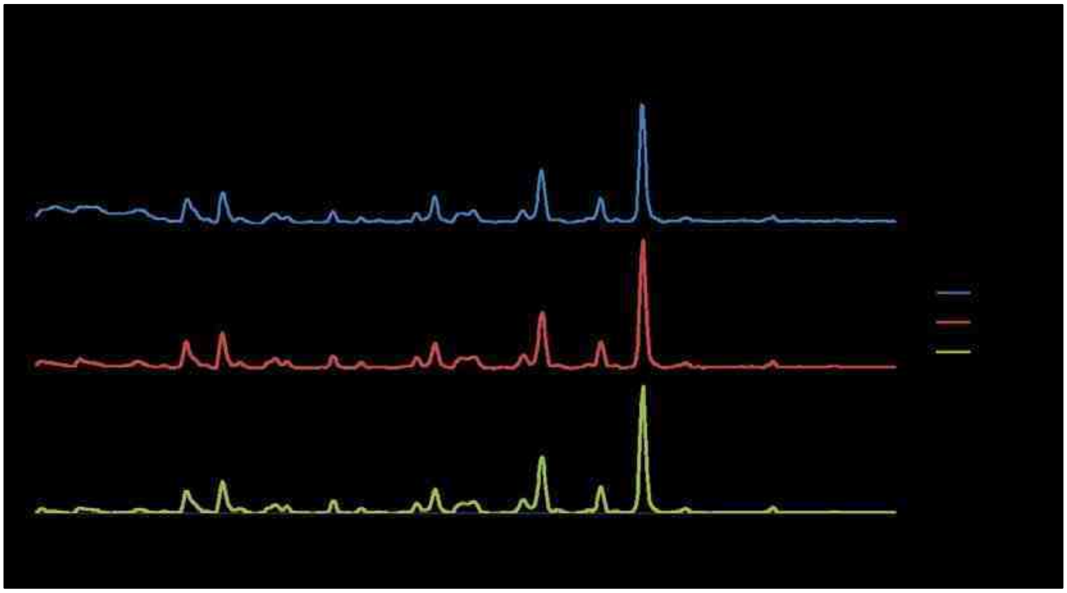


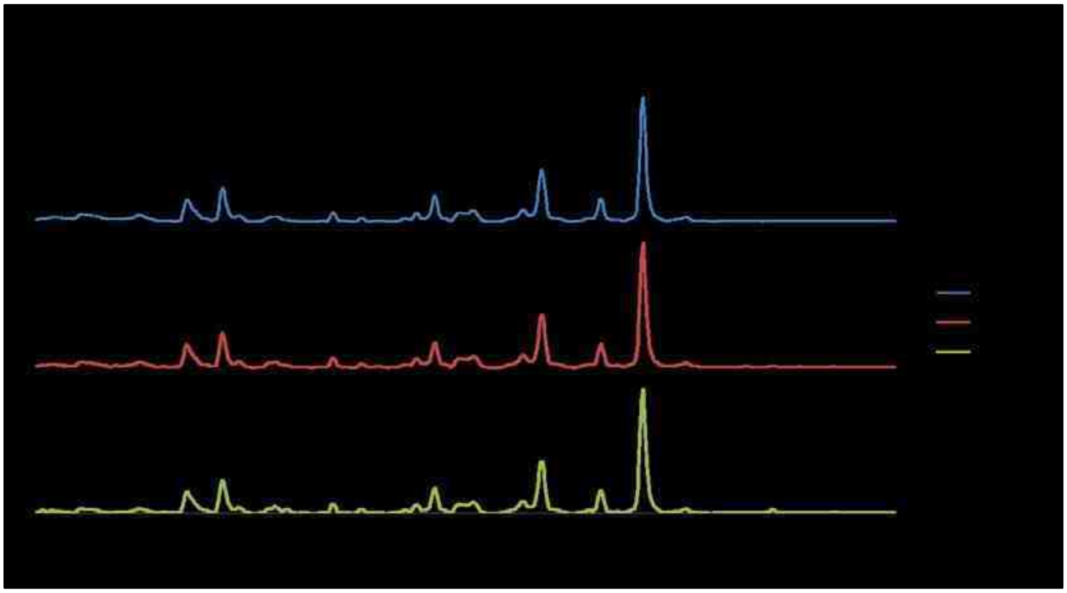


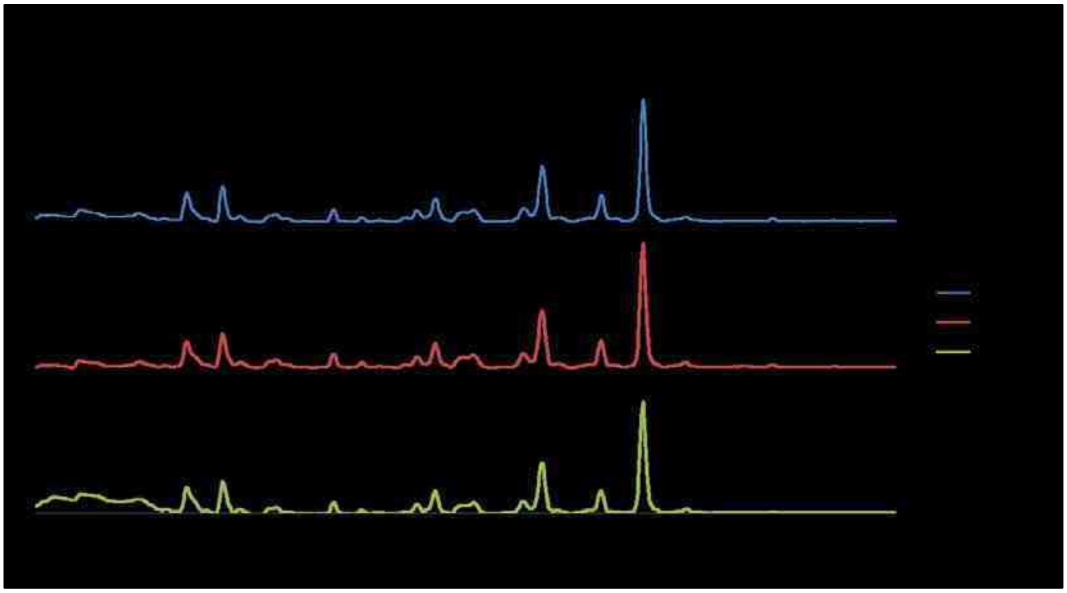


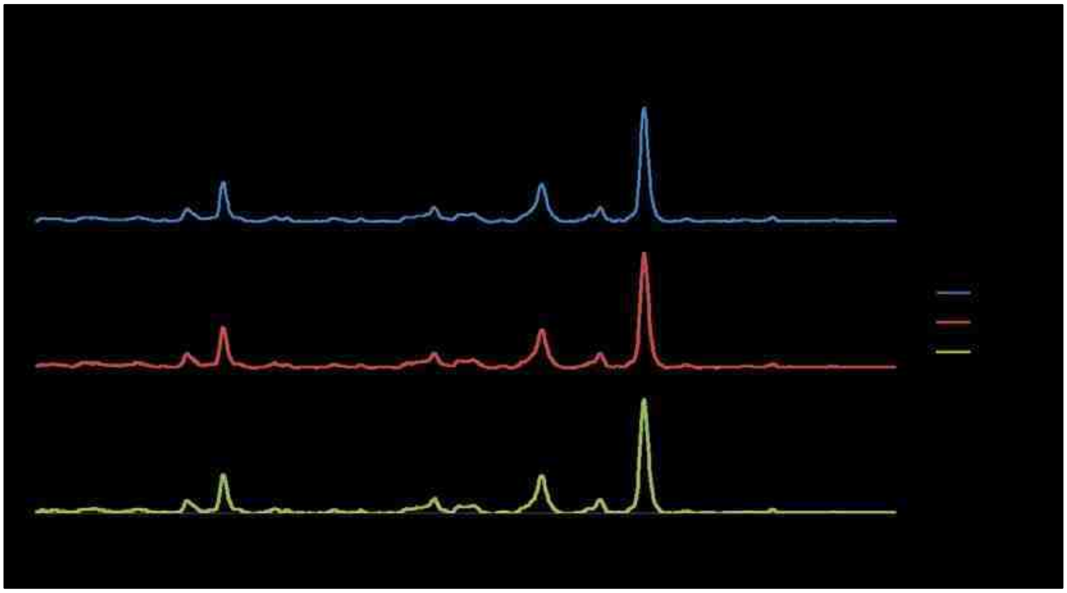


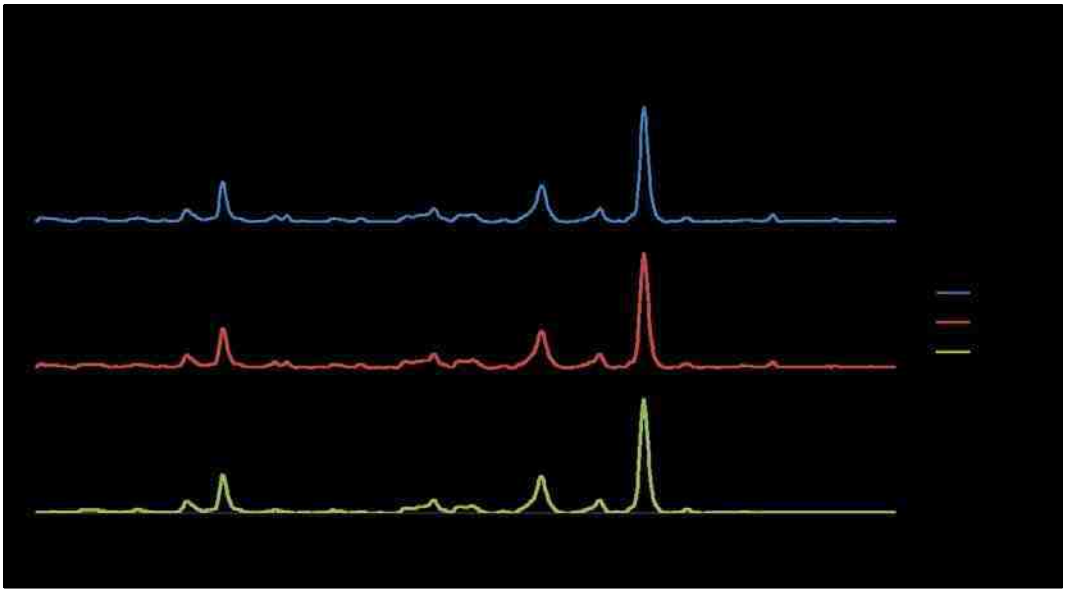


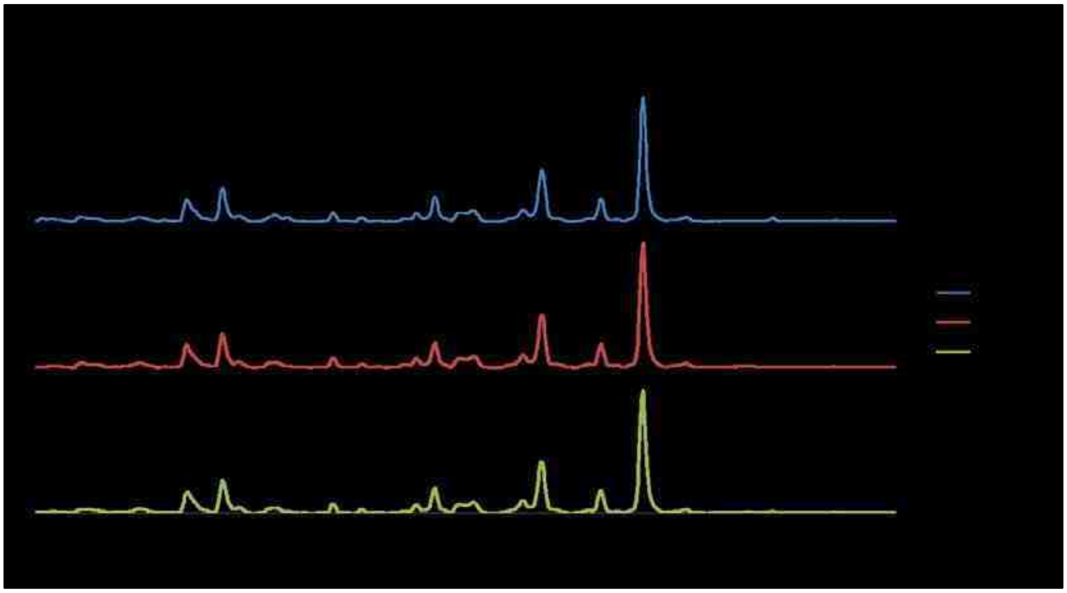


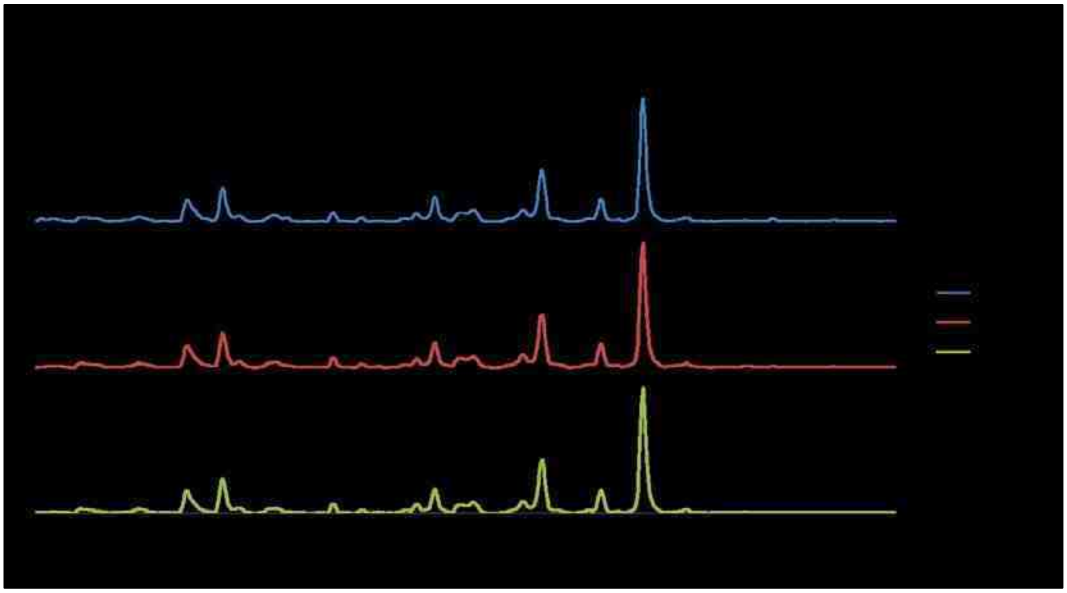


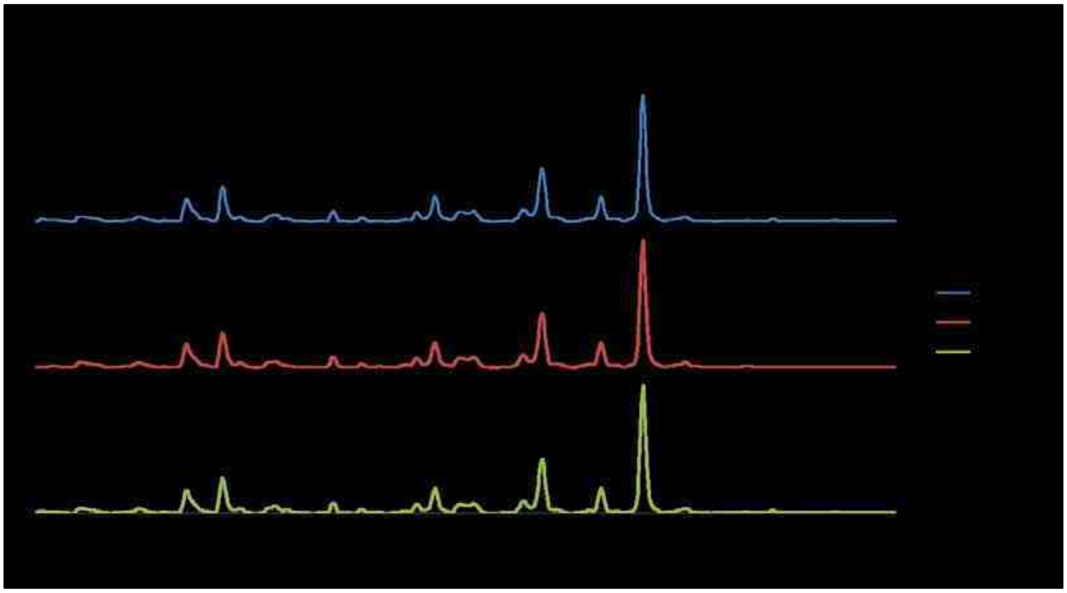


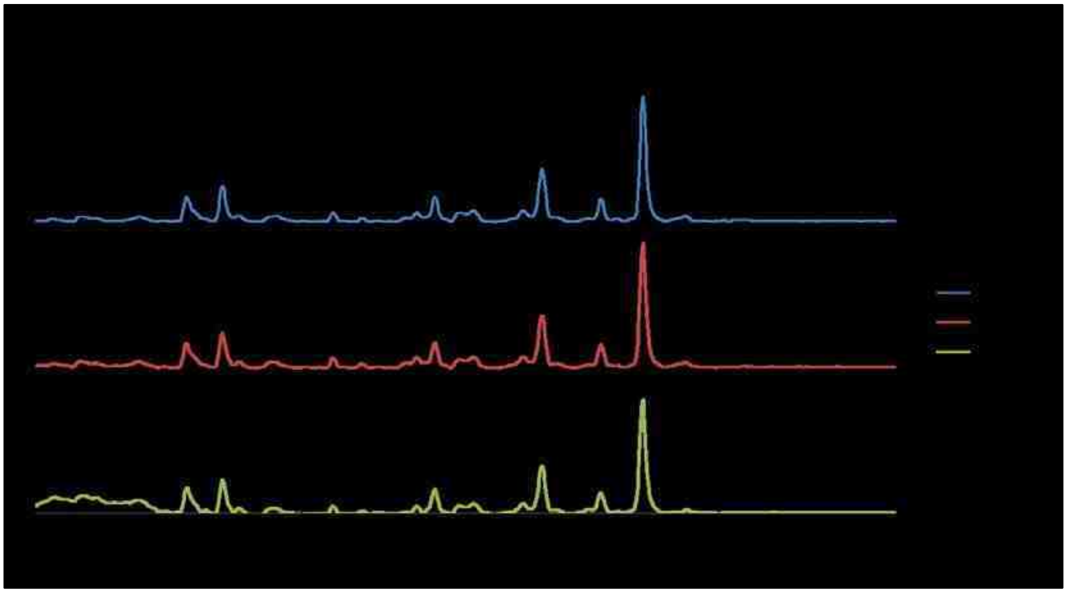


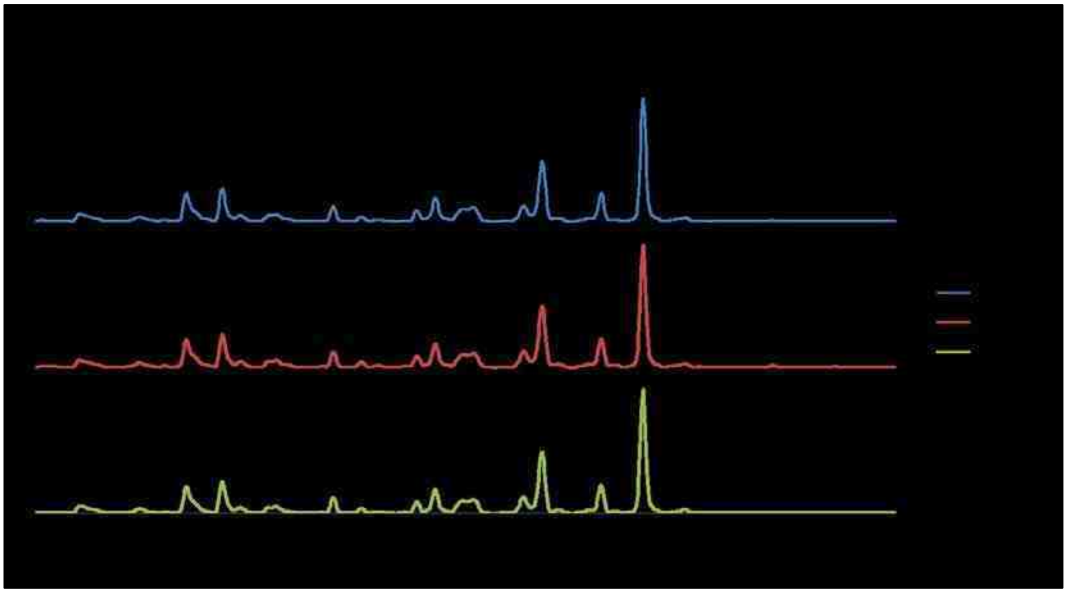


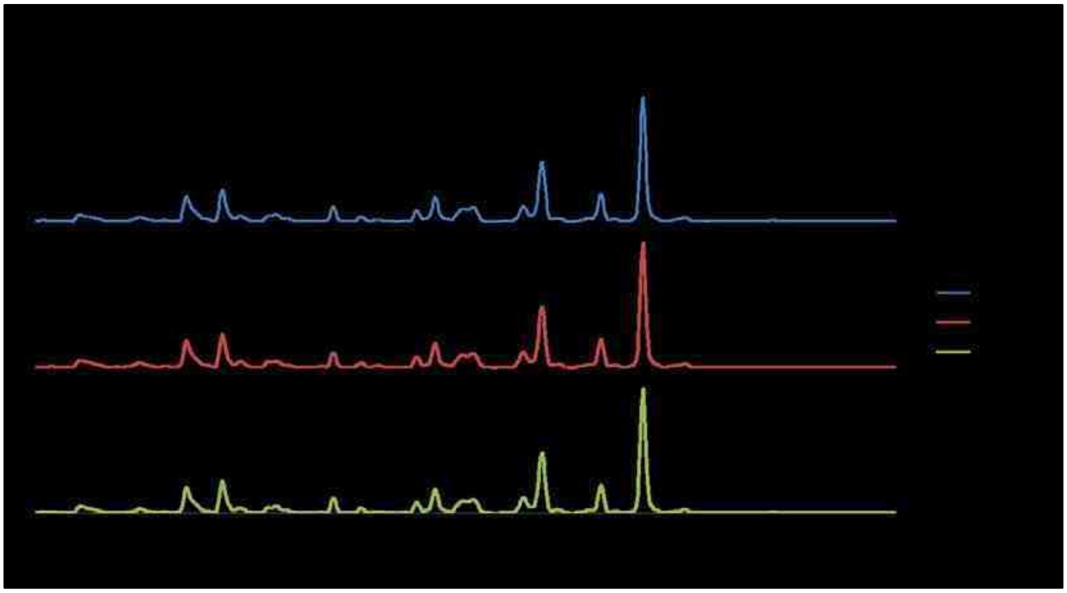


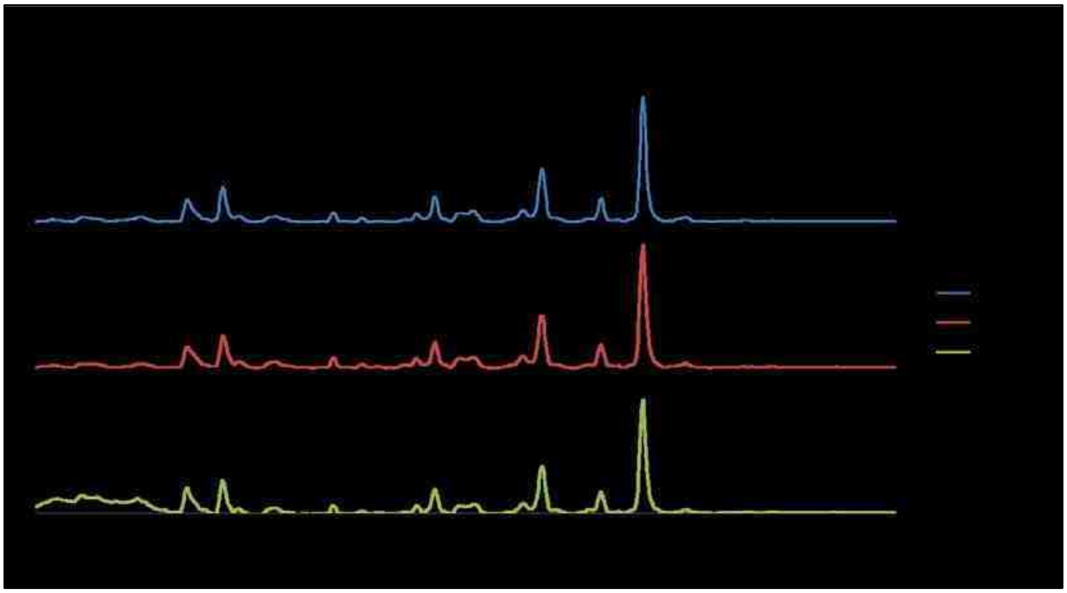


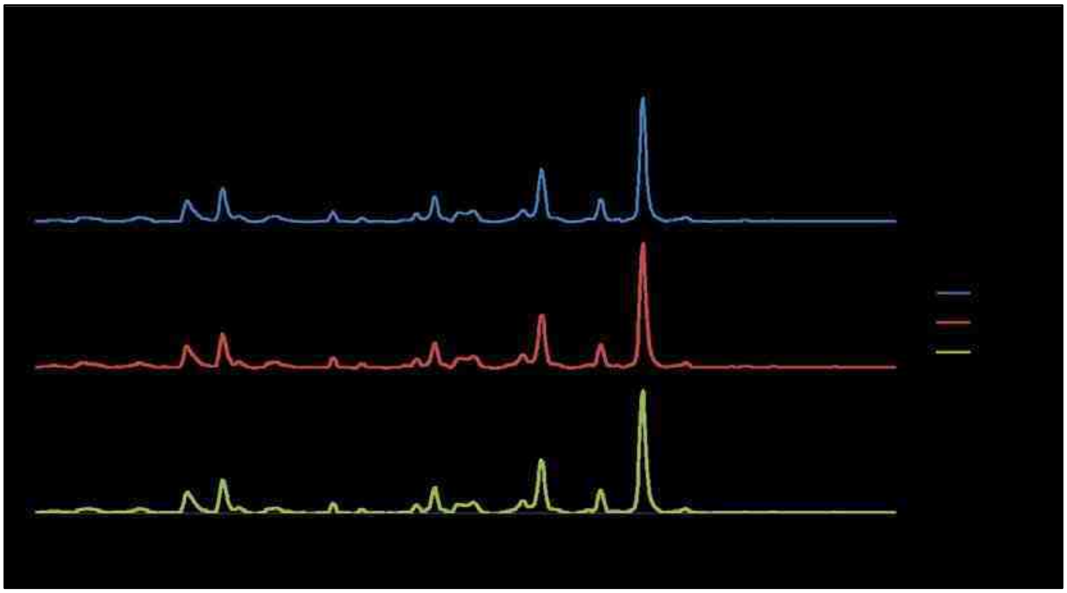


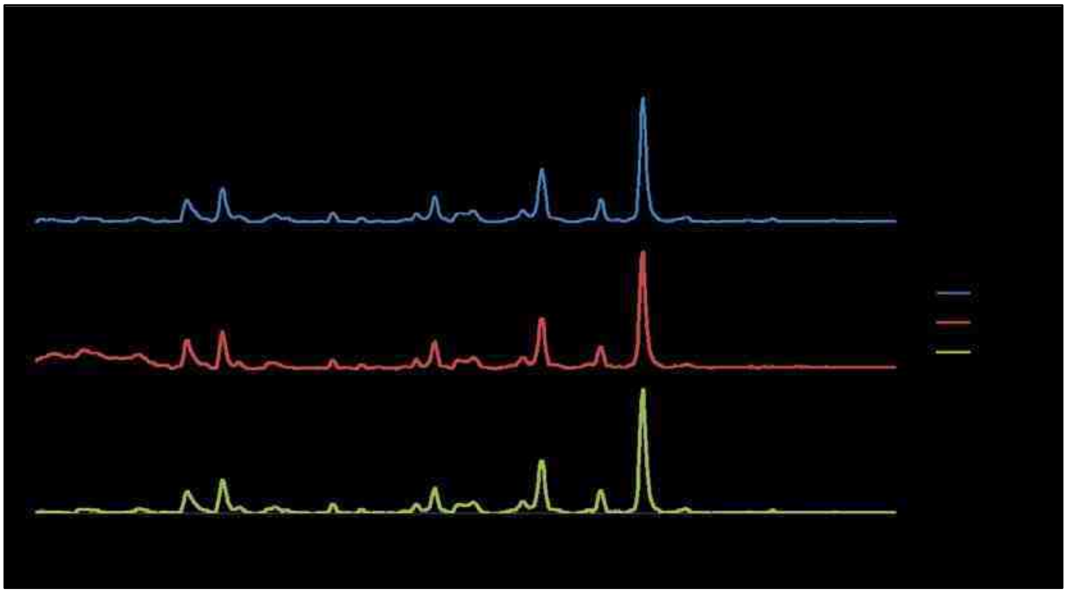


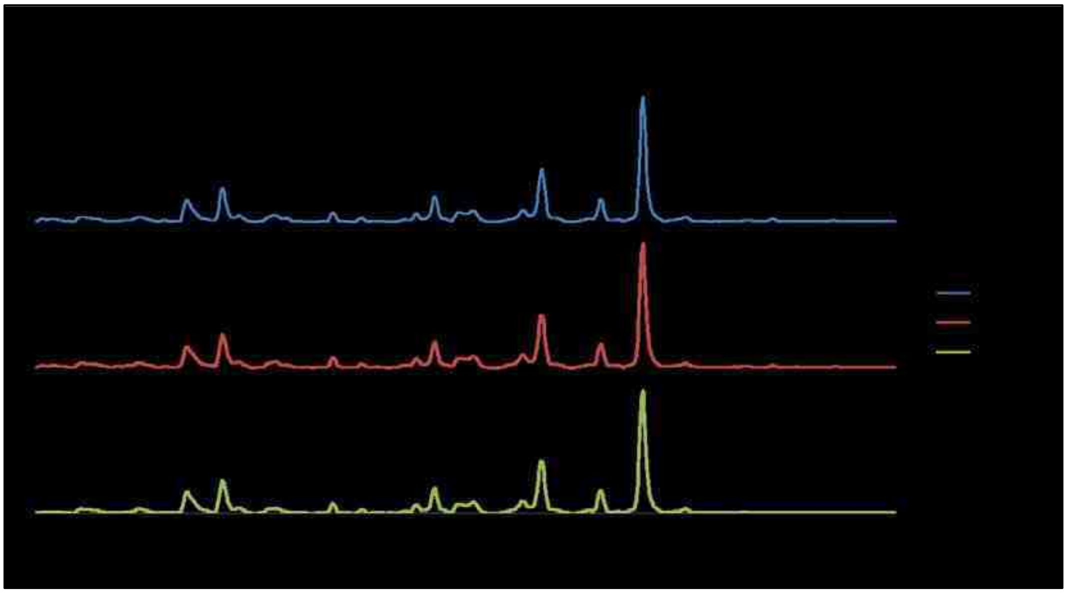


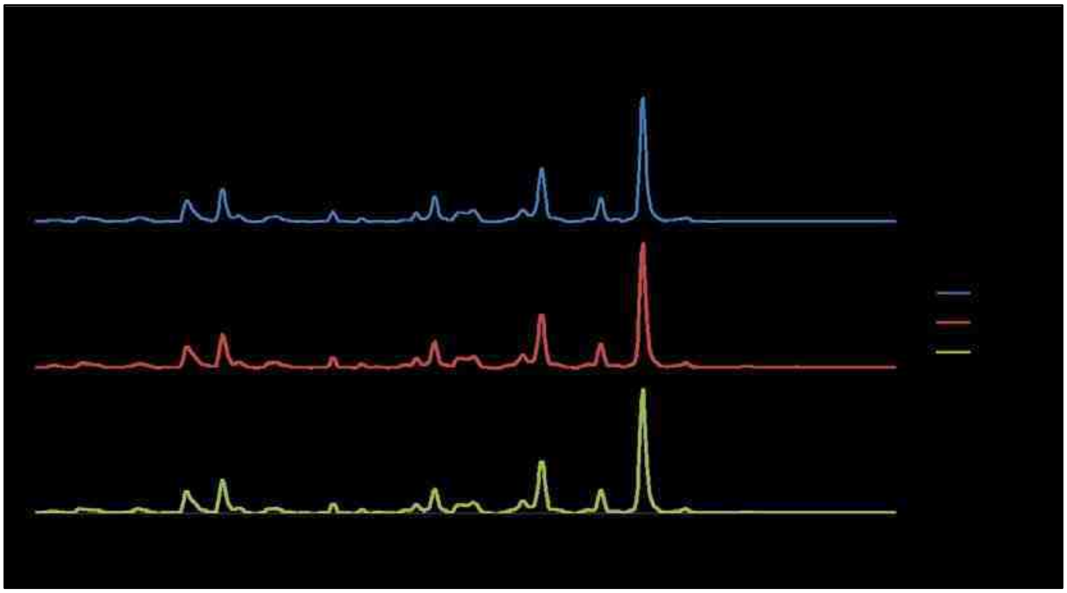


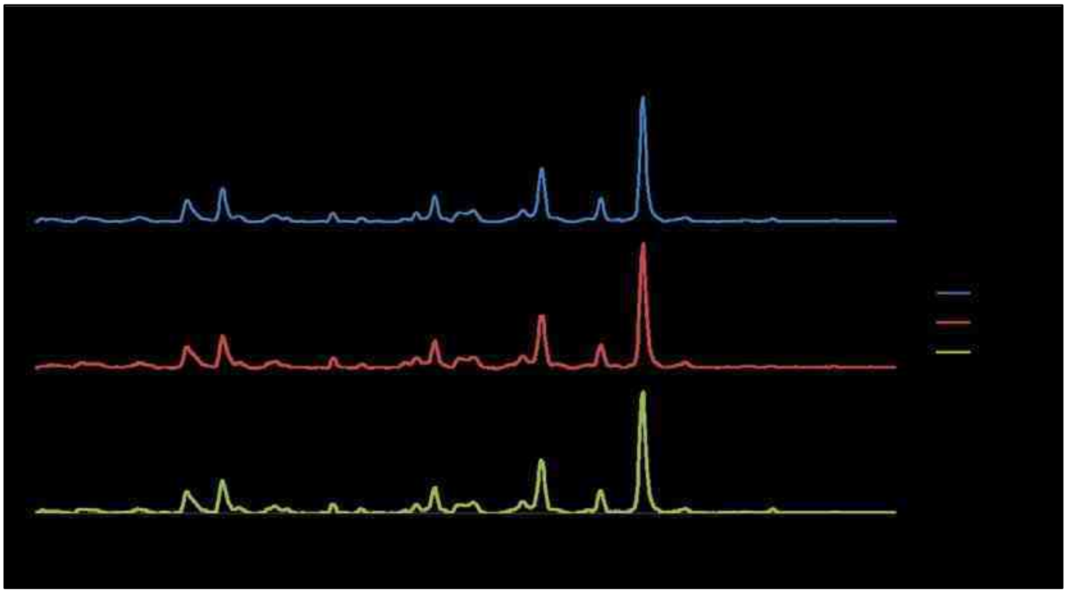


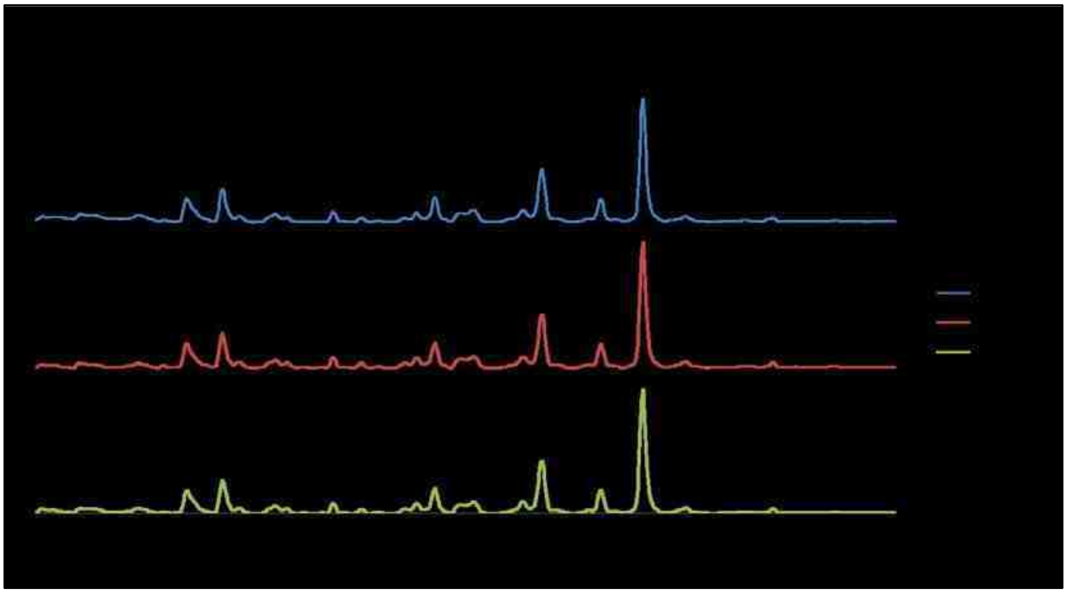


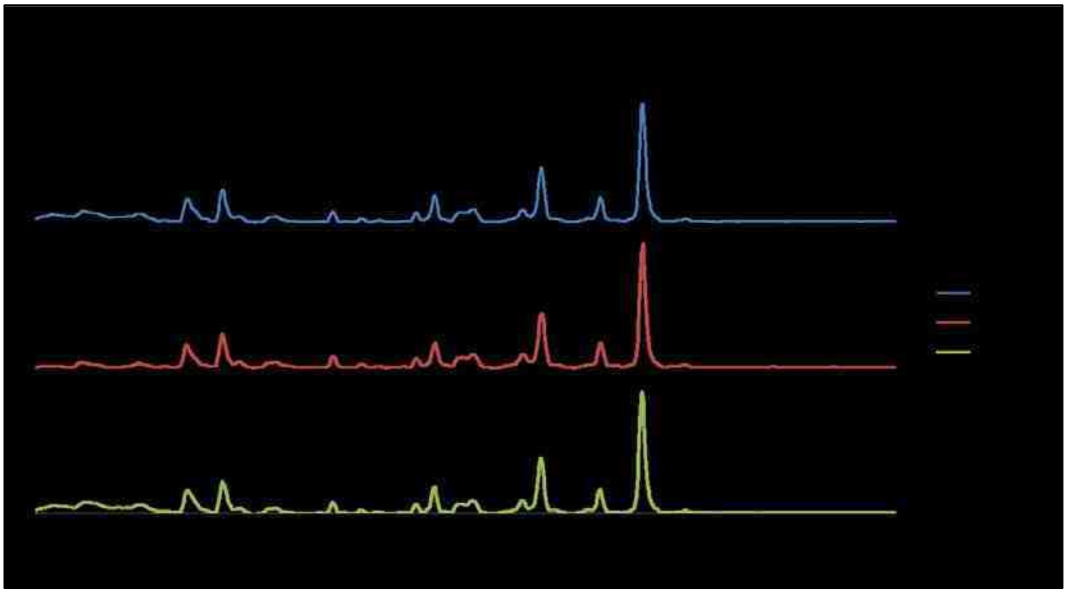


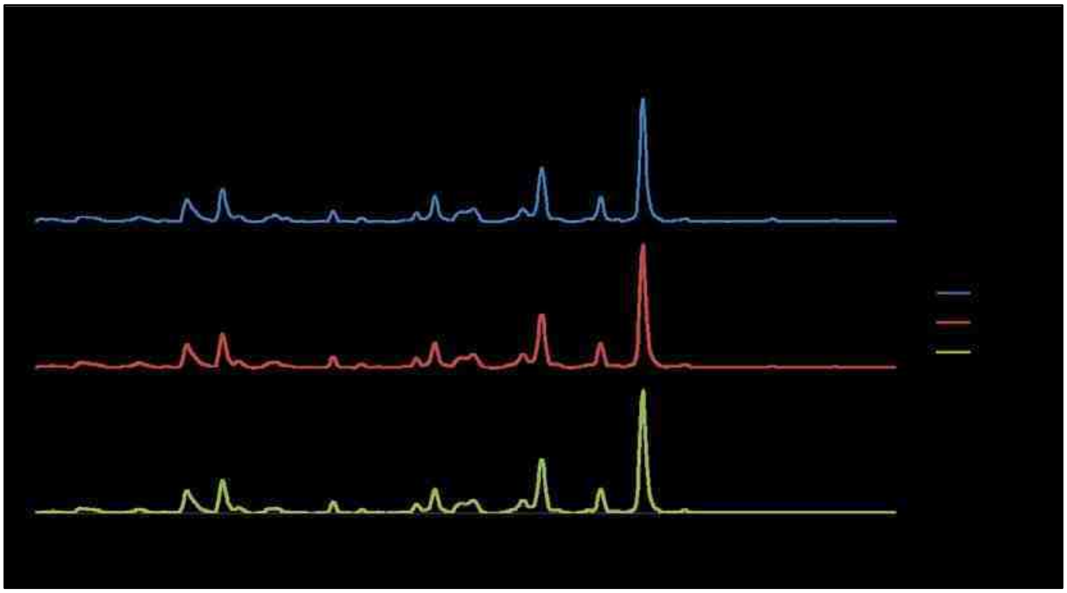


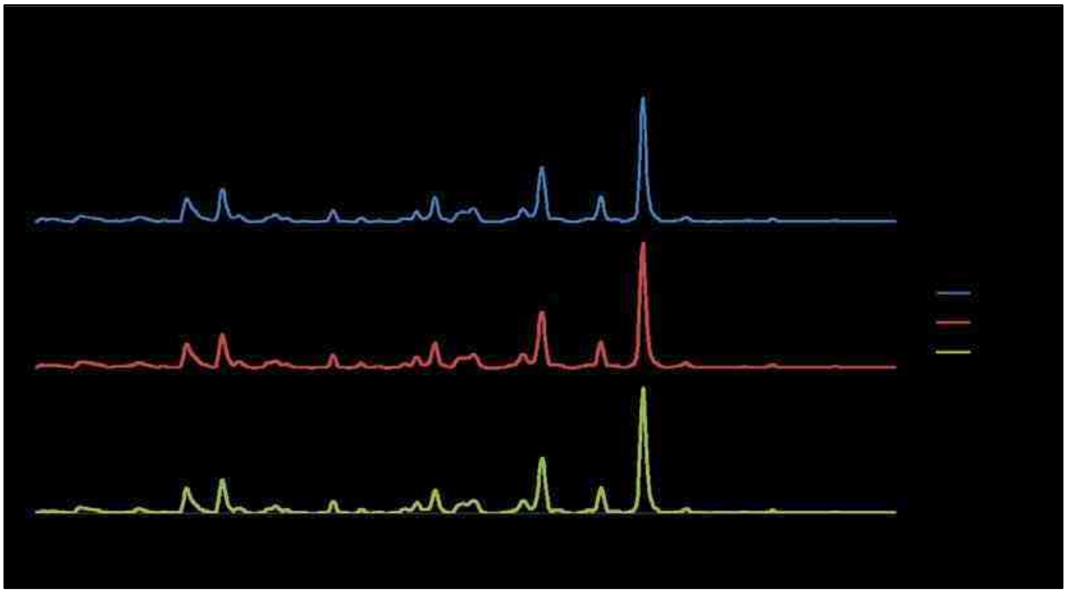


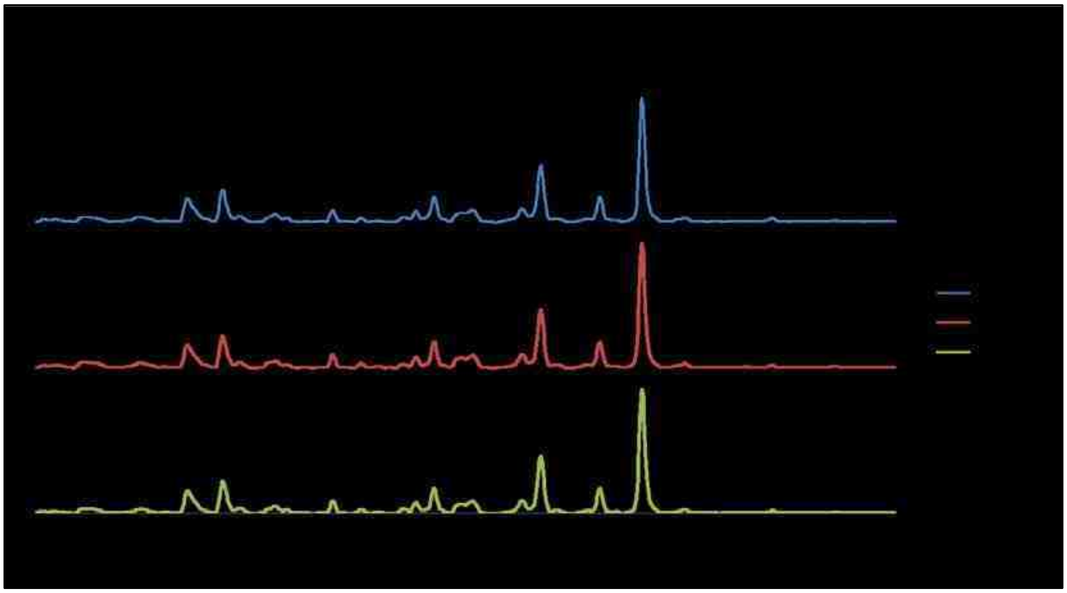


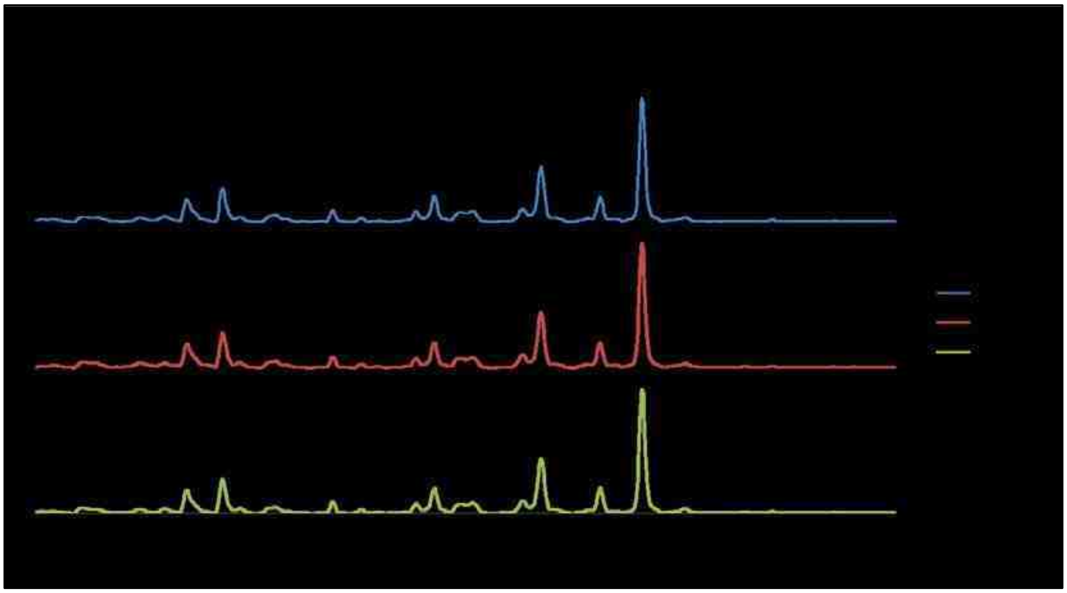


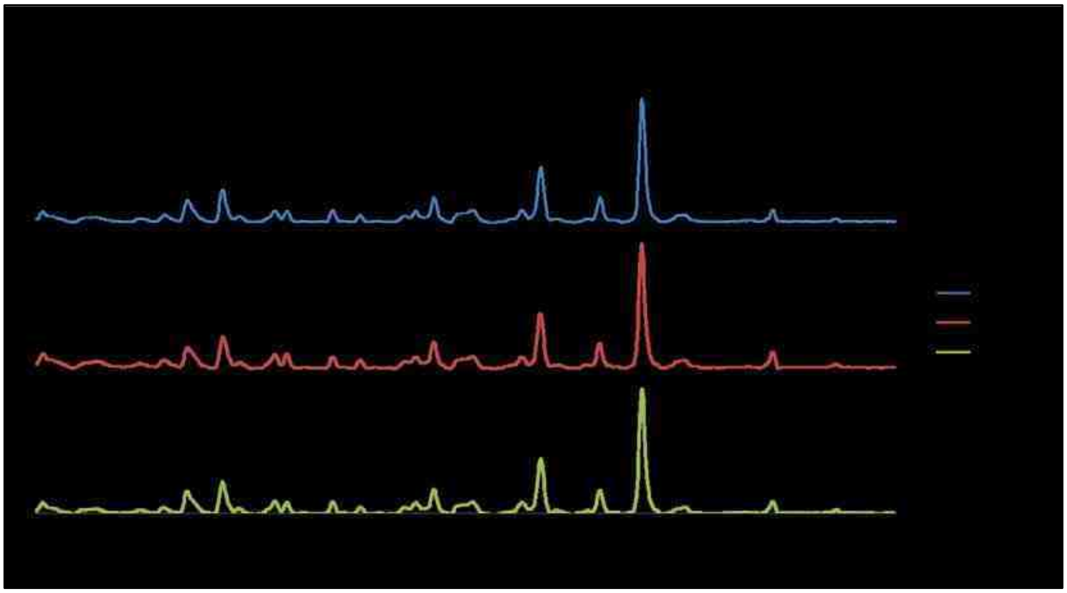


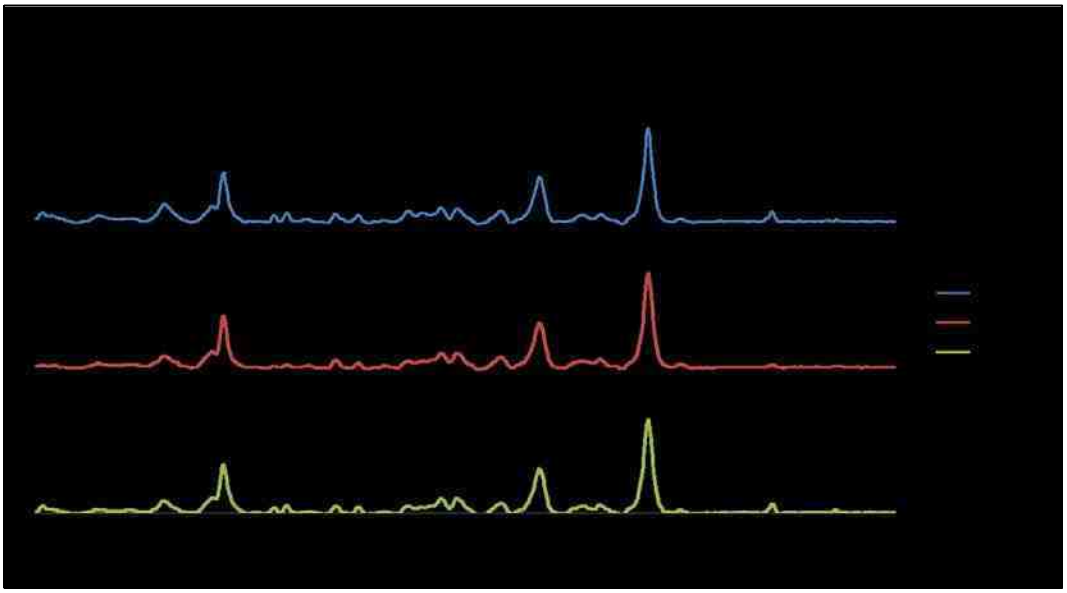


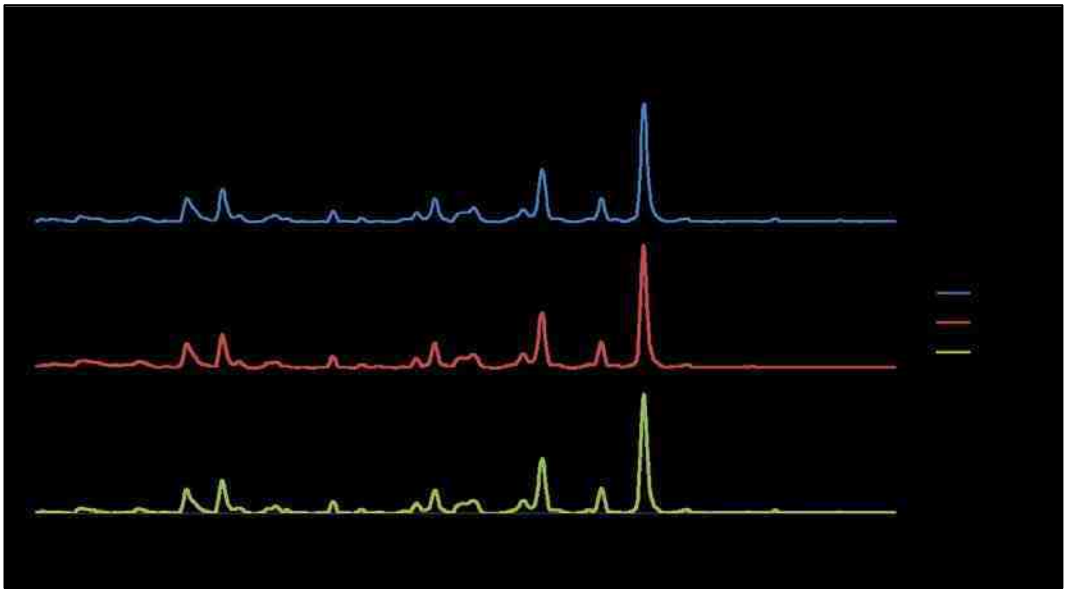


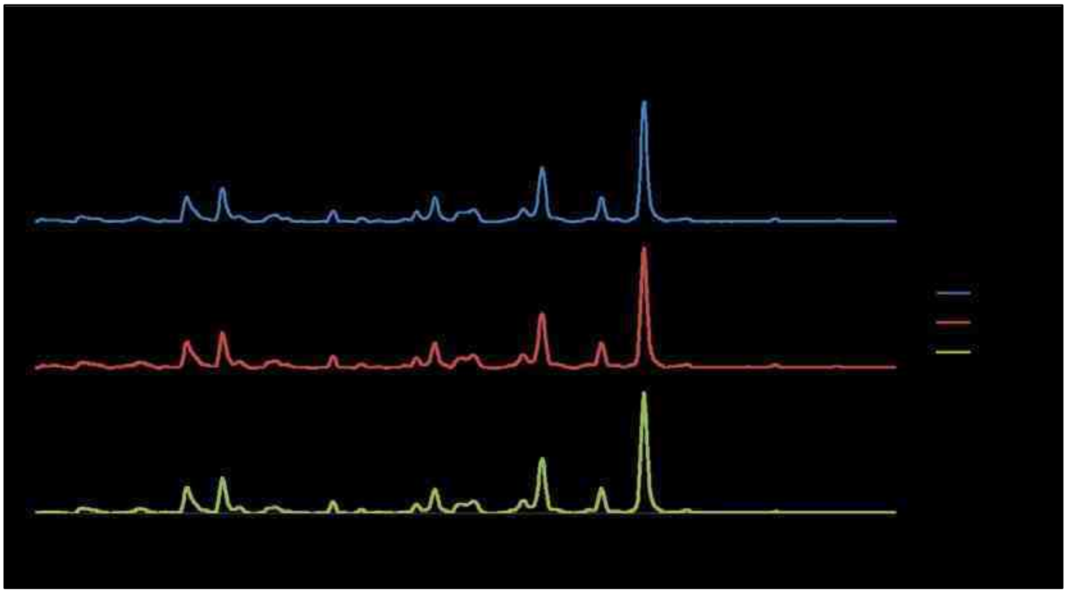


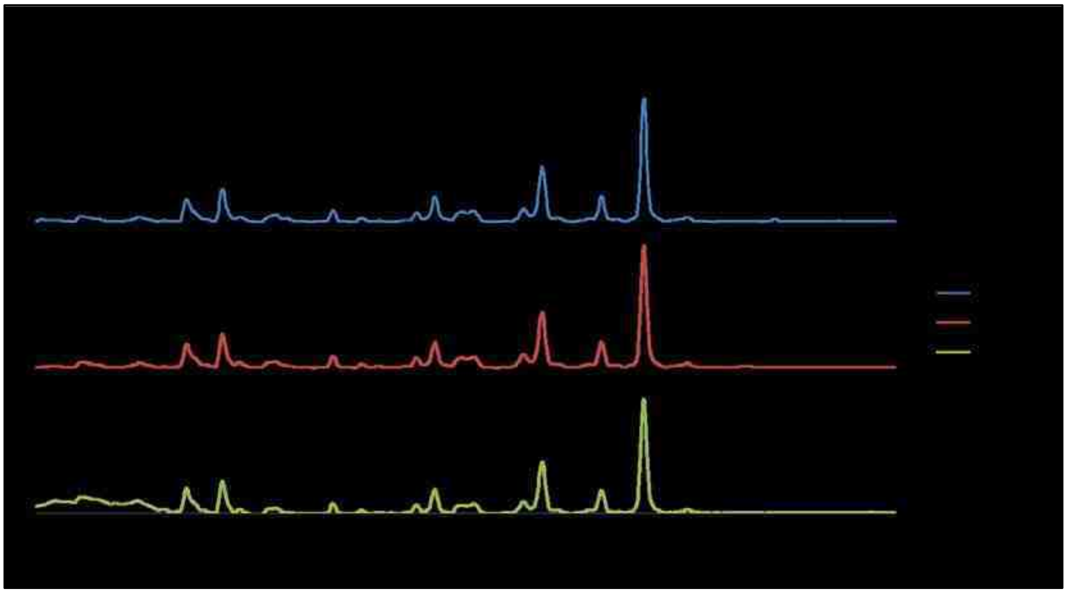


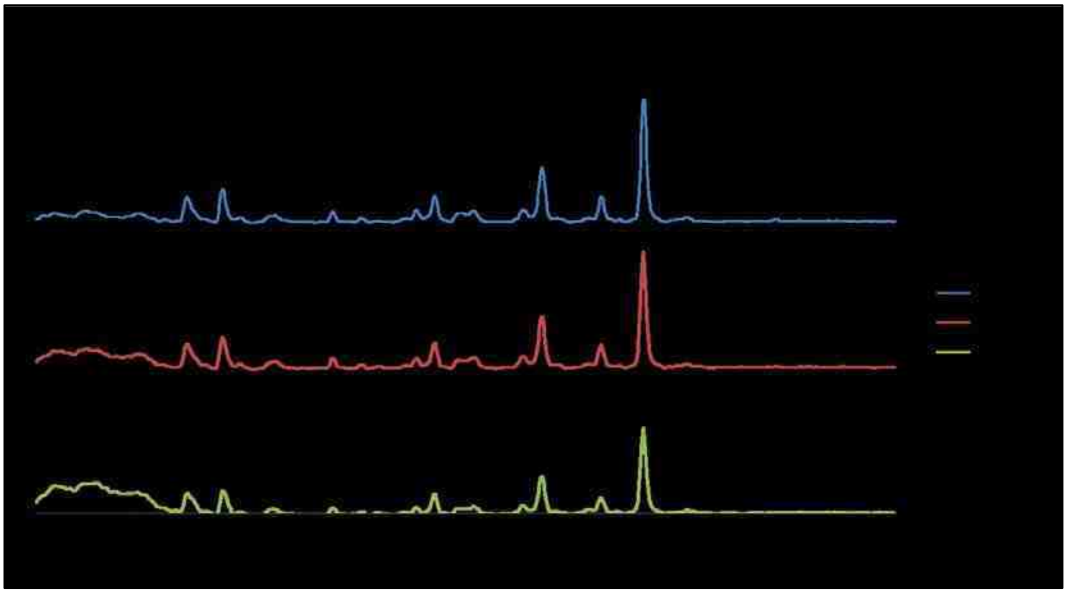


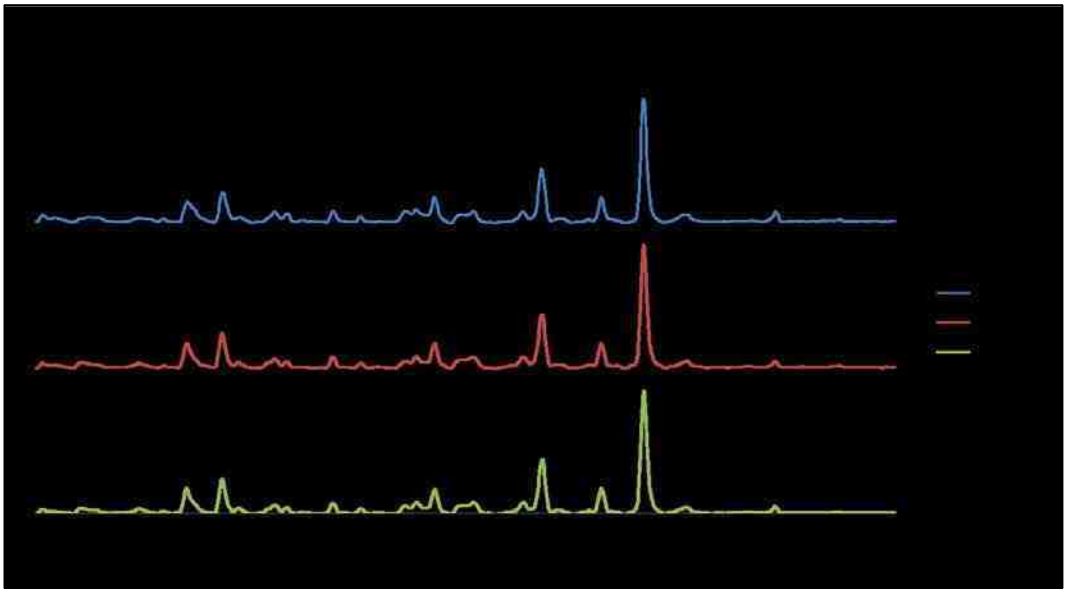


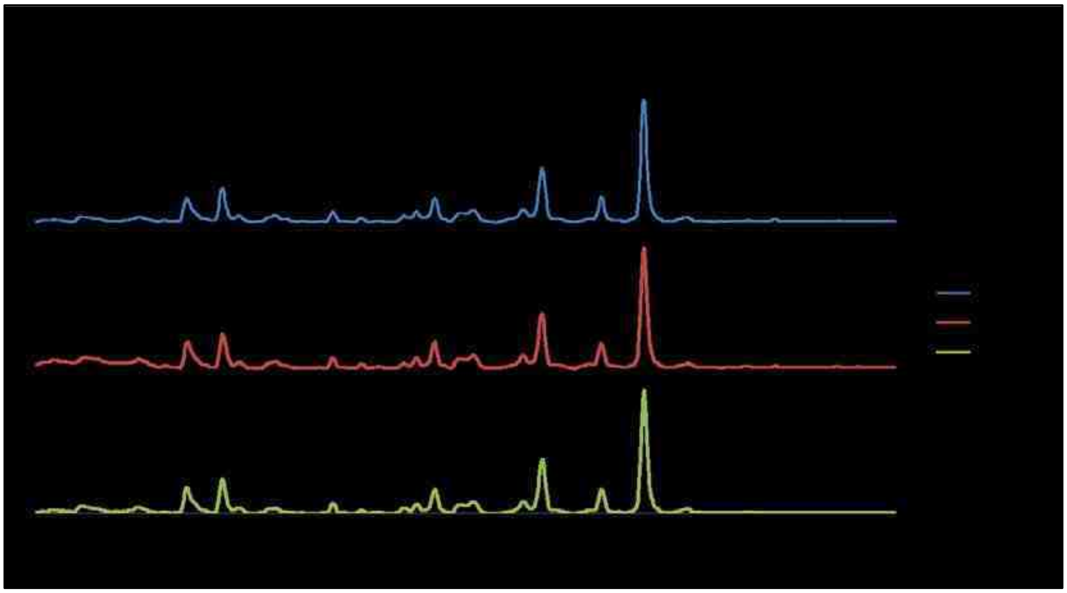


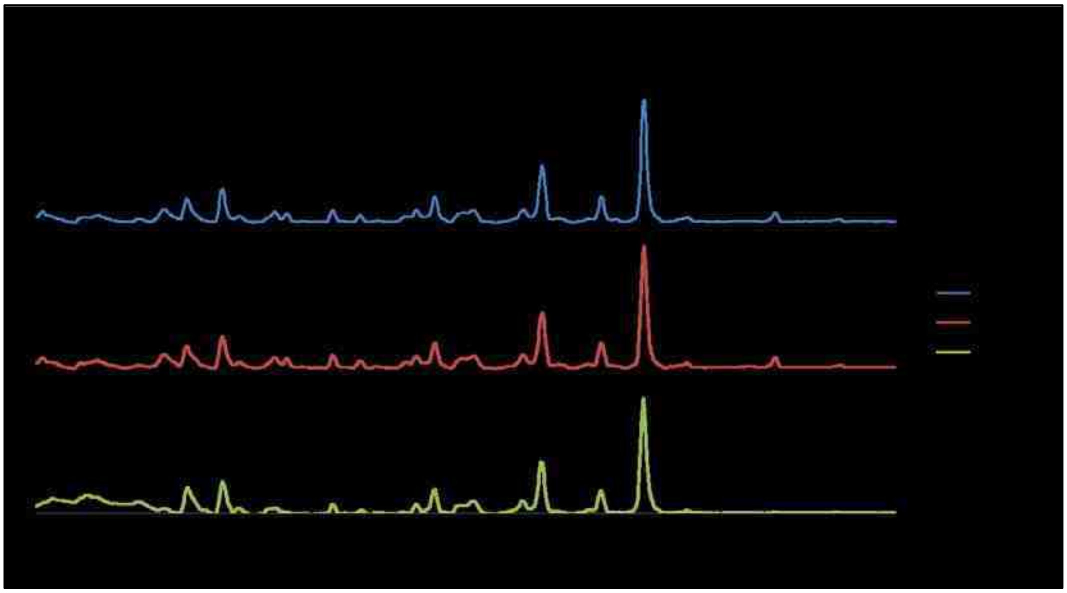


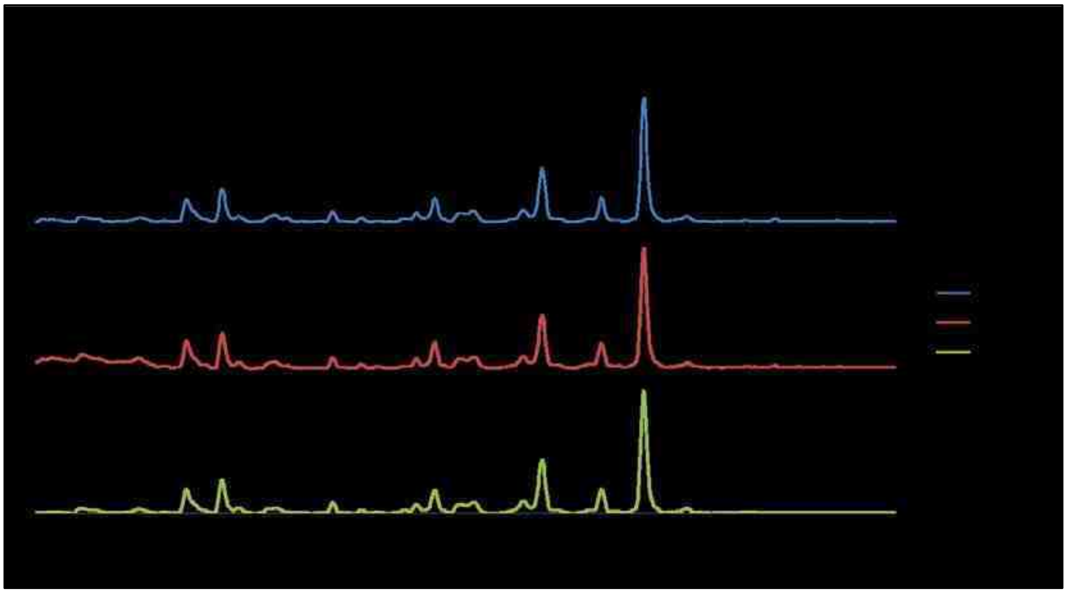


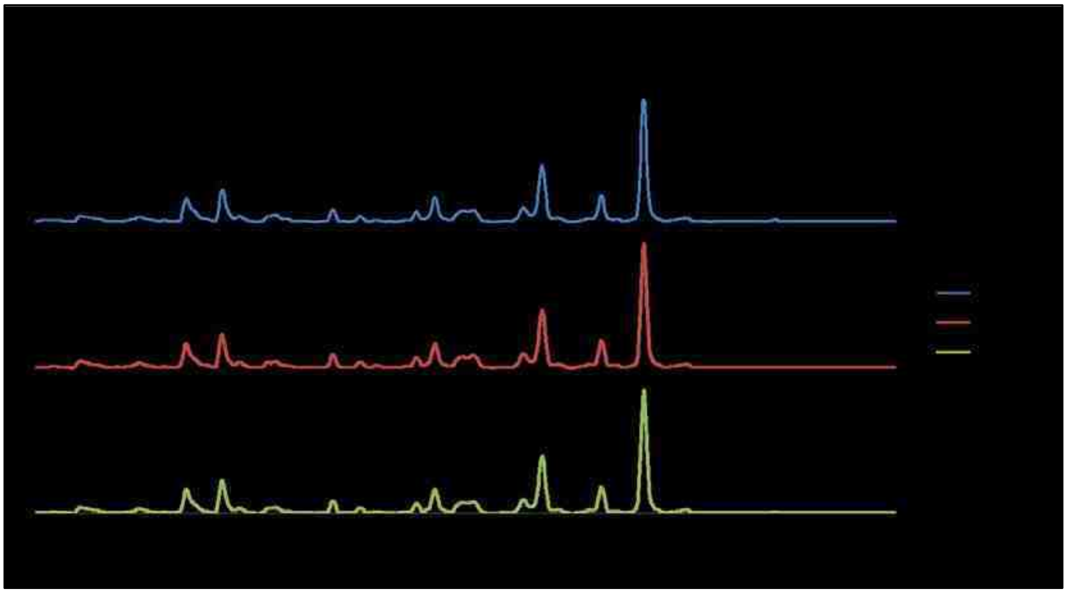


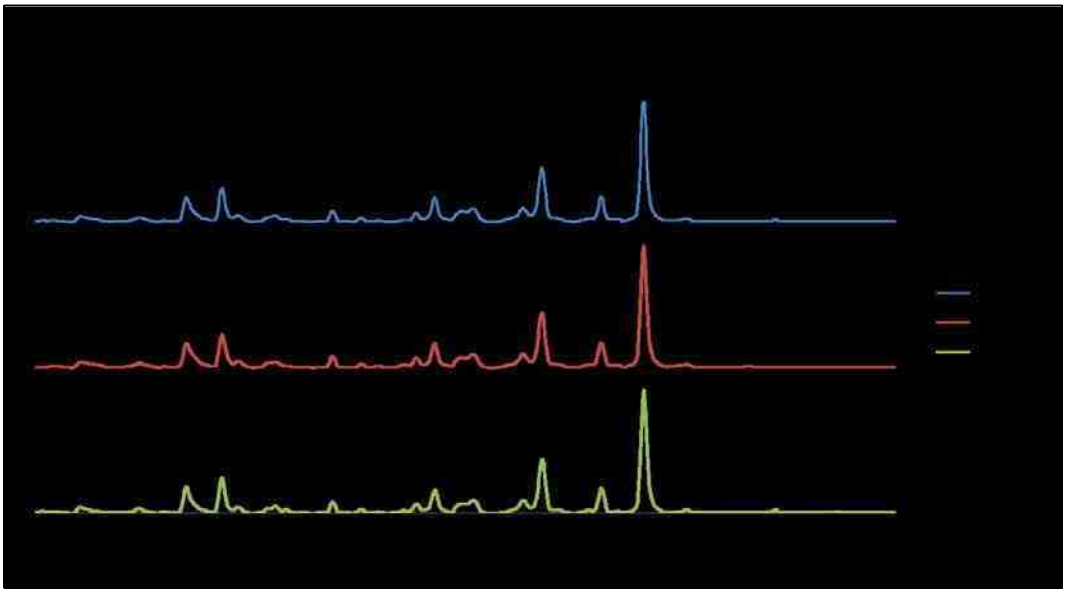


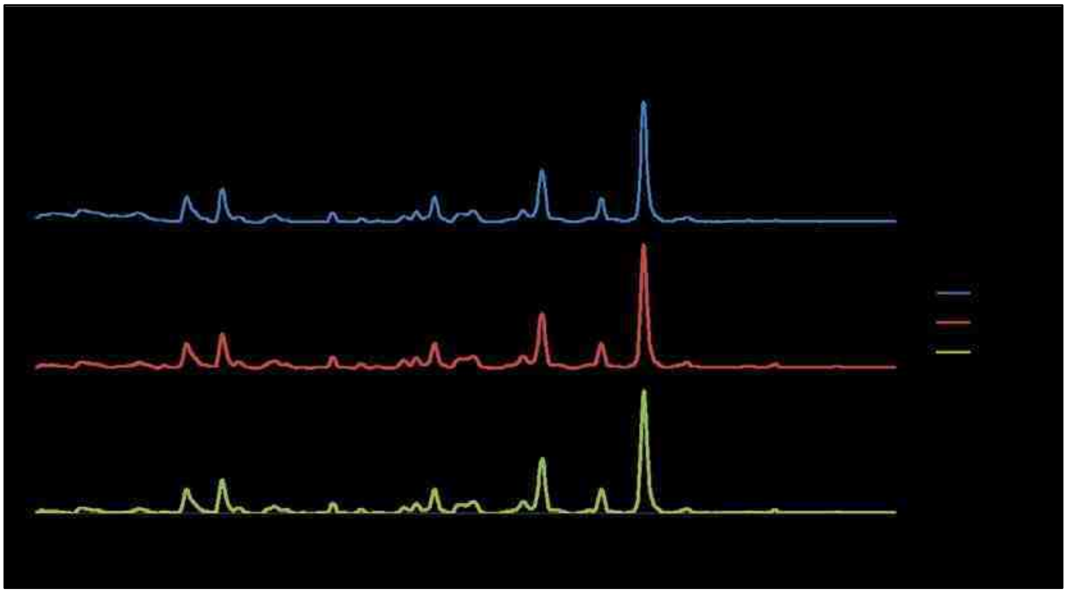


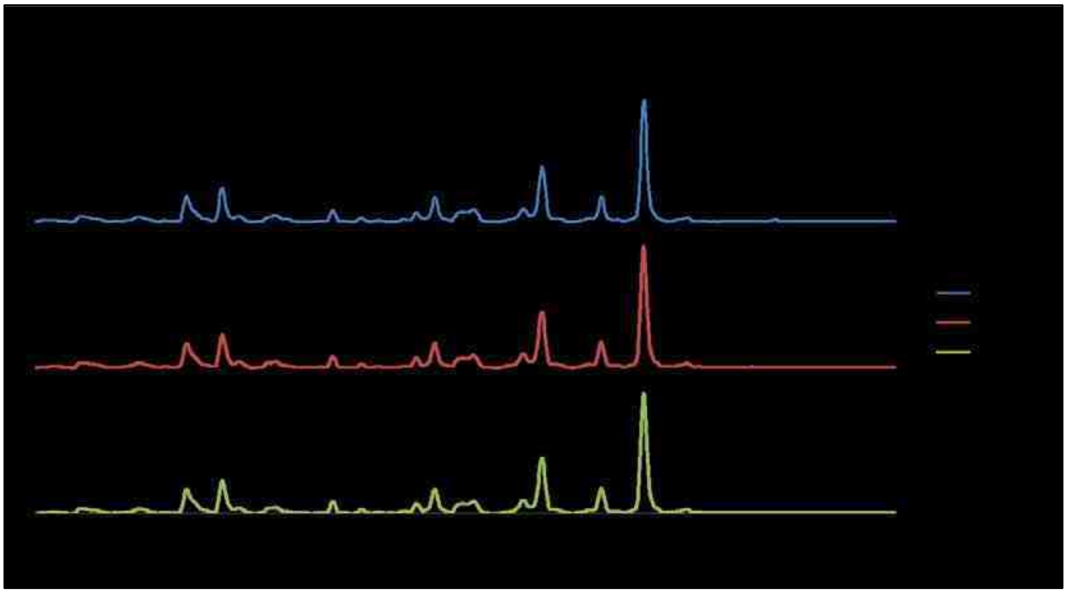


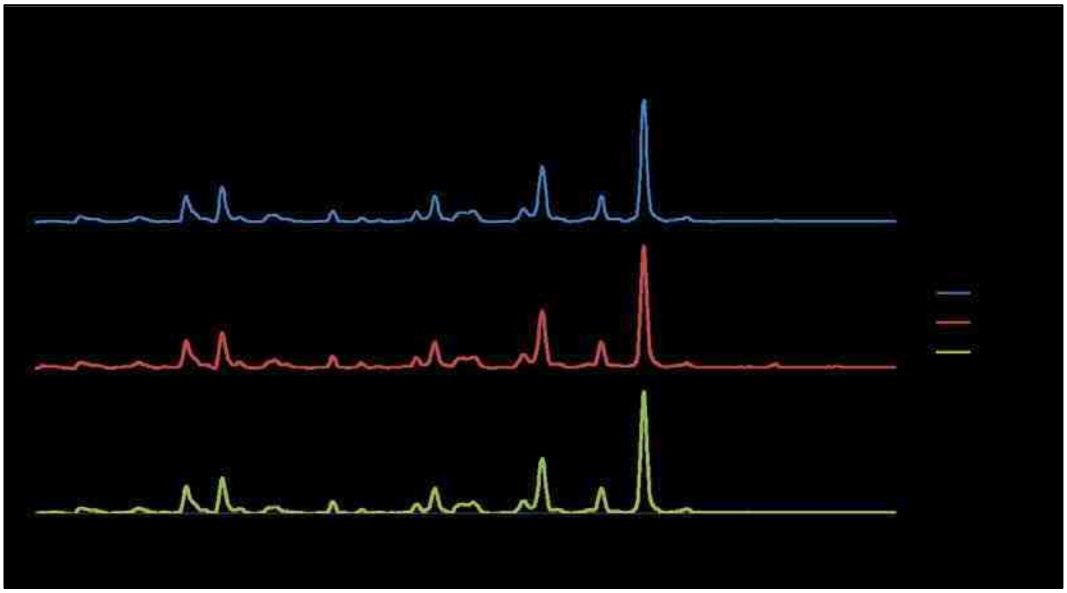


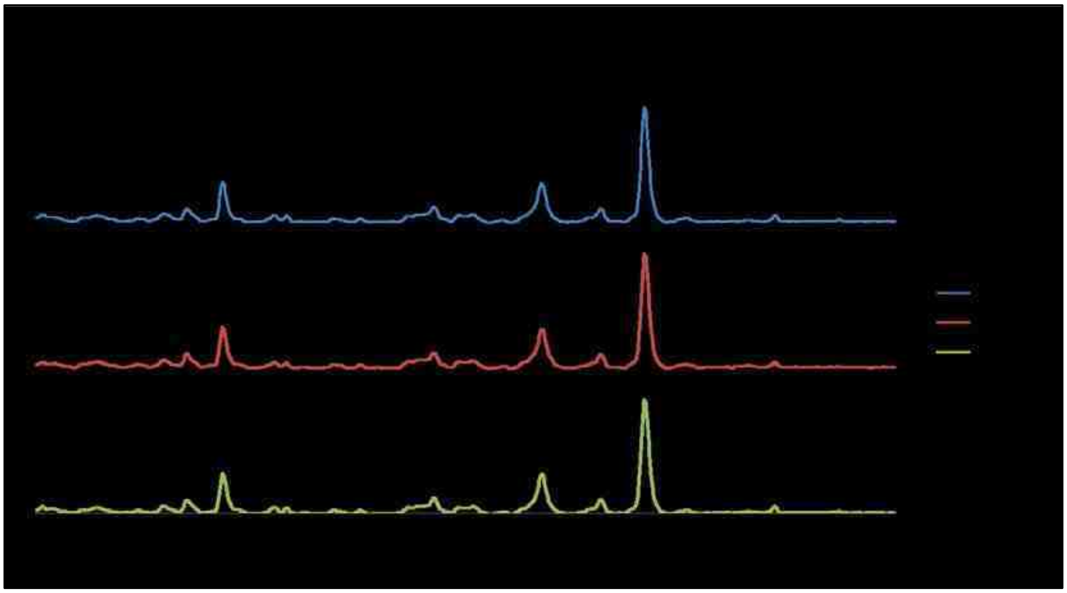




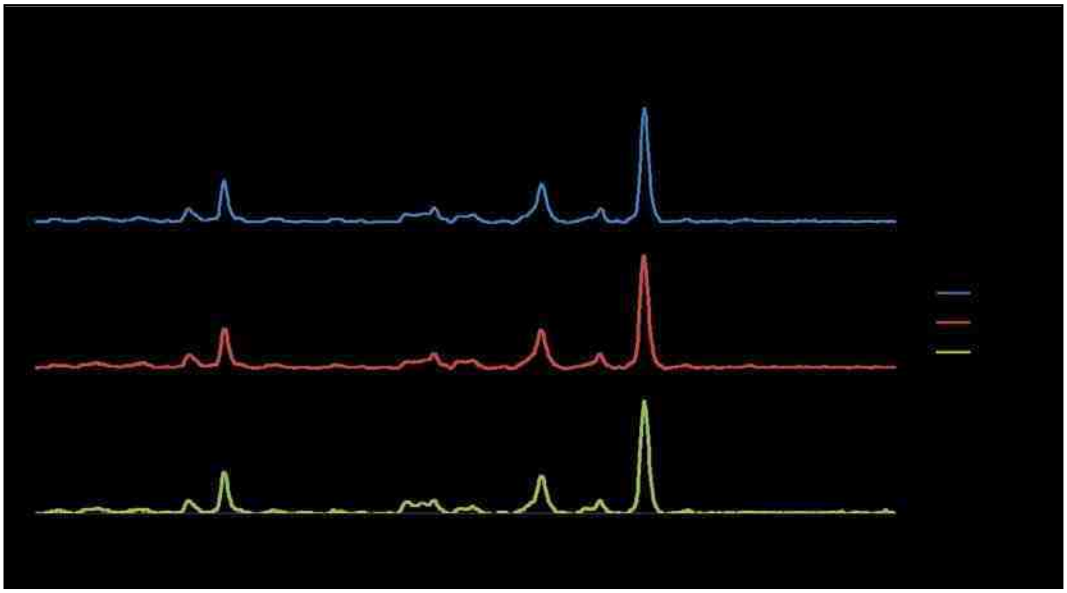


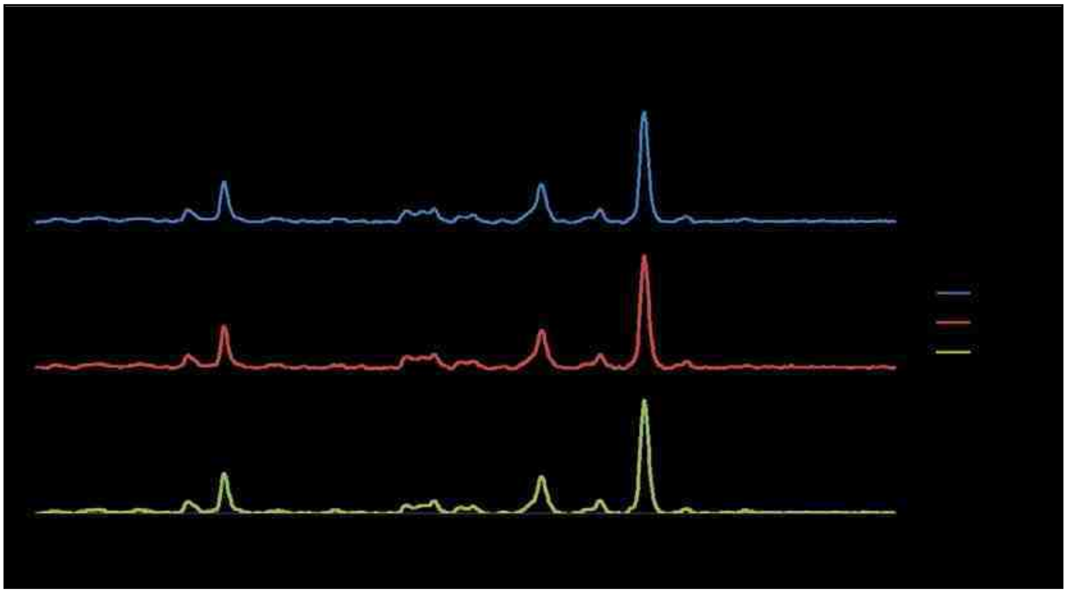


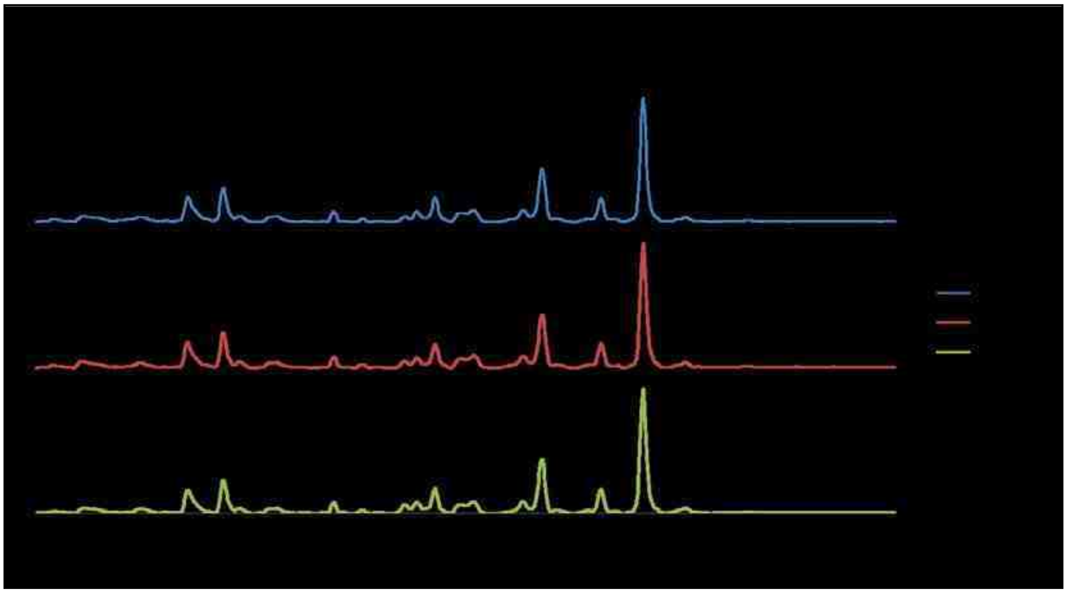


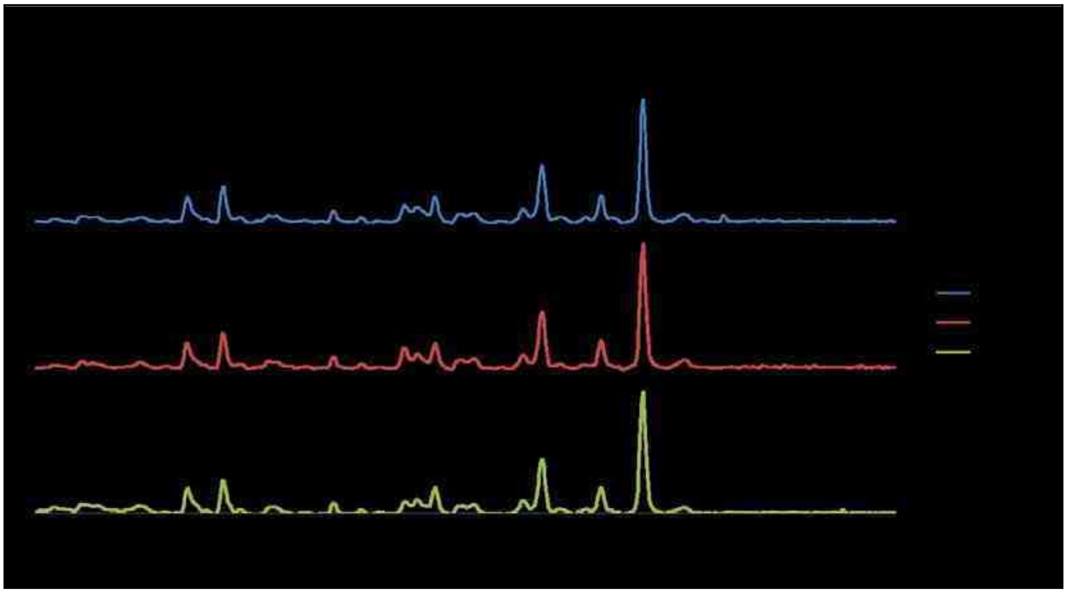


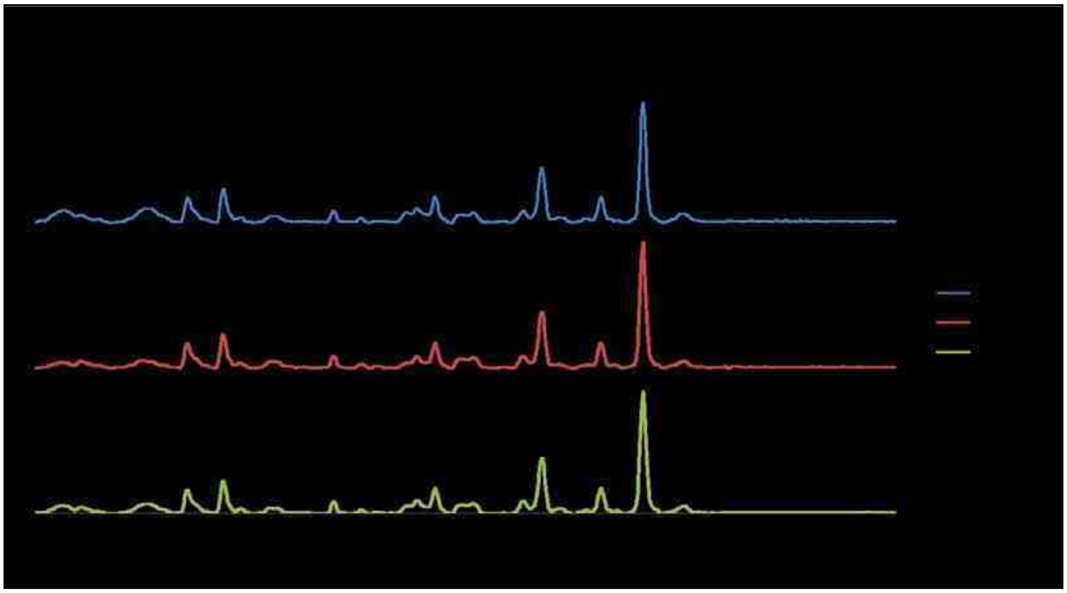
A.2 External Validation

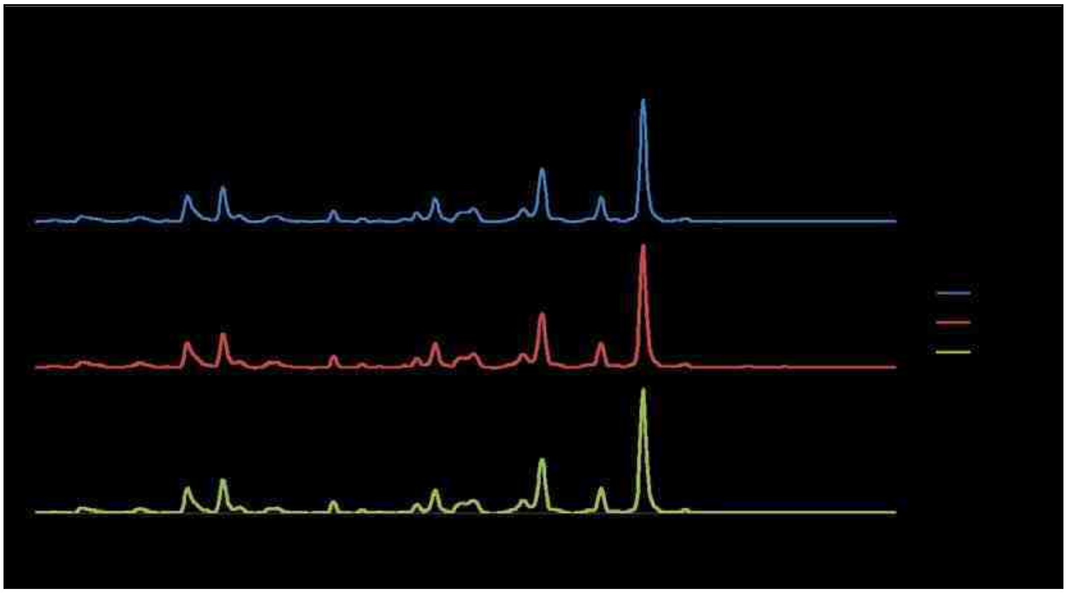


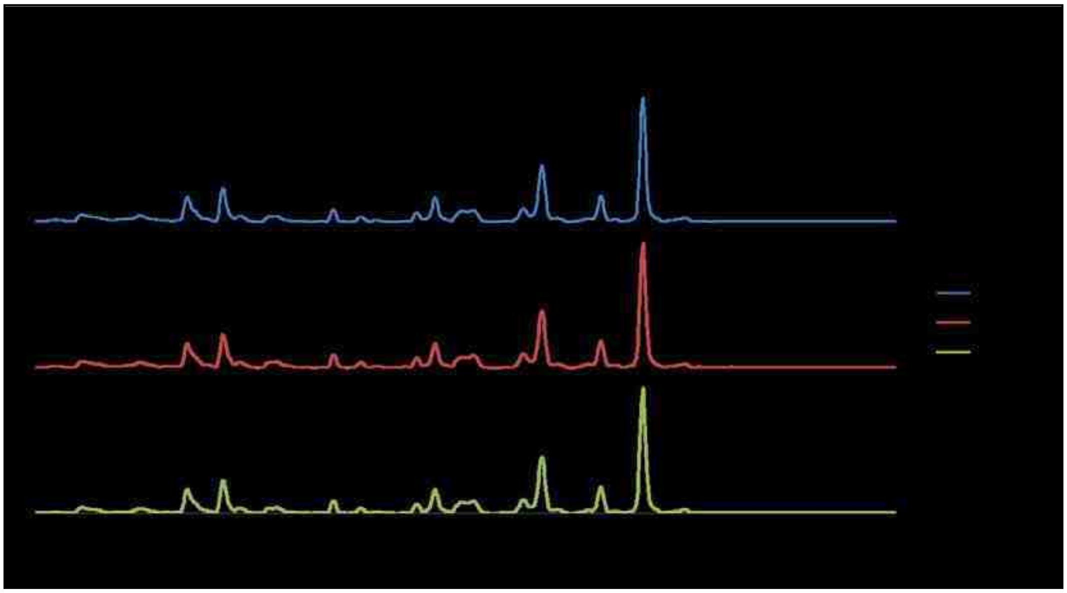


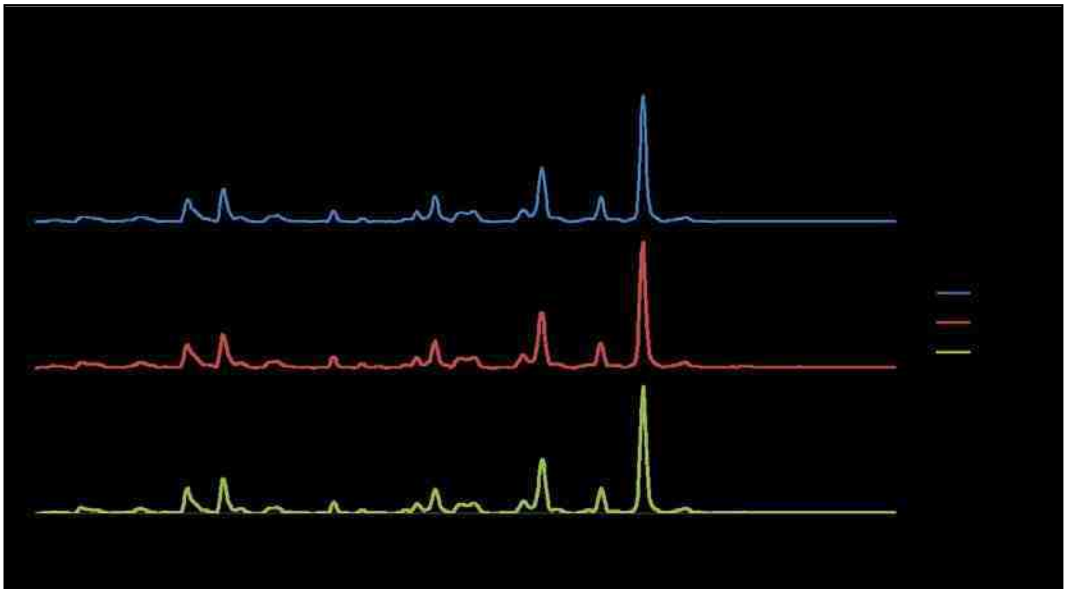


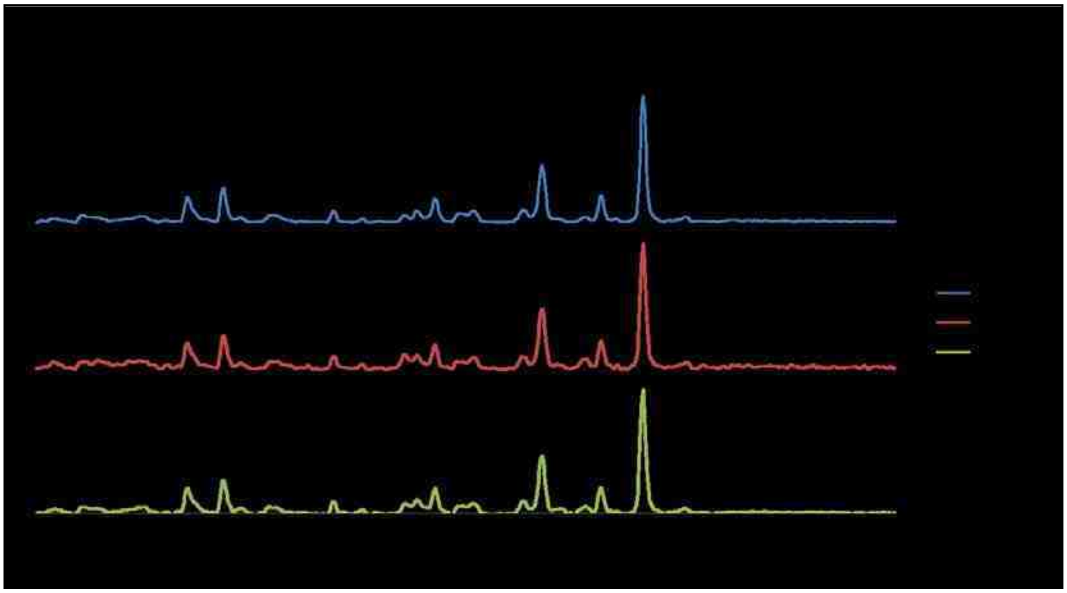


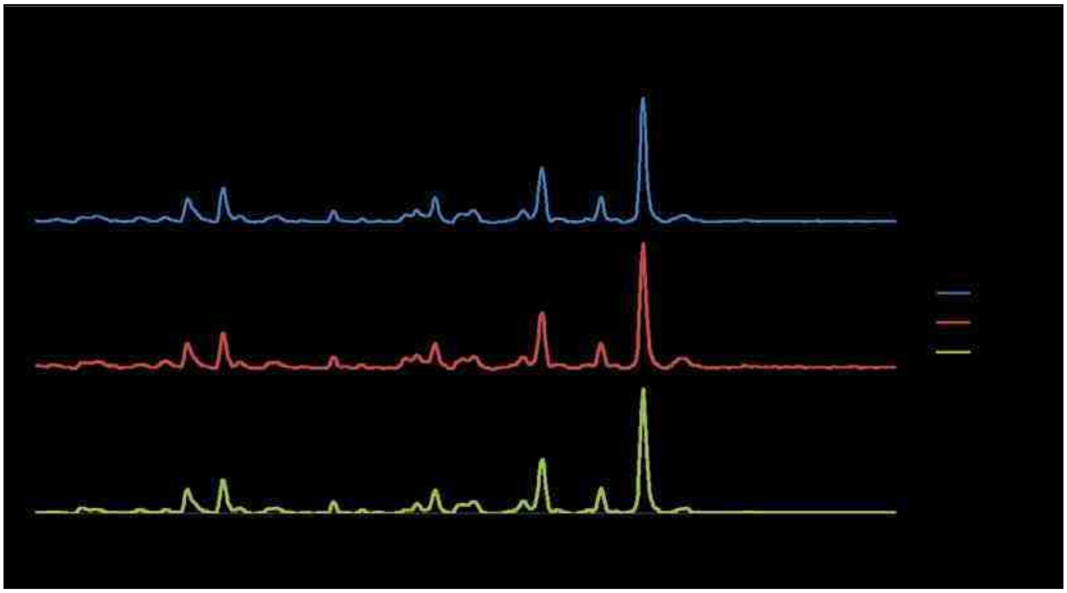


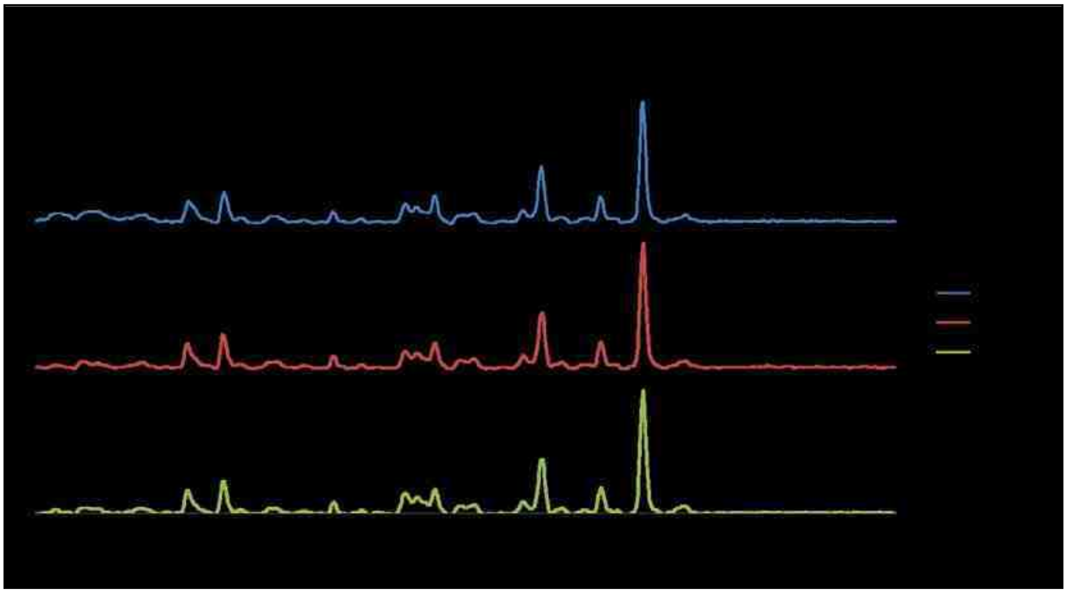


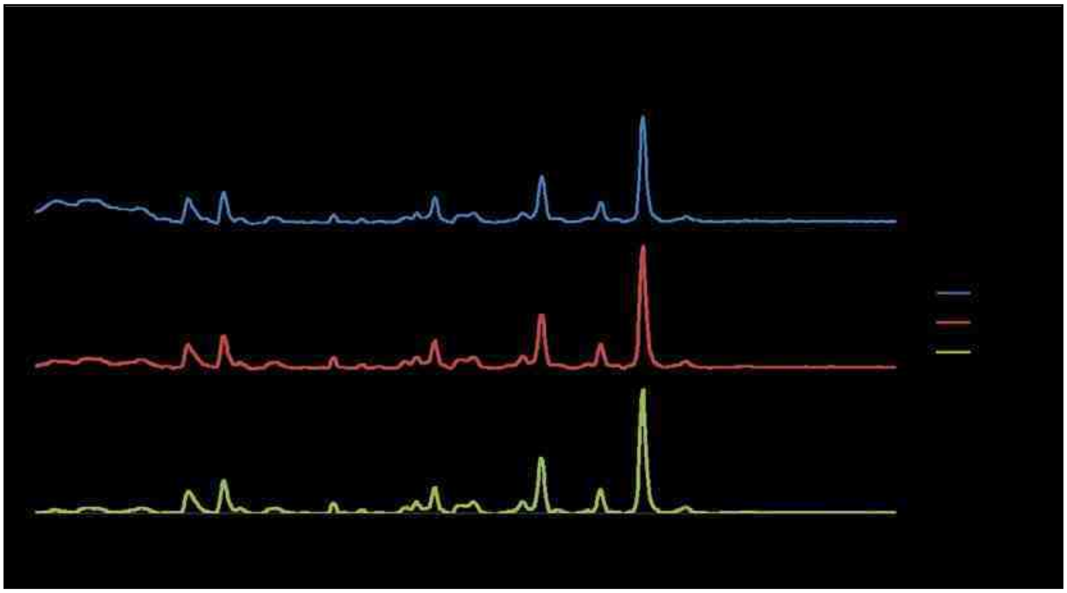


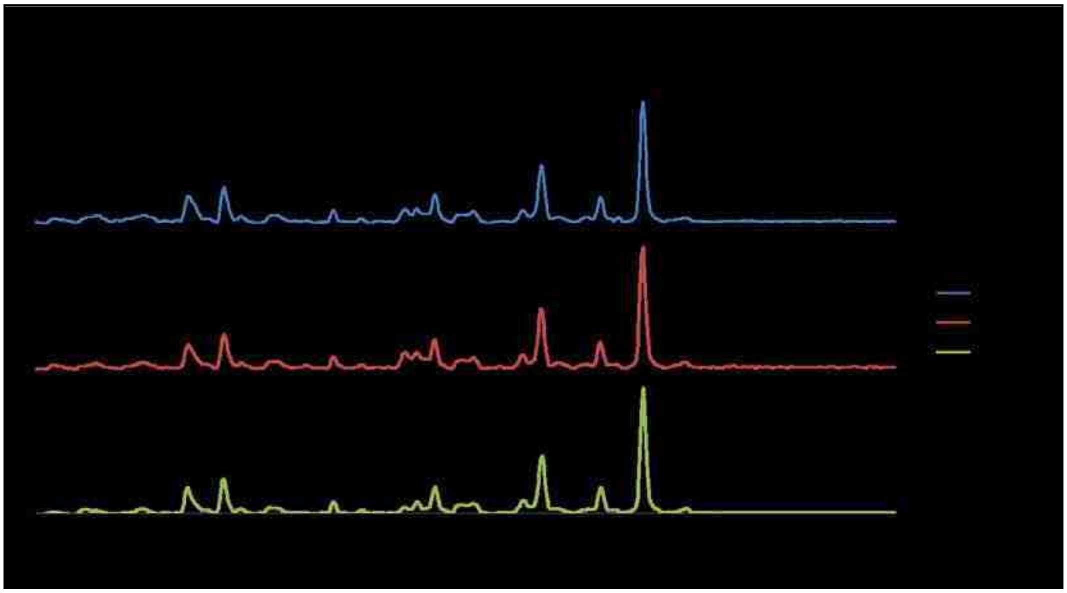


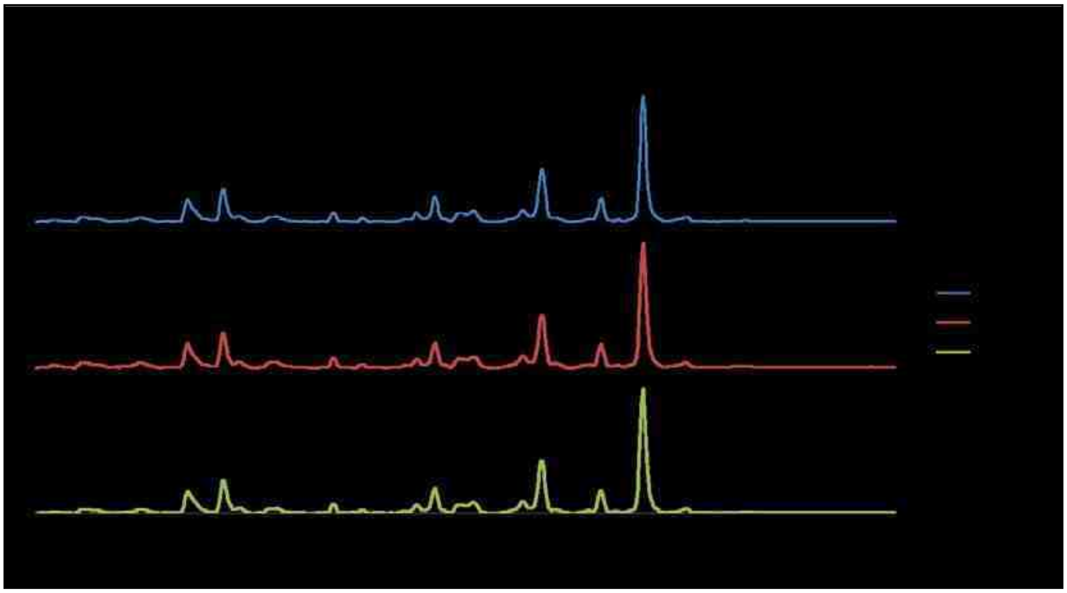


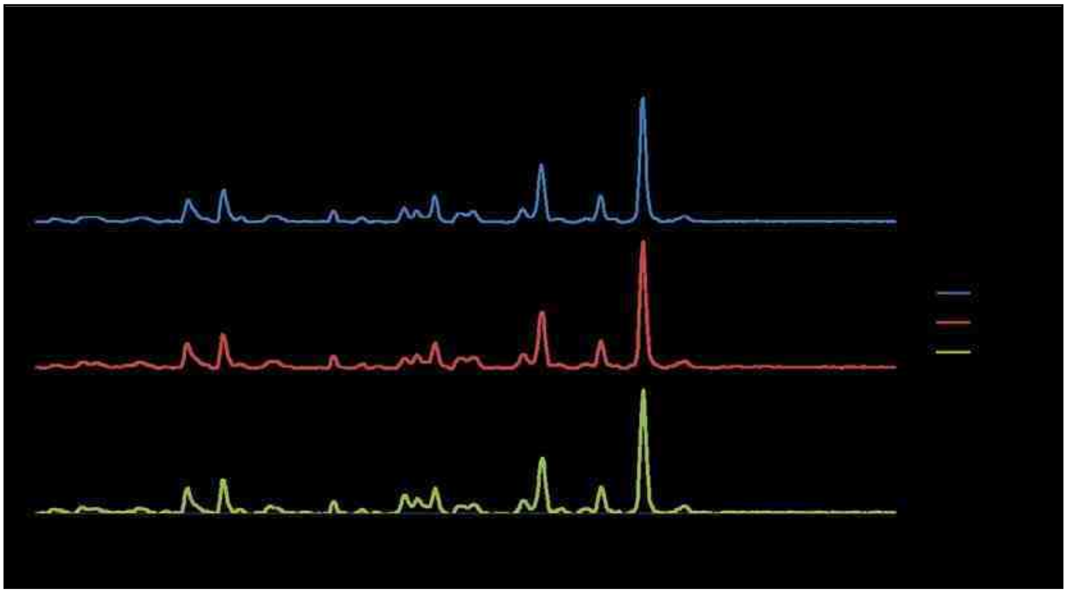


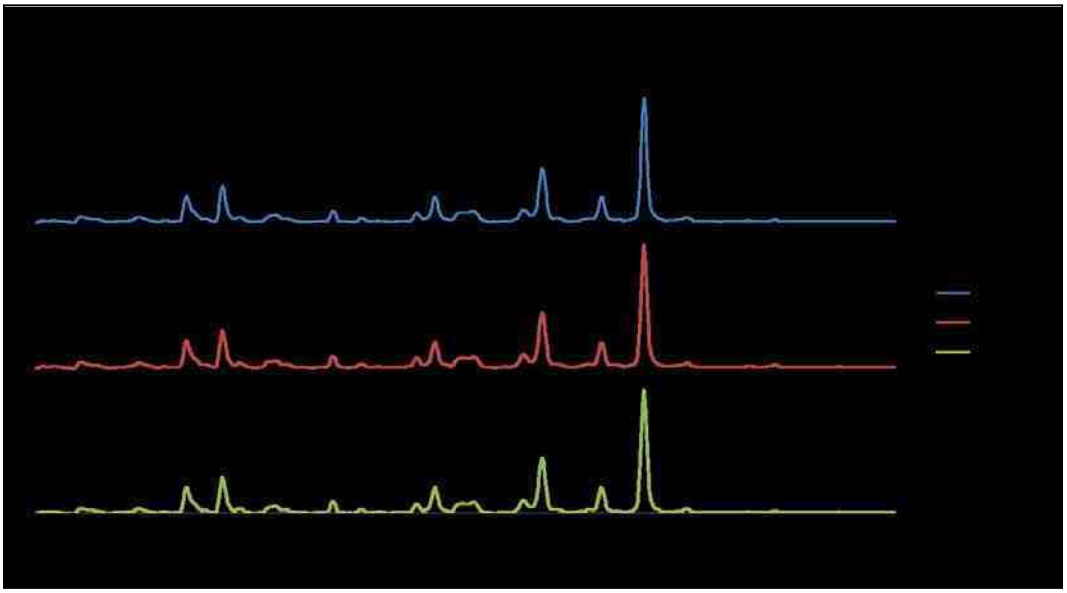


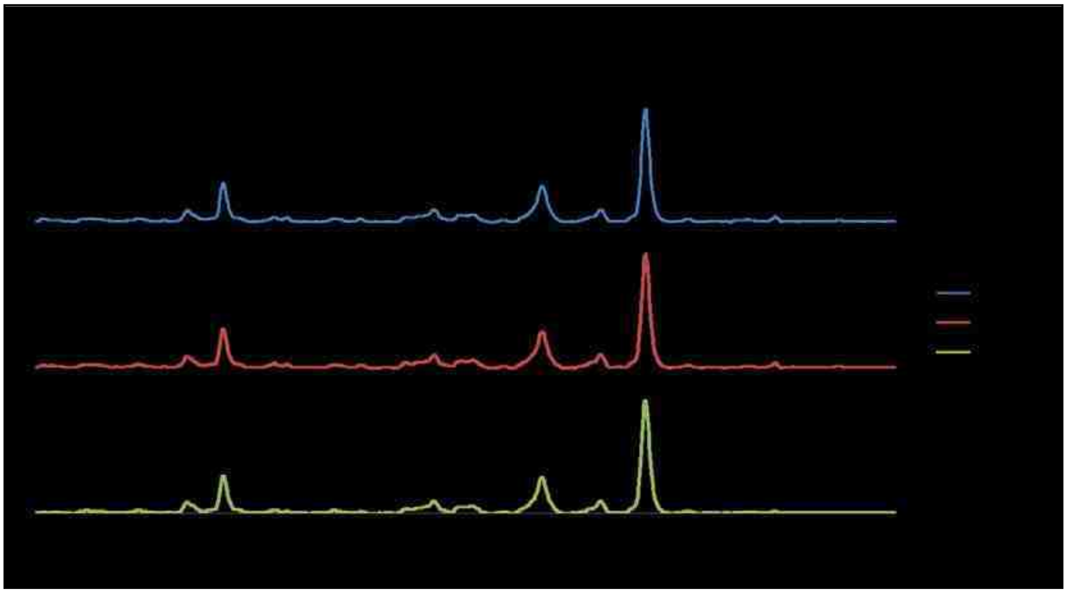


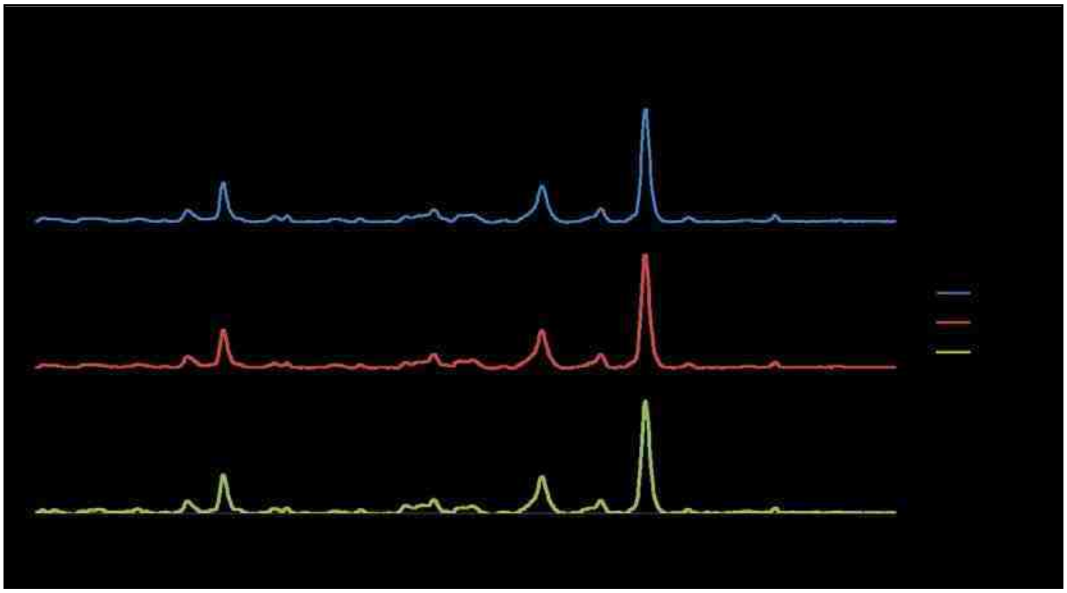


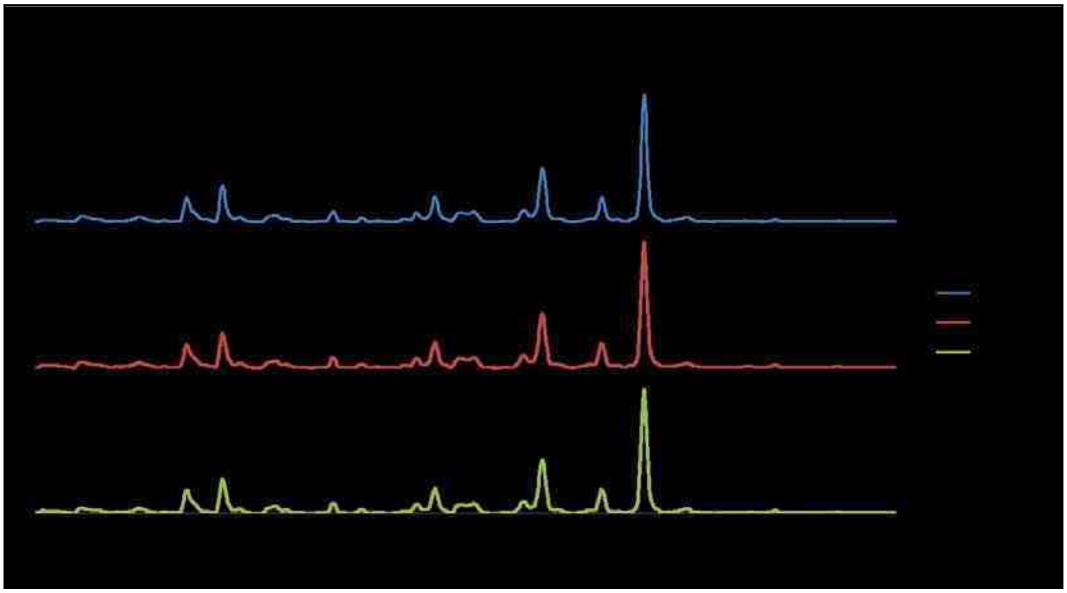


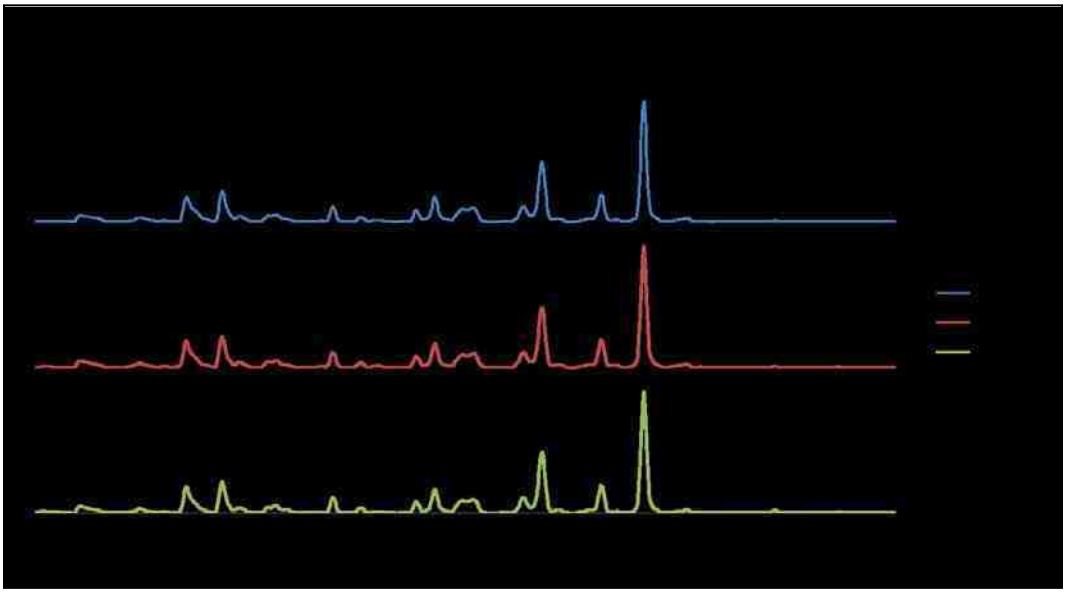


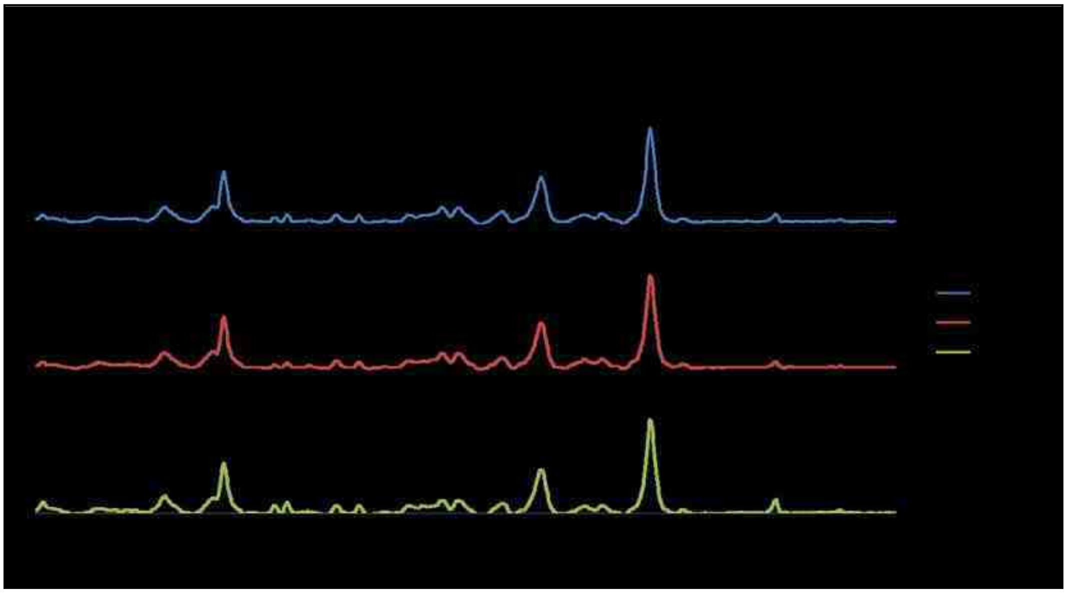


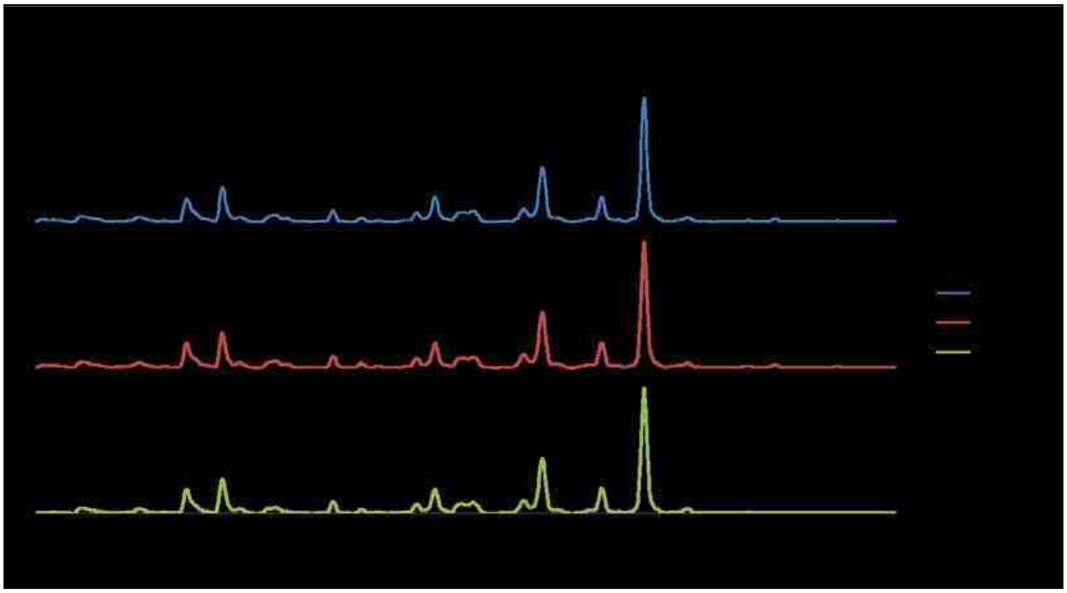


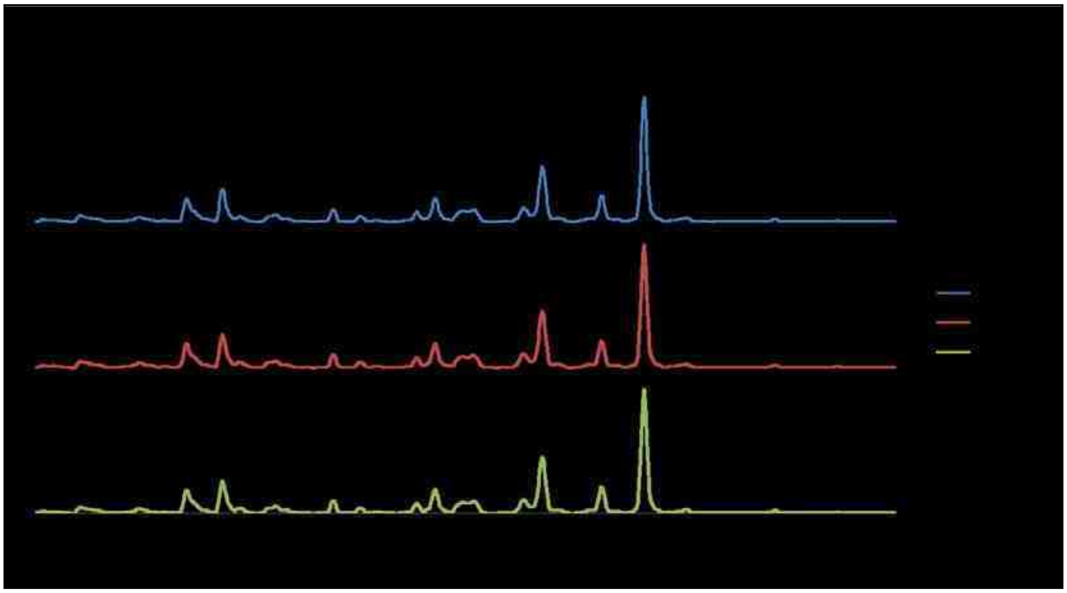


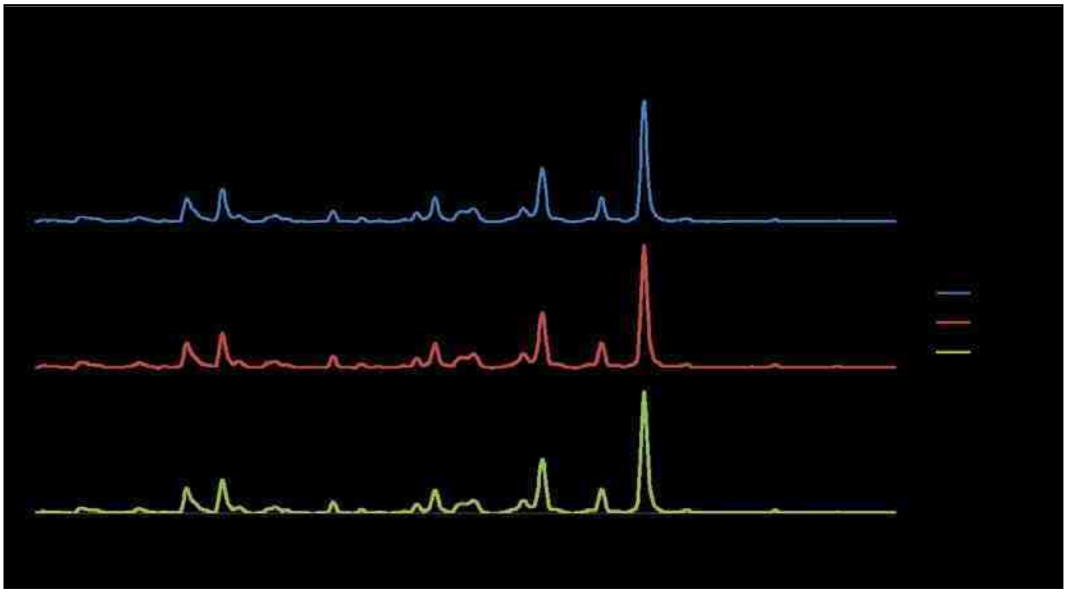


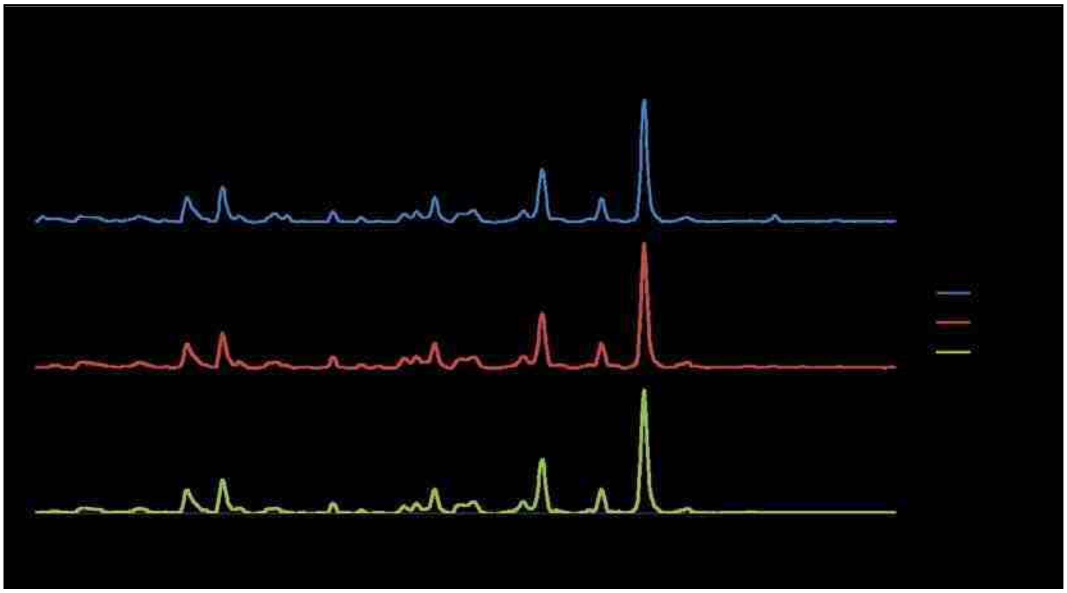










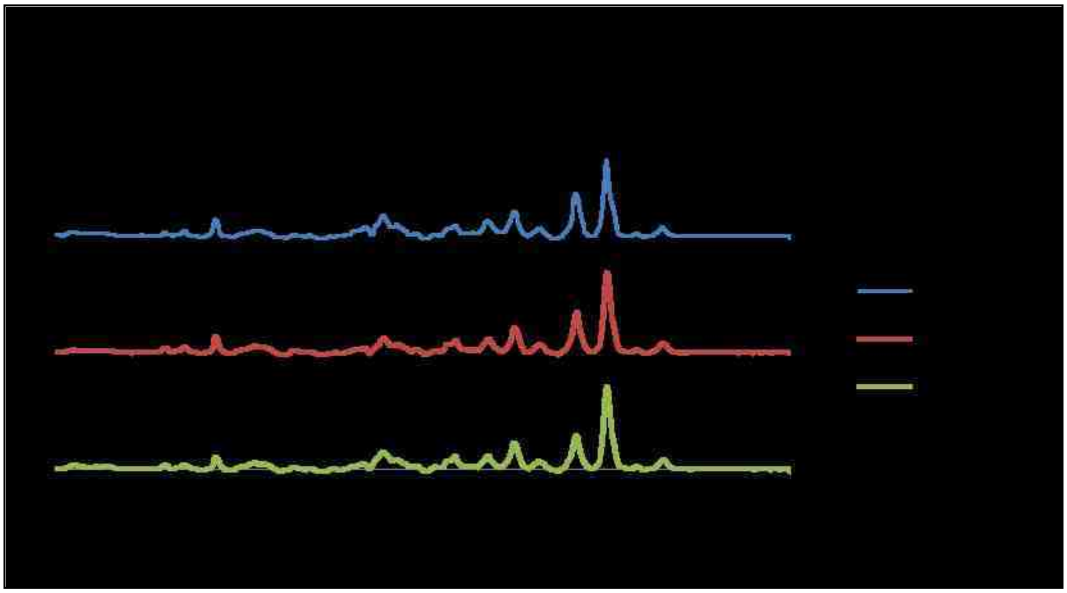


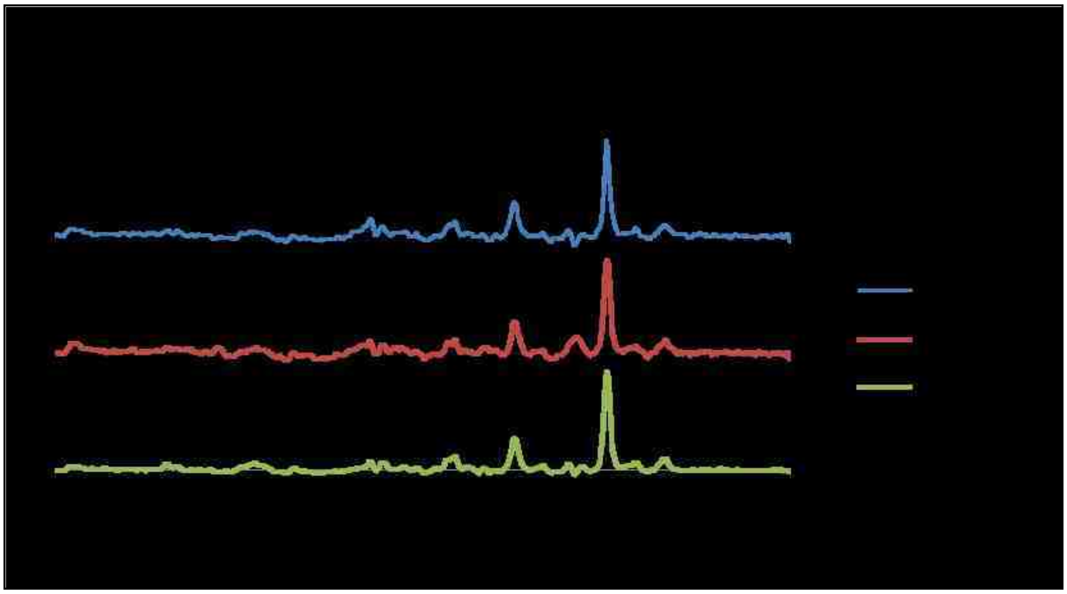
A.3 External Validation Key

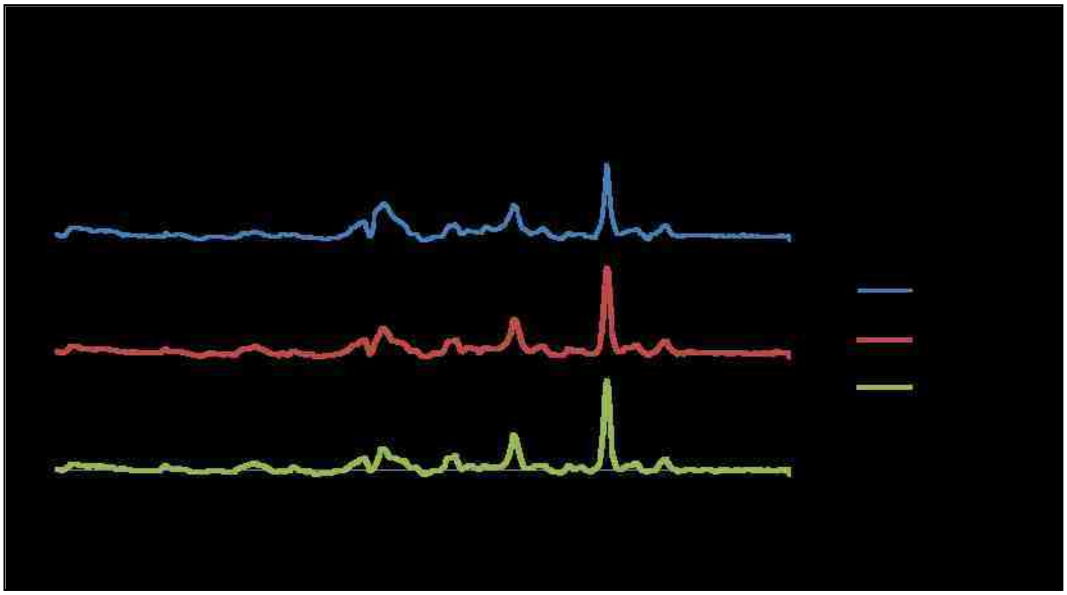
External Validation Sample	**Original Sample**
EVB01	2-Canon
EVB02	13-Lexmark
EVB03	15-Mita
EVB04	16-Ricoh
EVB05	17-Ricoh
EVB06	21-QMS
EVB07	22-Panasonic
EVB08	23-Xerox
EVB09	28-Ricoh
EVB10	31-Toshiba
EVB11	36-Ricoh
EVB12	42-Lexmark
EVB13	43-Konica
EVB14	46-Okidata
EVB15	48-Ricoh
EVB16	53-Canon CLC 800
EVB17	58-Canon CLC 950
EVB18	63-IBM Network Color Printer
EVB19	66-Okidata C711
EVB20	68-Hewlett Packard Color Laser Jet 5550
EVB21	85-Xerox 1005
EVB22	88-Xerox 5765/5760/5760ADF
EVB23	94-Panasonic FP C1/FP C1XE
EVB24	95-QMS MagiColor LX
EVB25	96-Mita Ci 7600

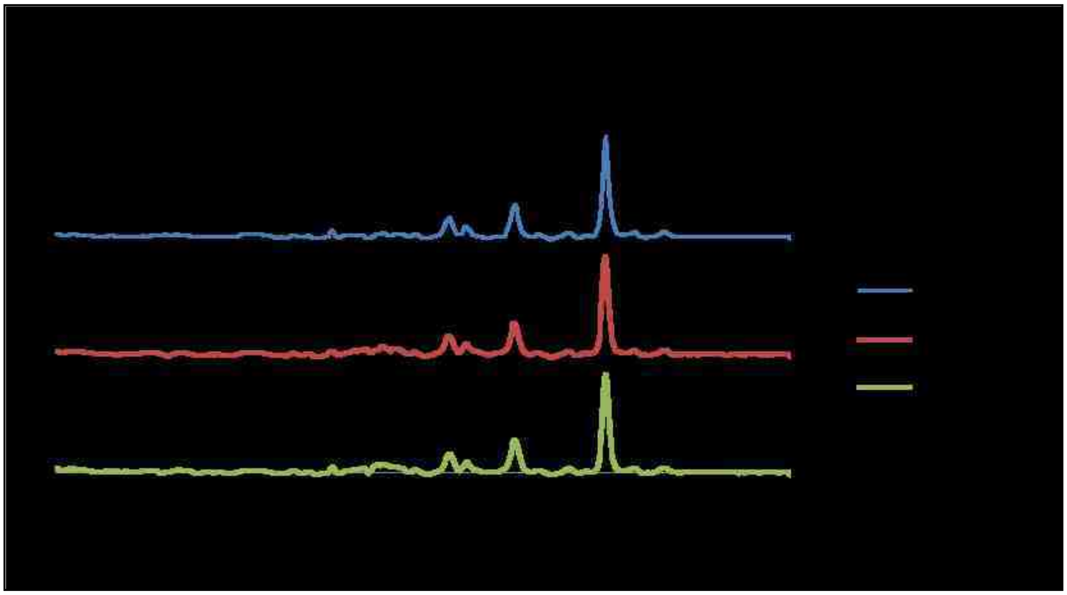
Appendix B. Yellow Toner Spectra by Raman Spectroscopy

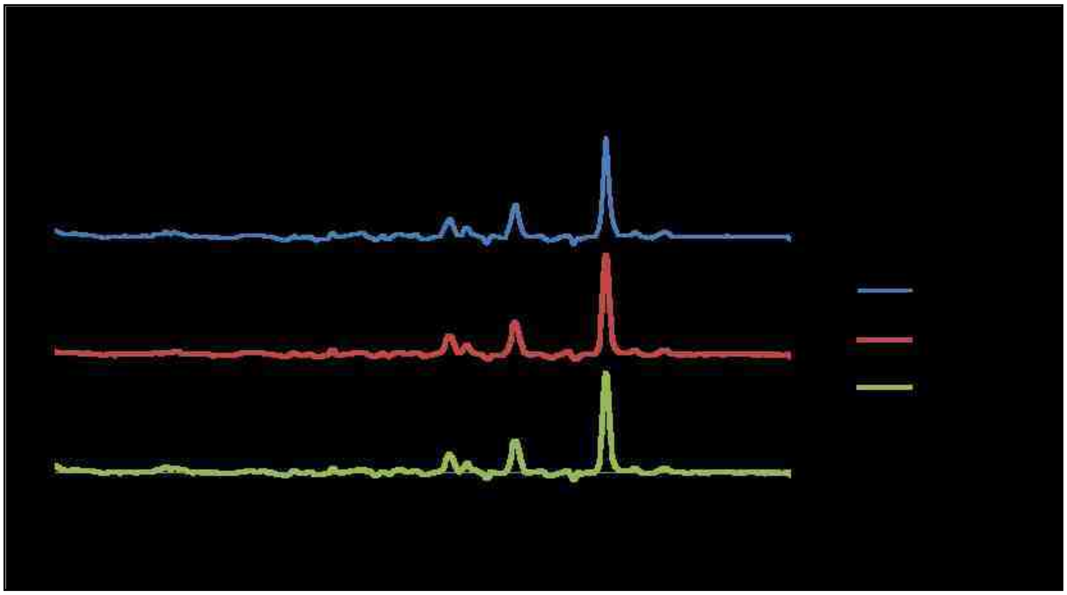
B.1 Training Samples

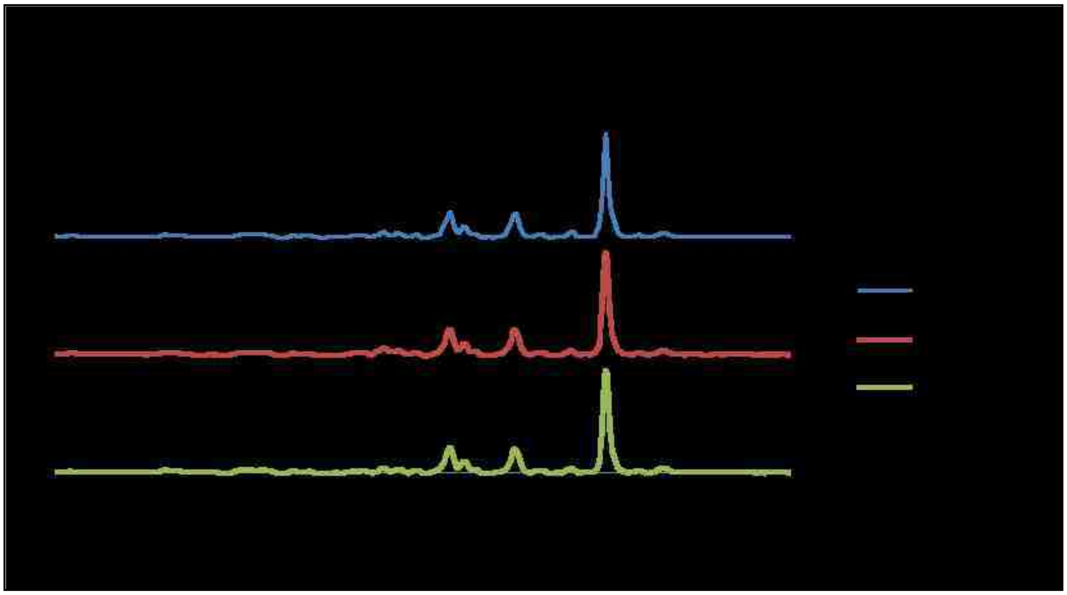


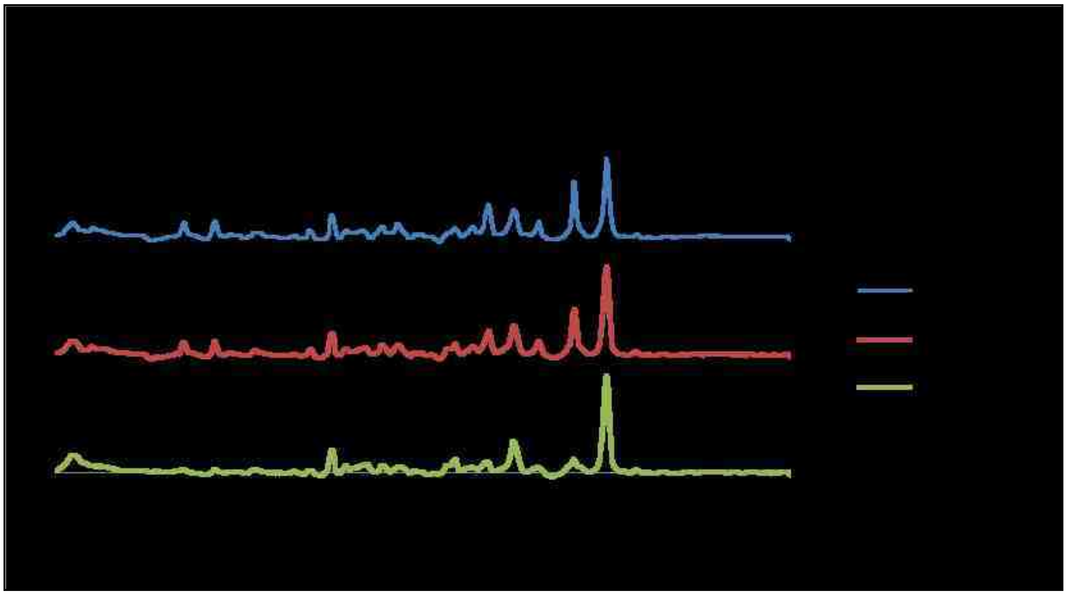


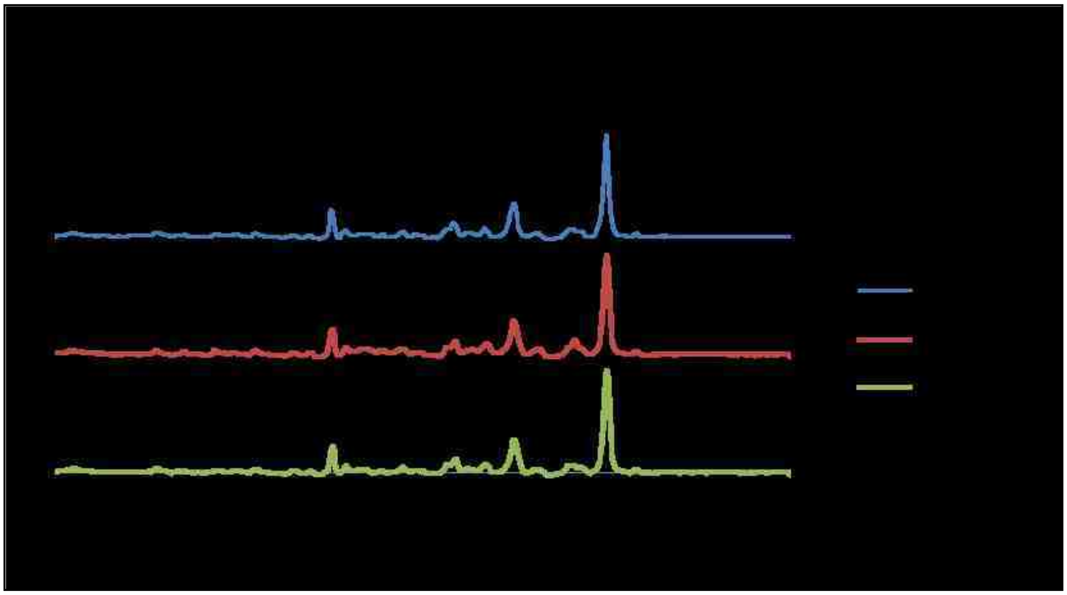


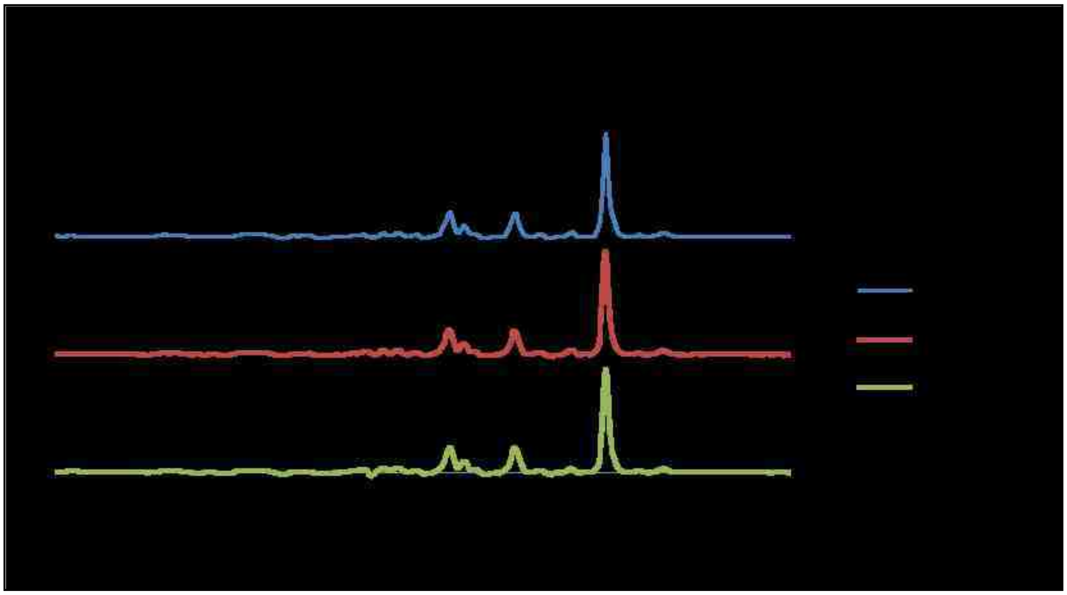


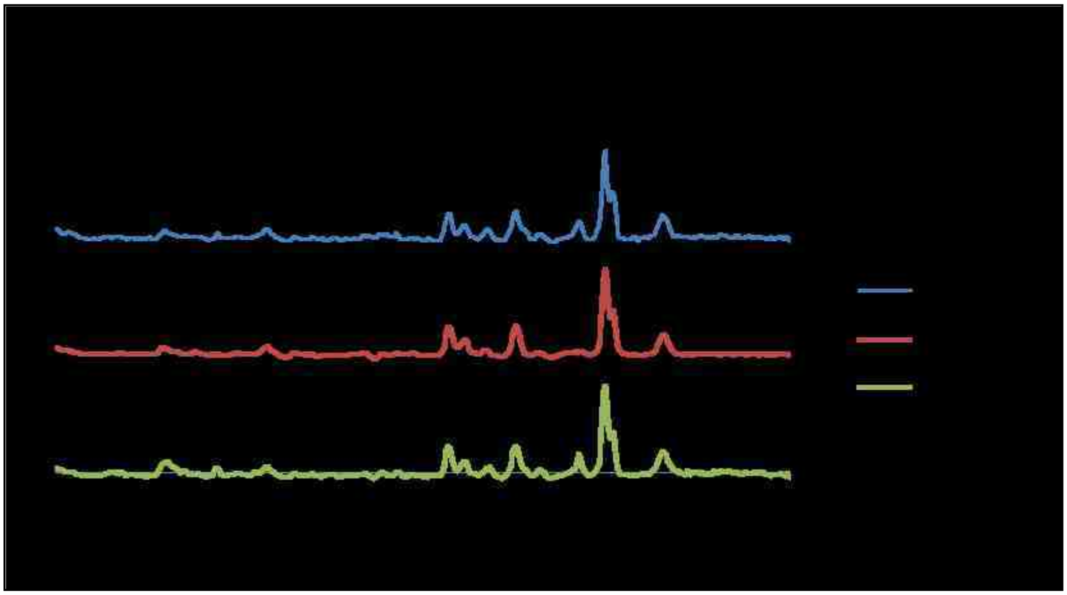


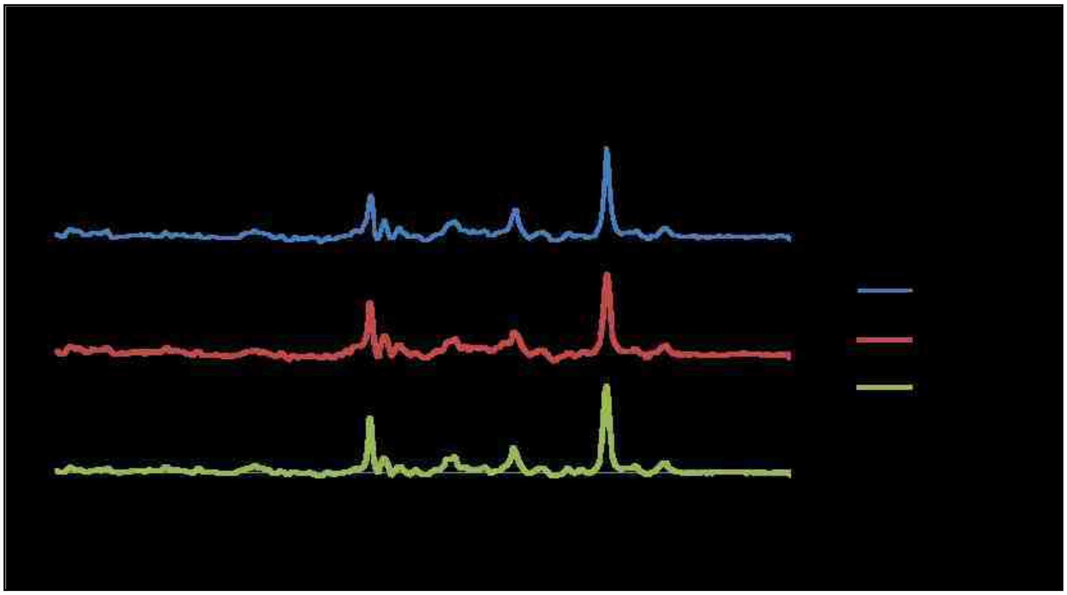


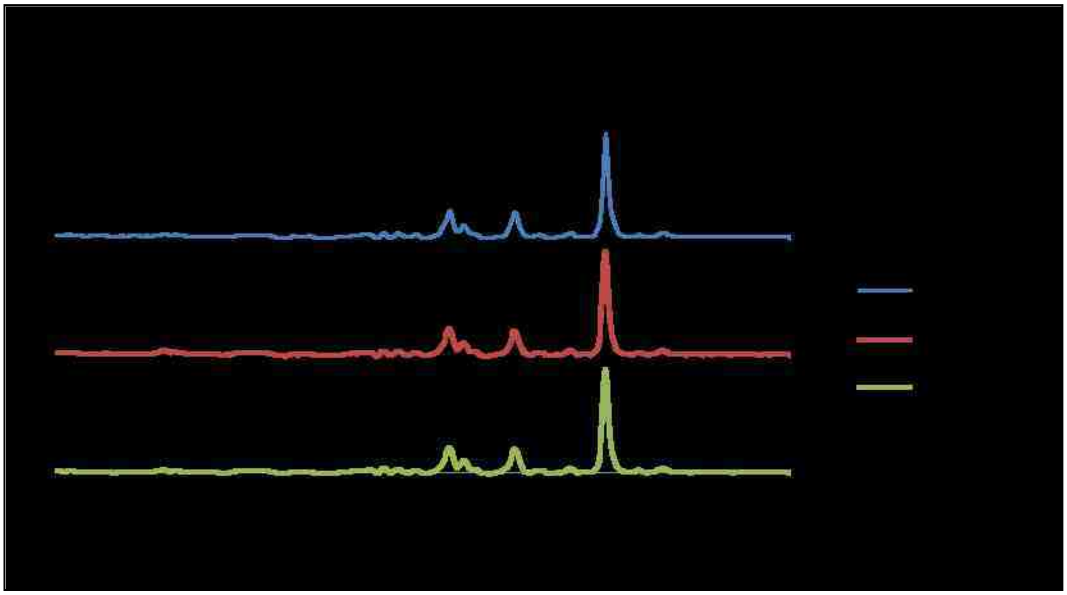


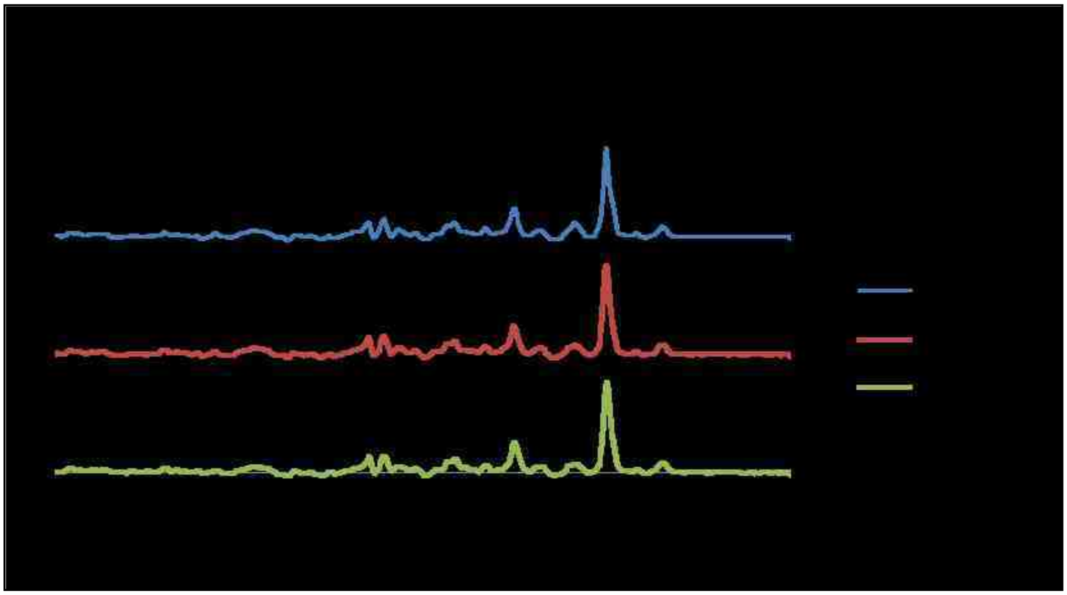


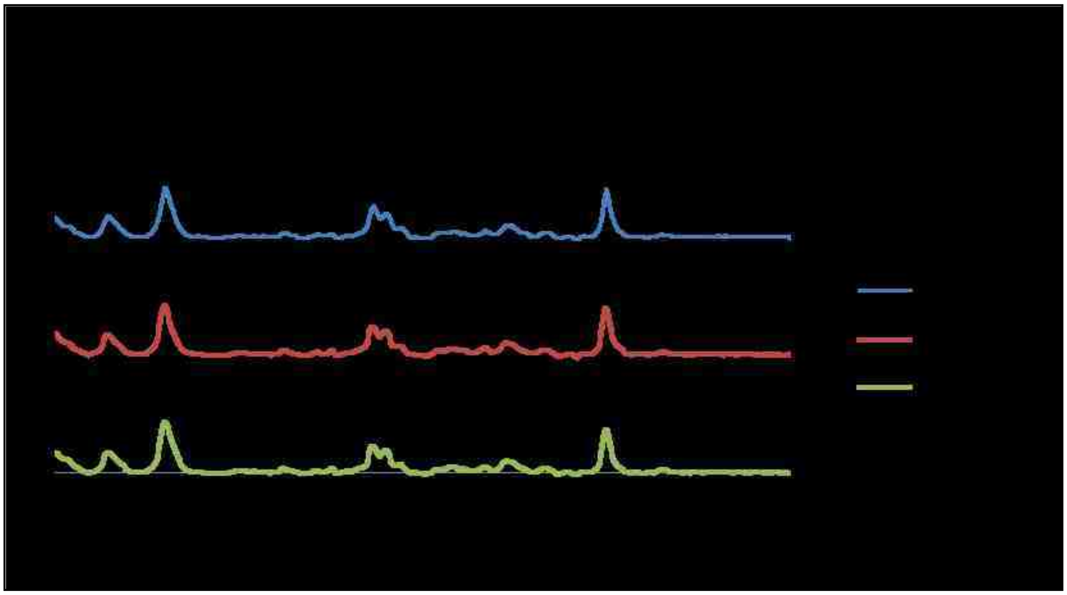


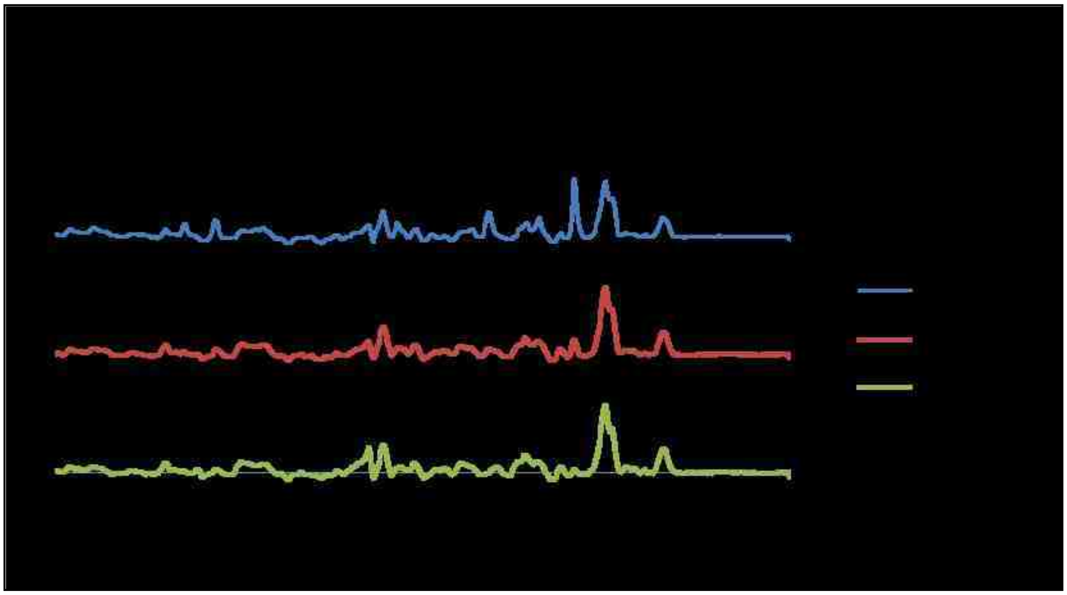


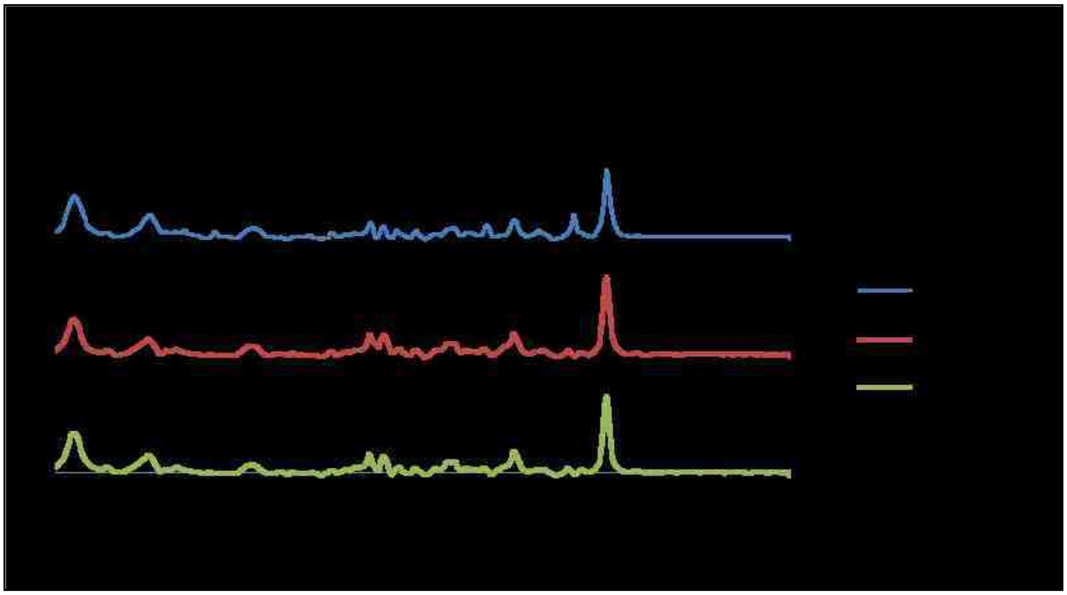


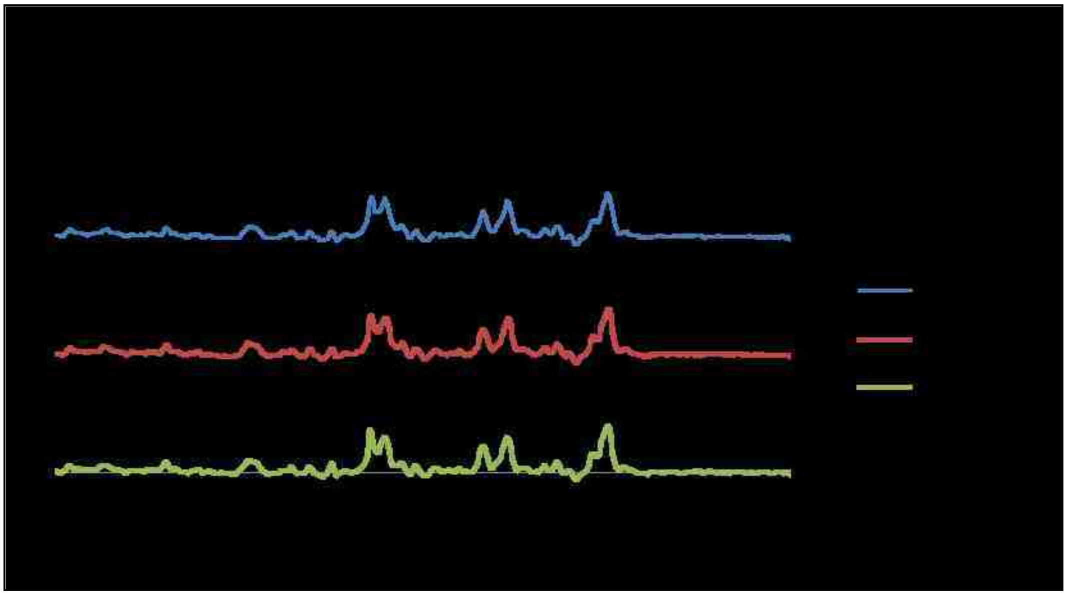


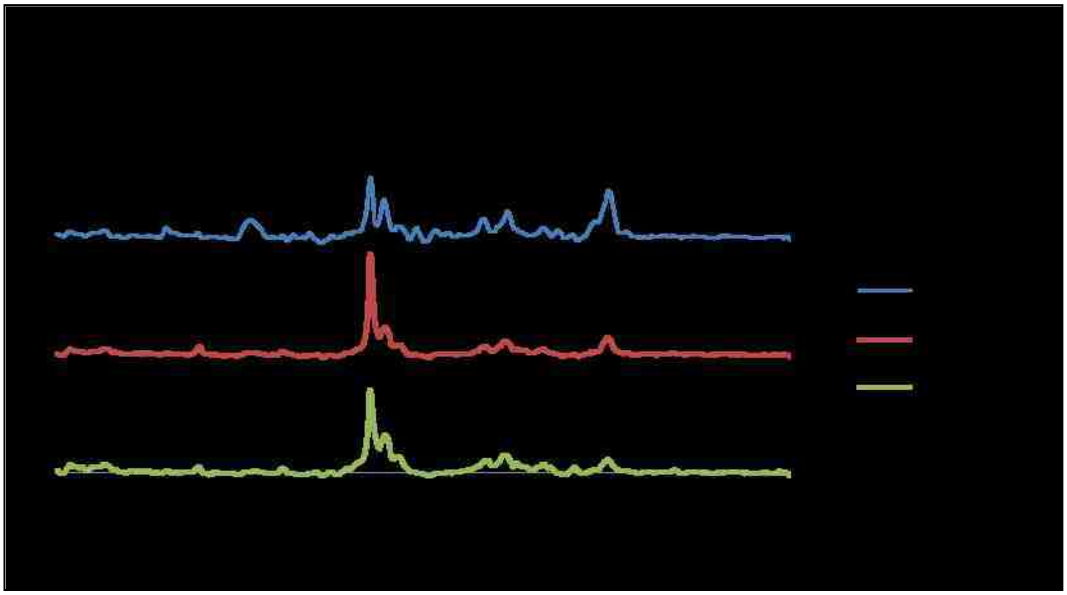


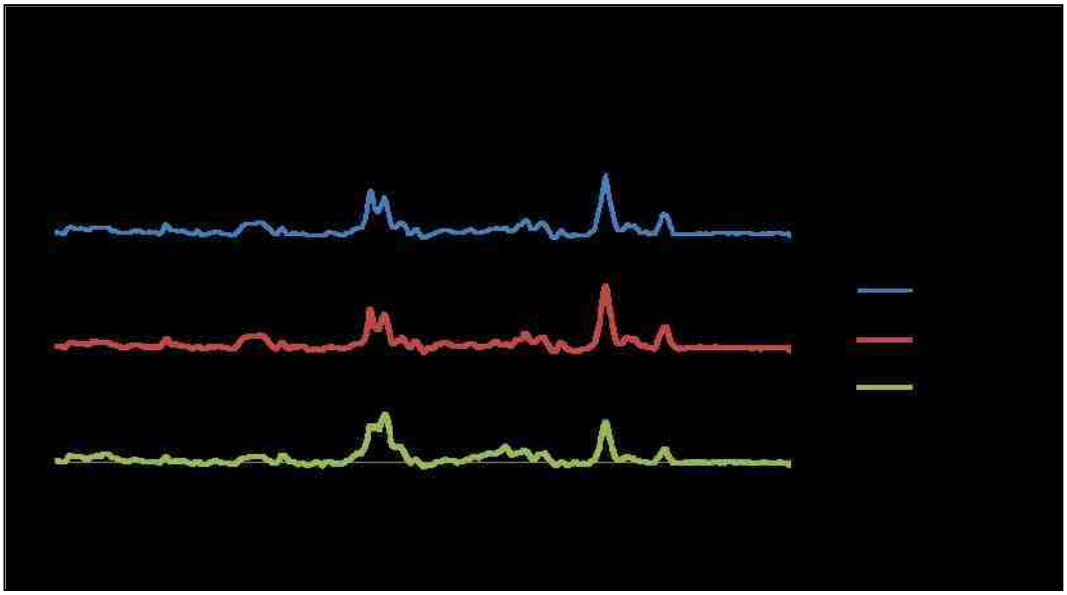


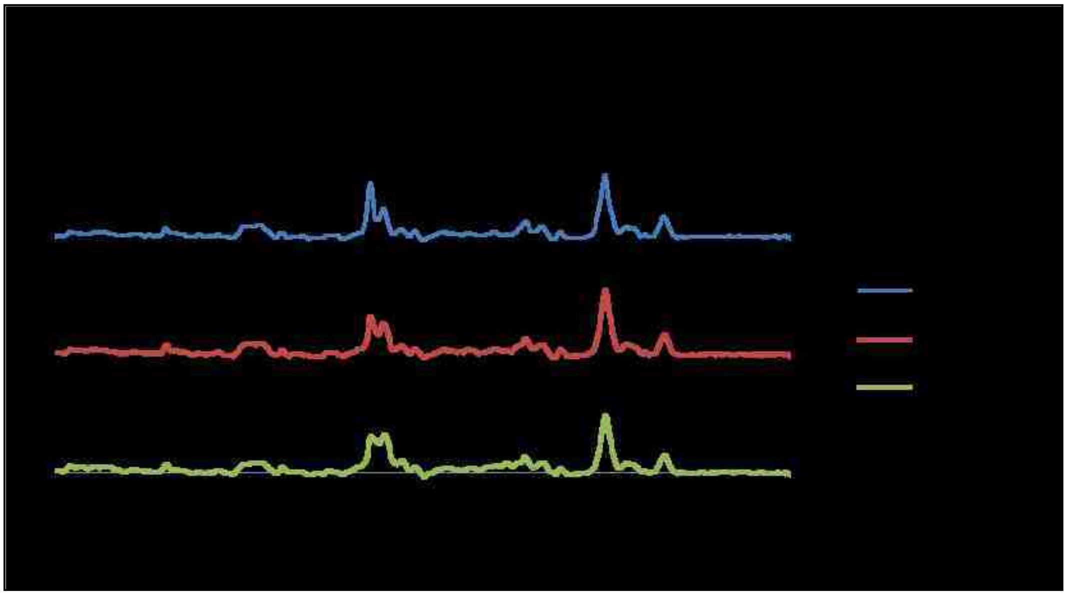


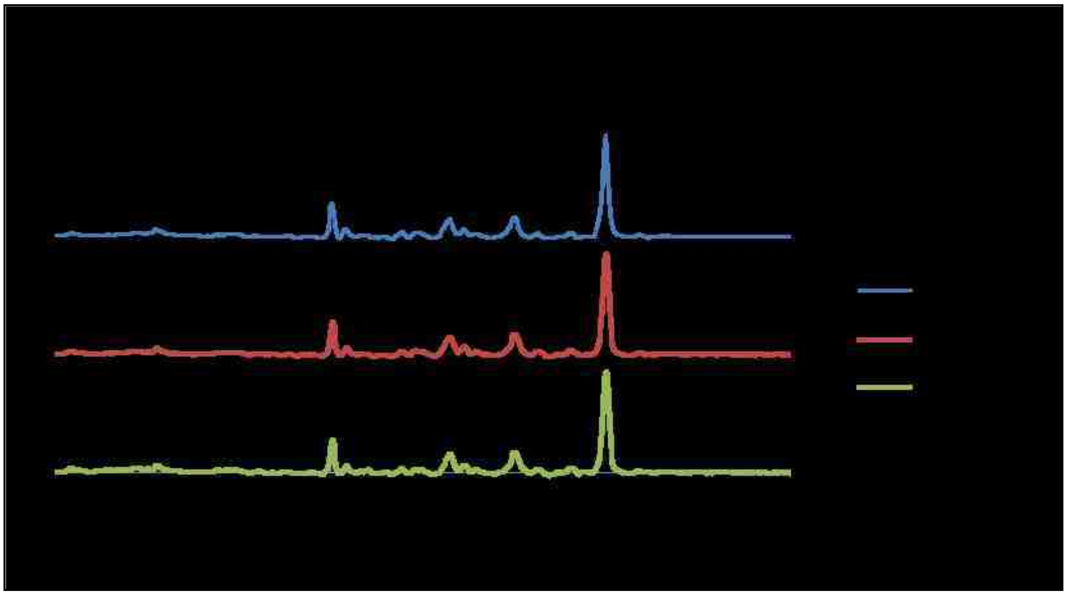


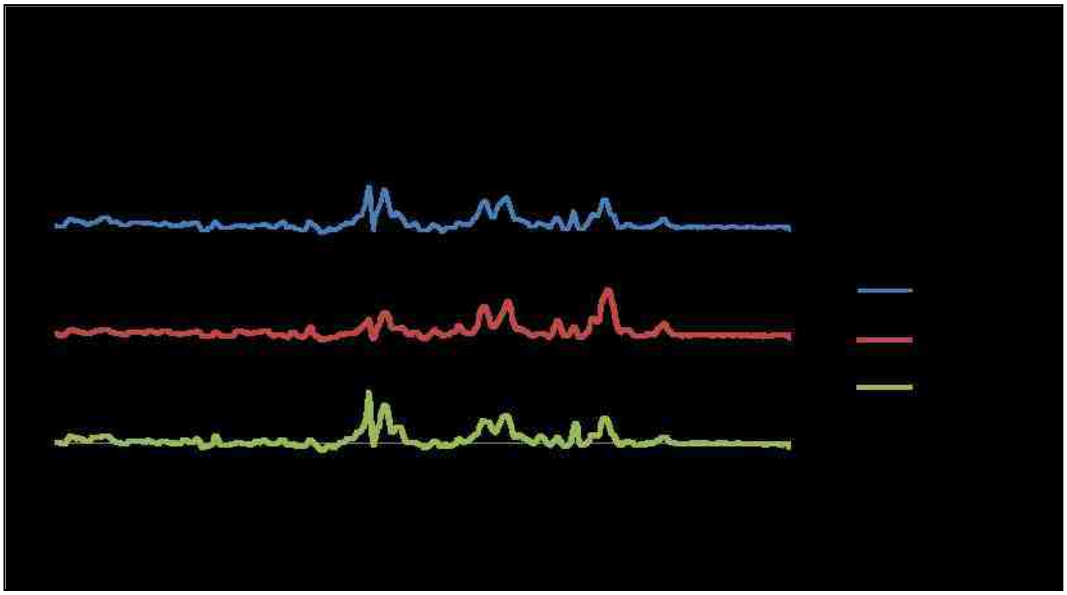


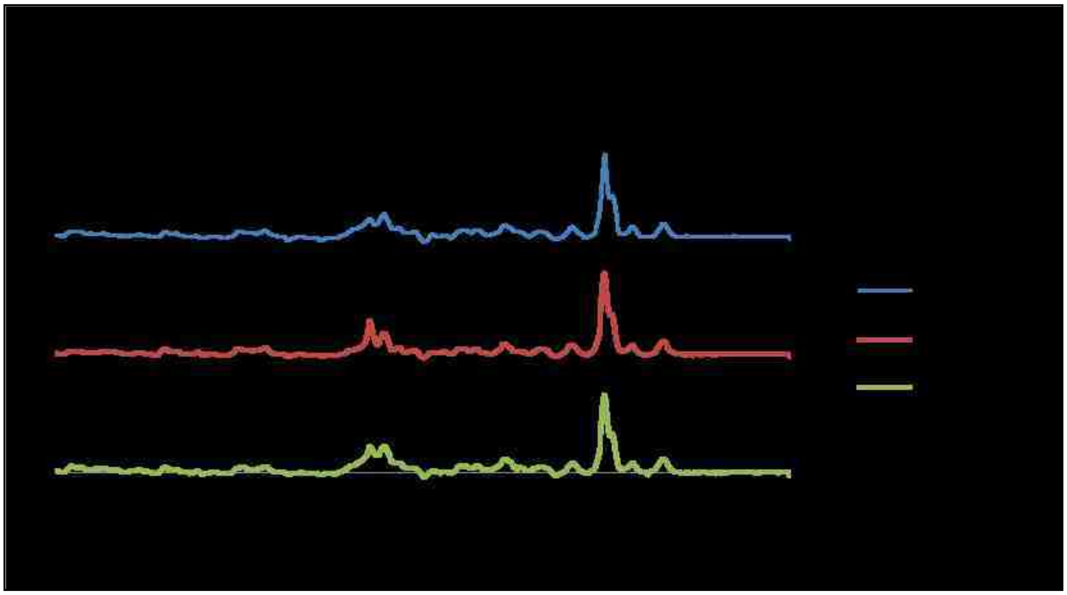


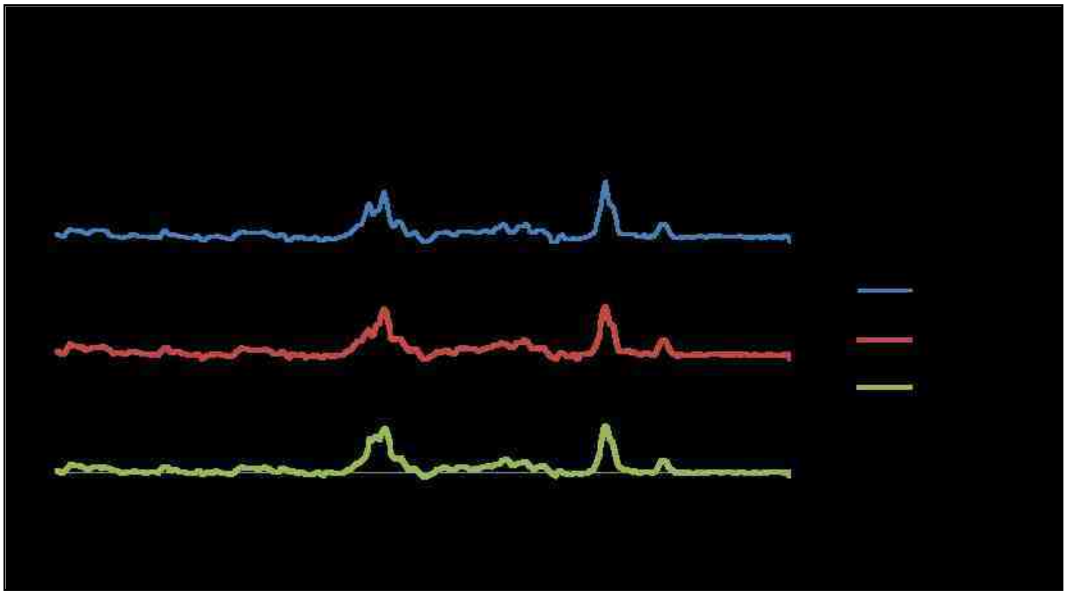


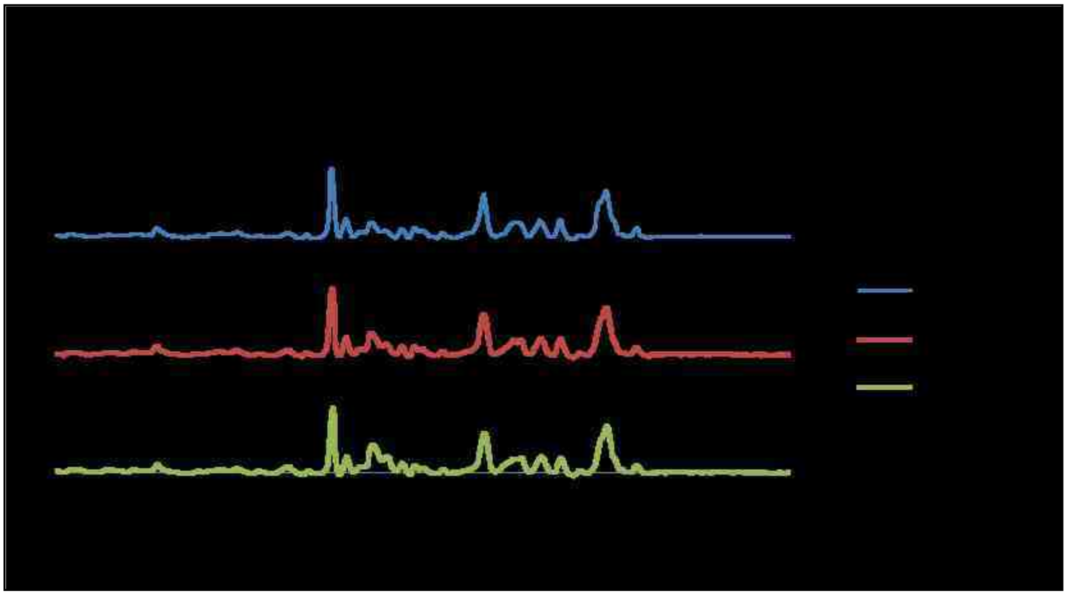


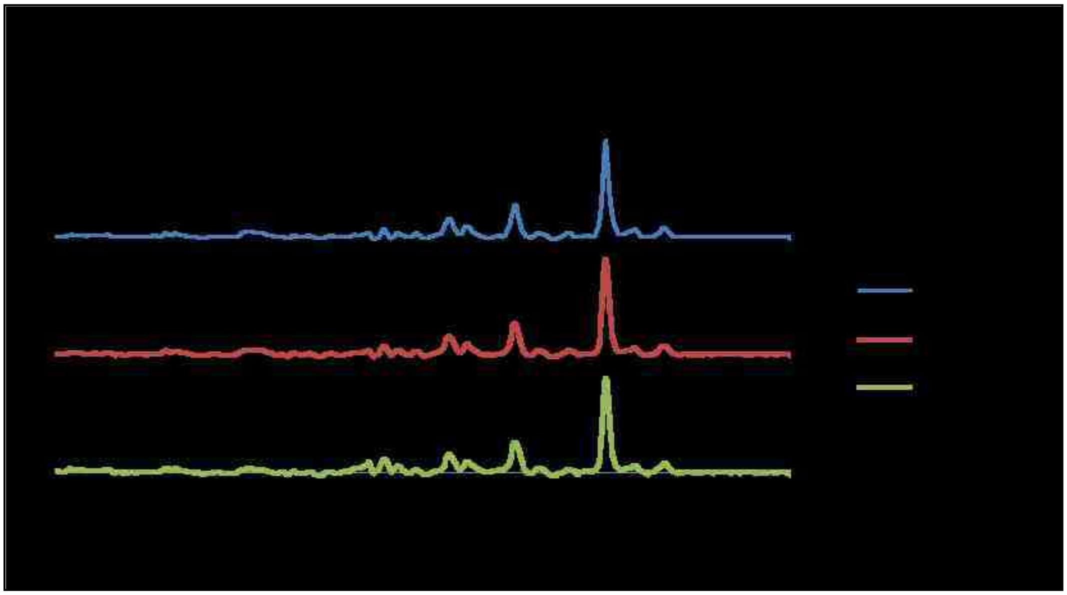


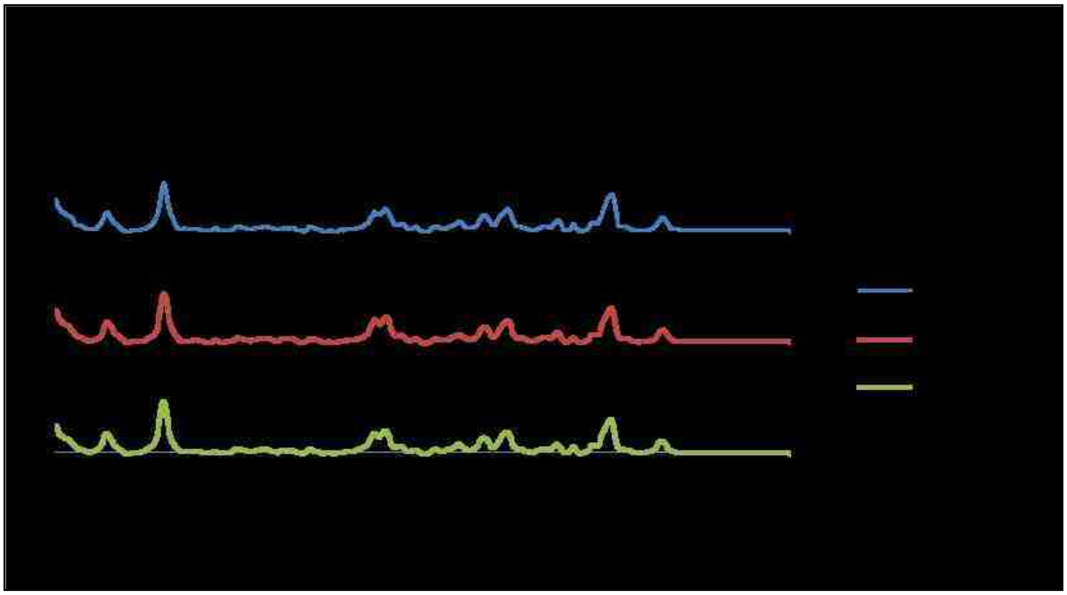


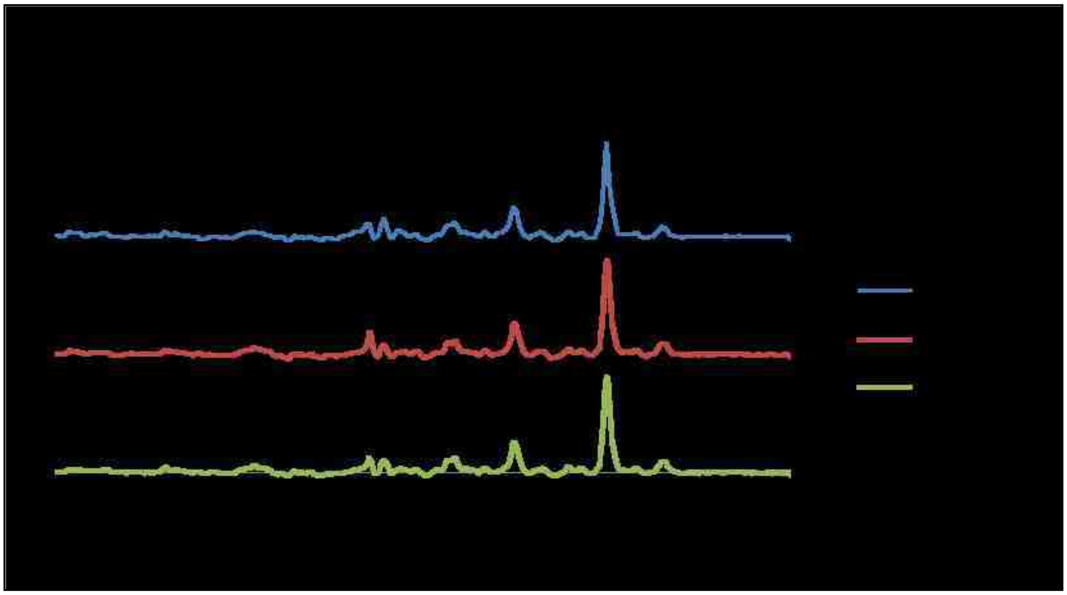


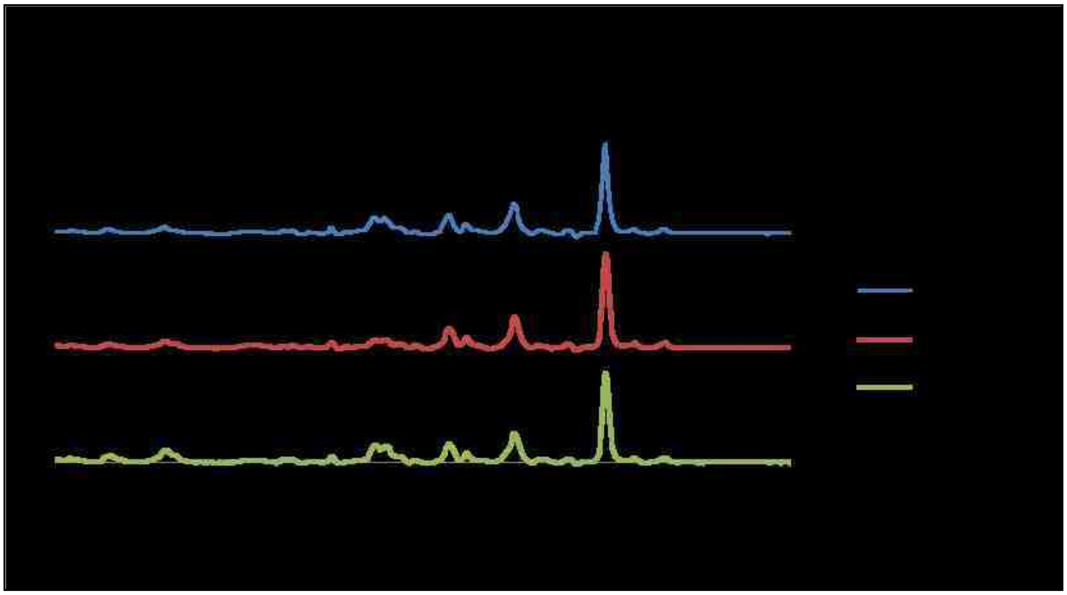


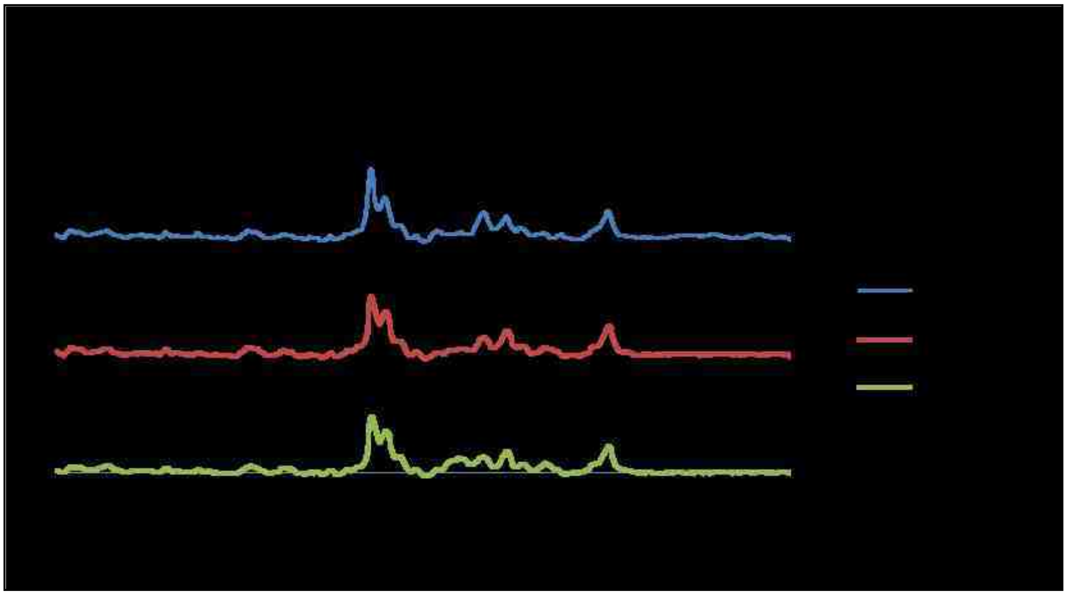


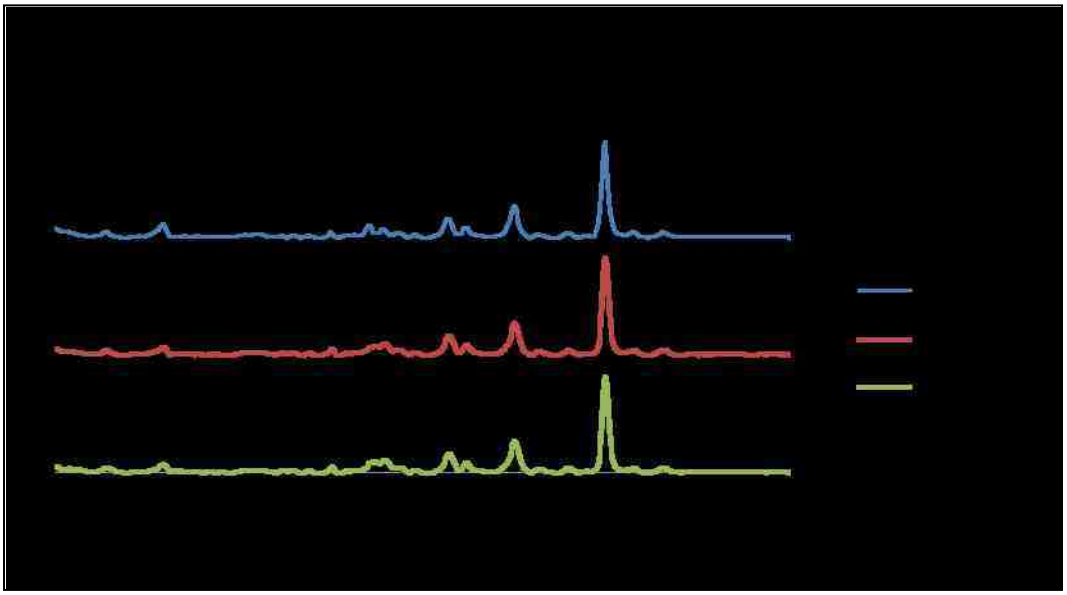


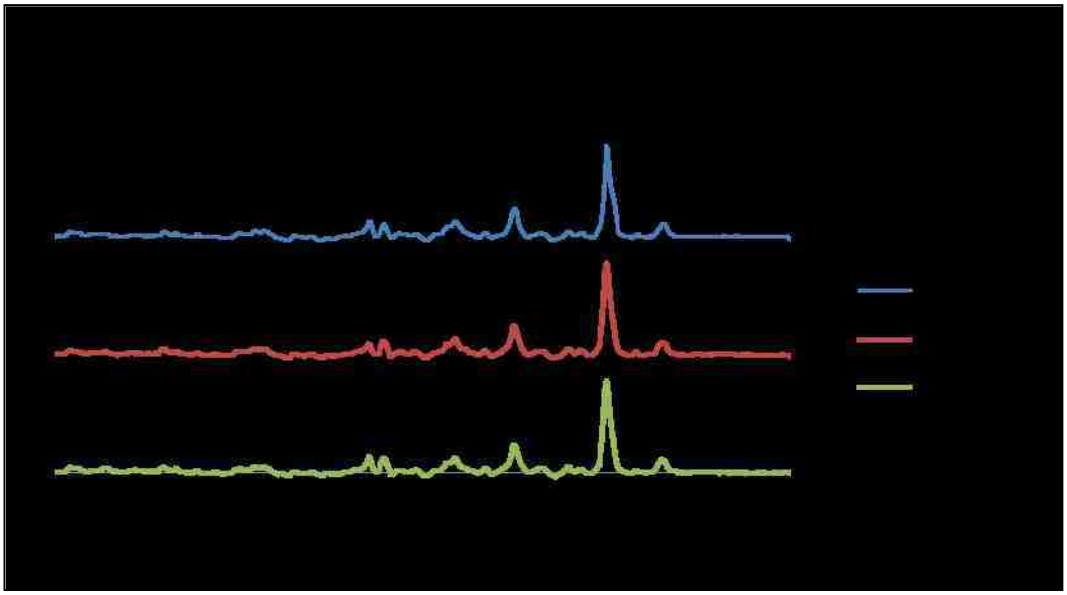


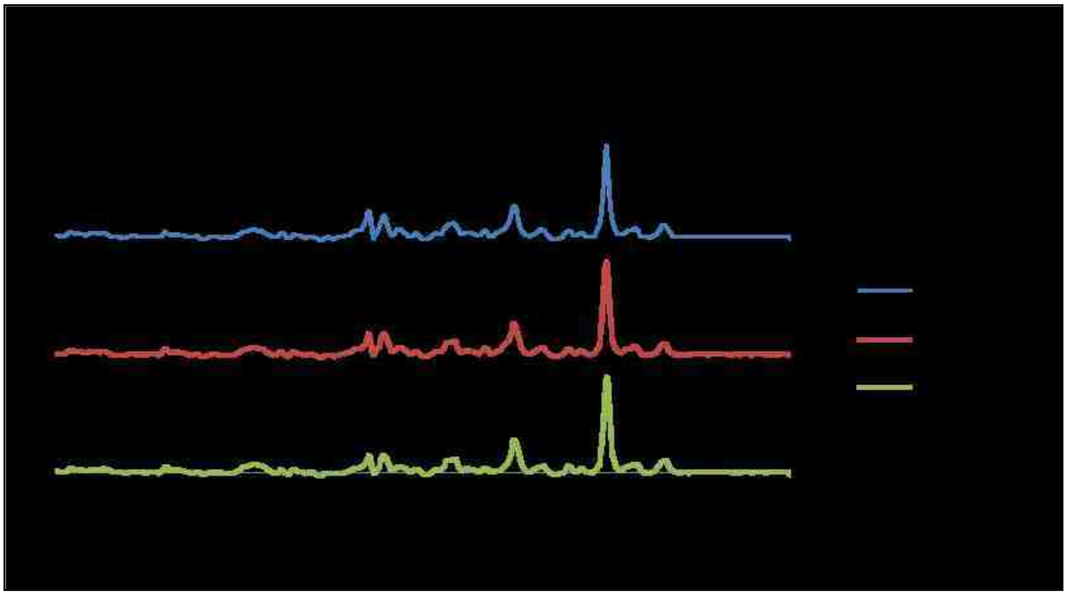


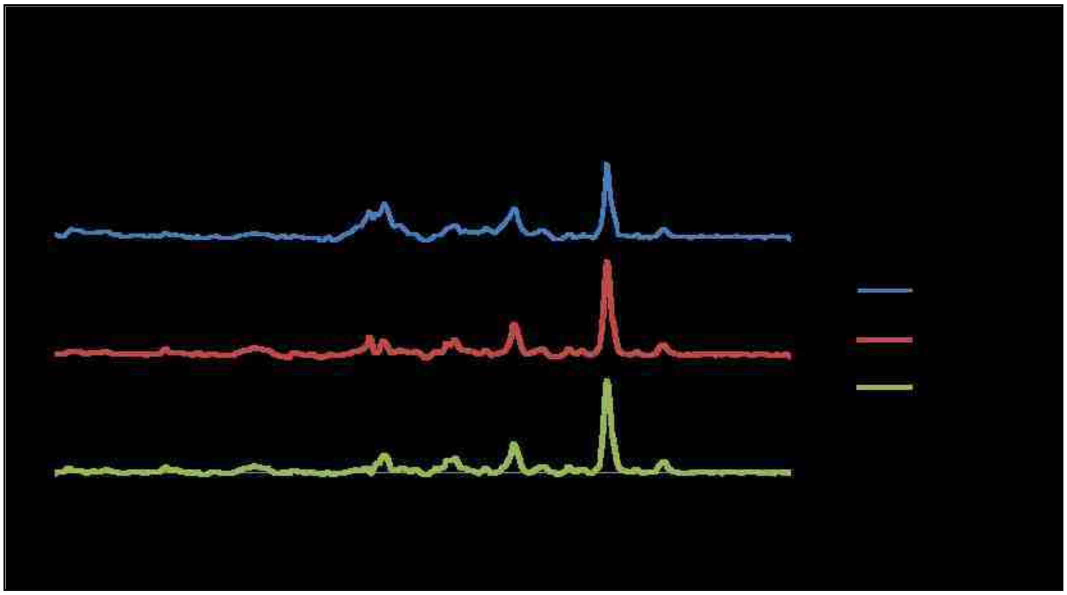


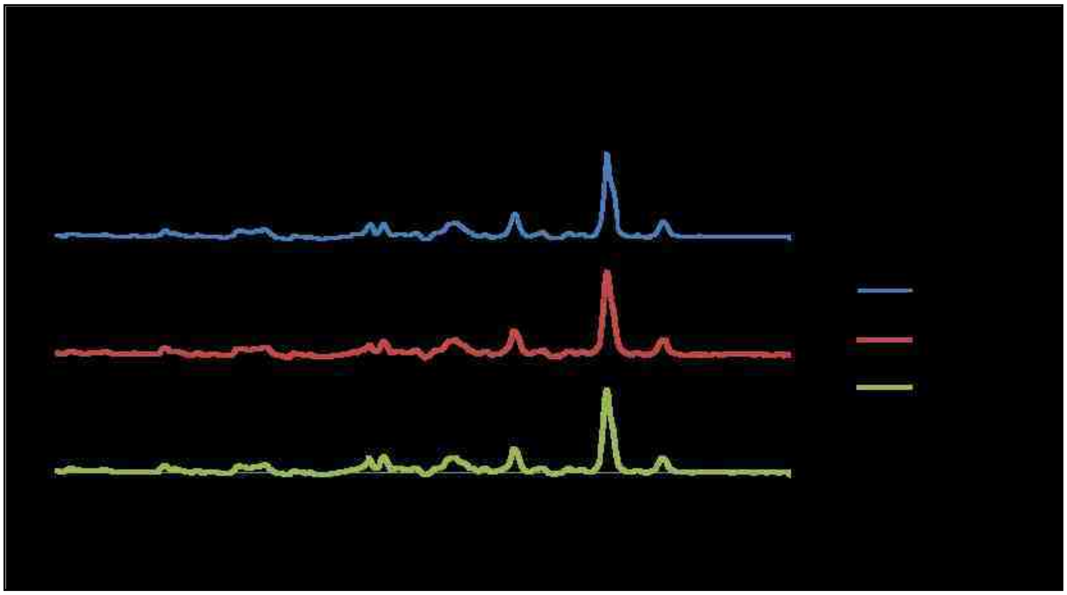


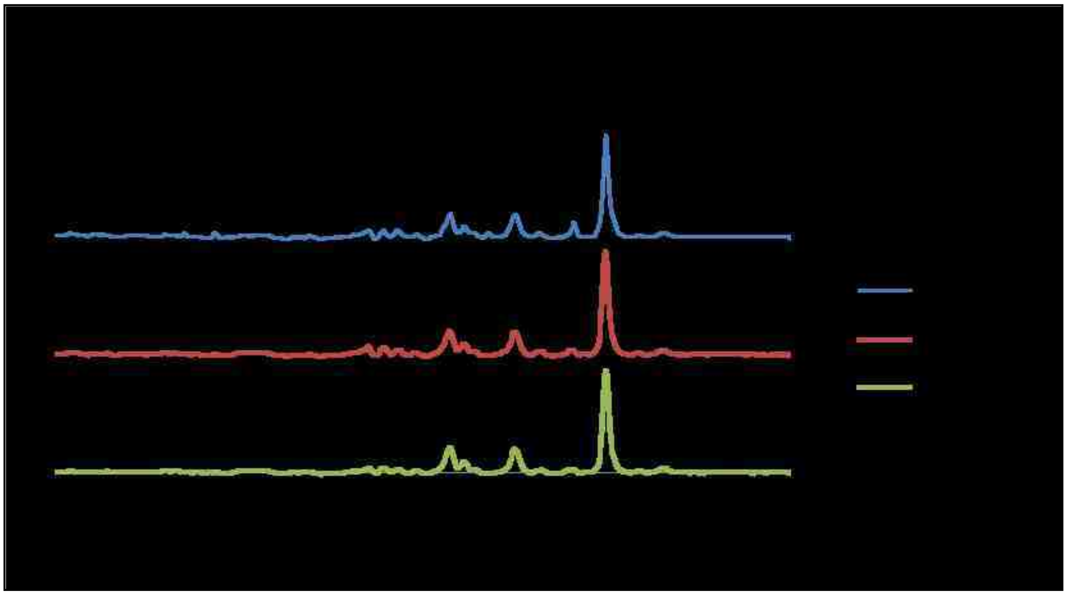


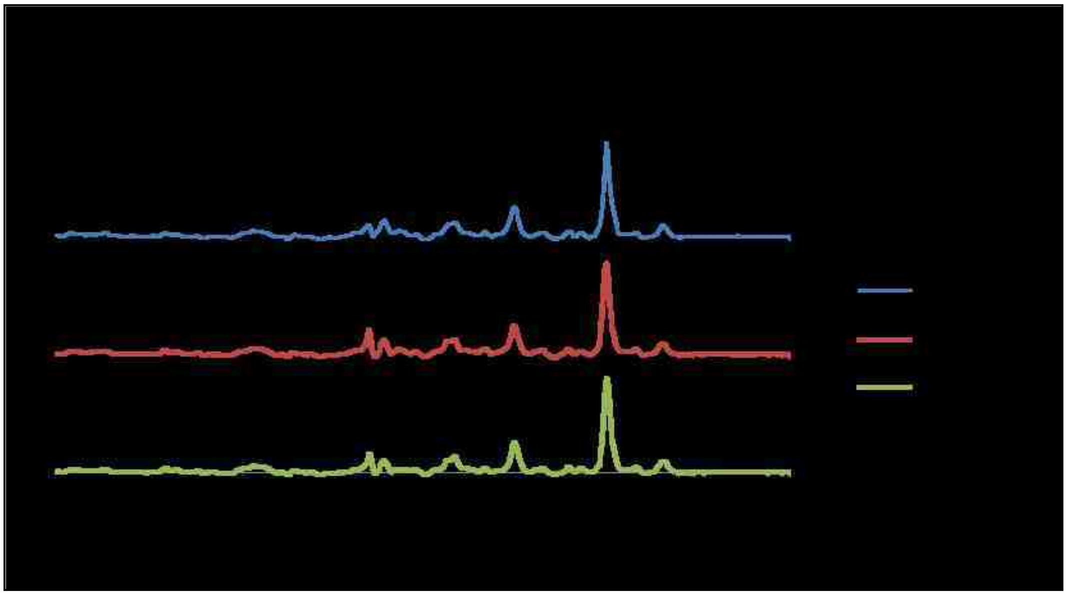


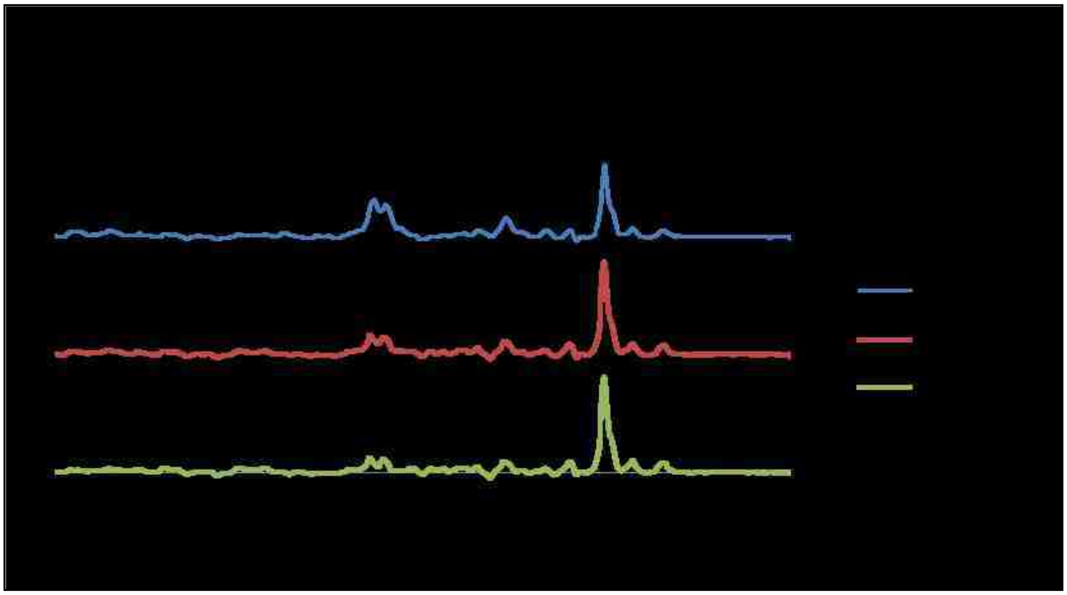


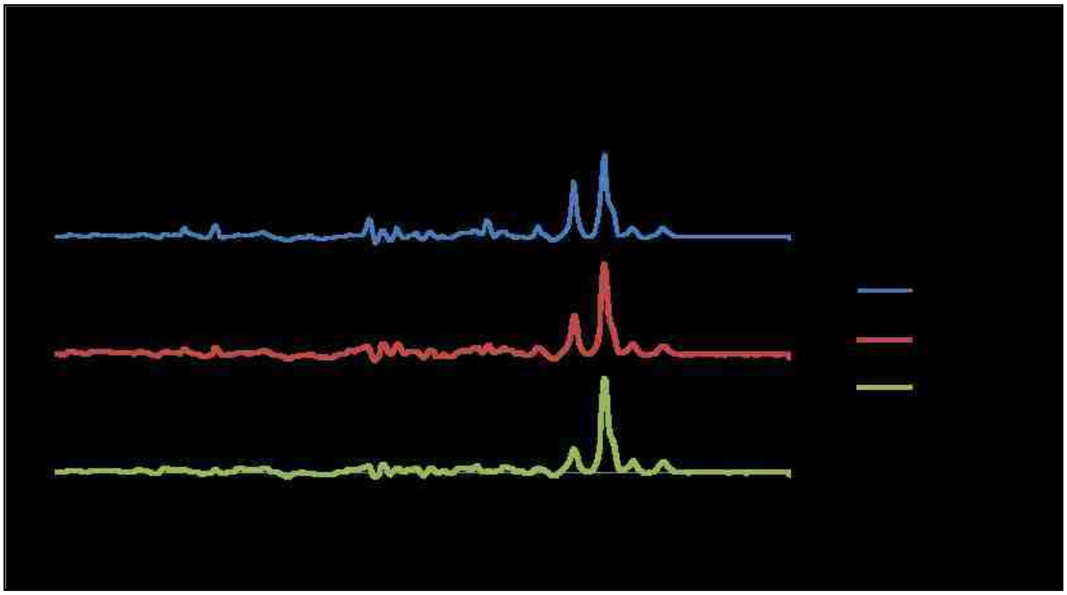


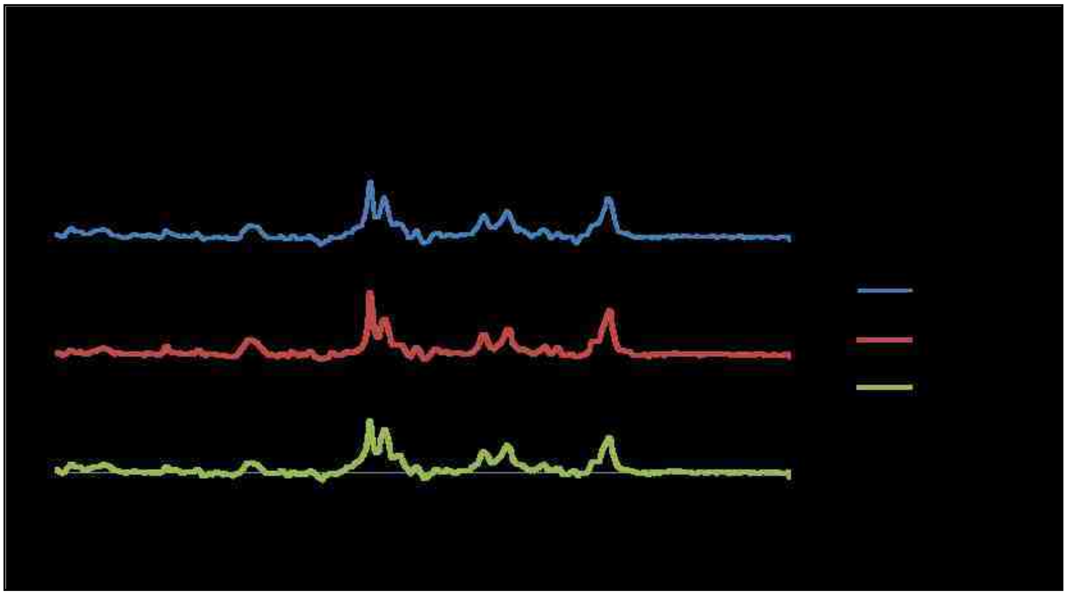


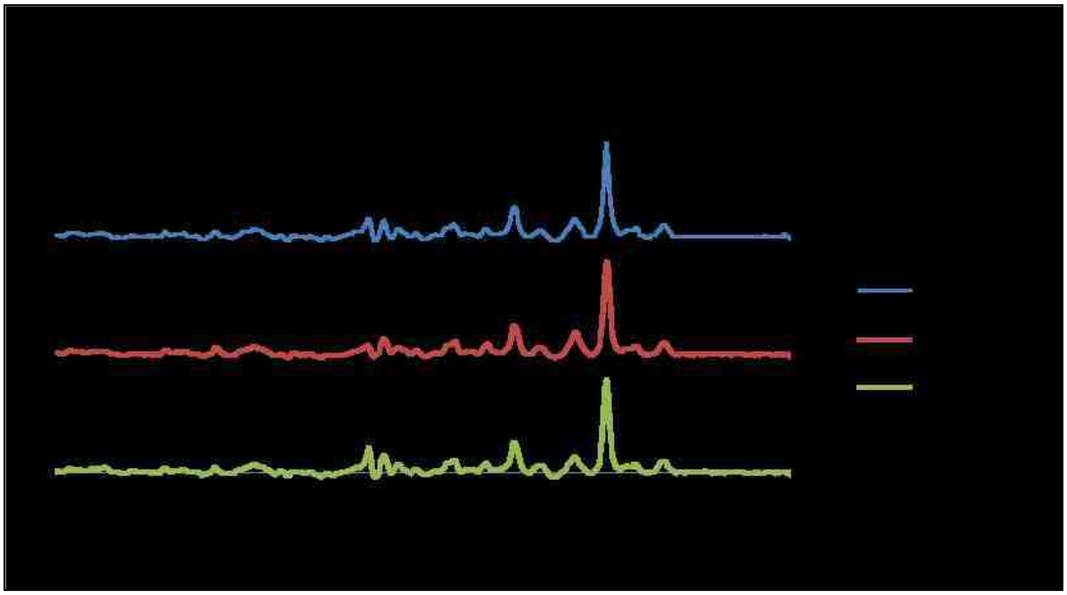


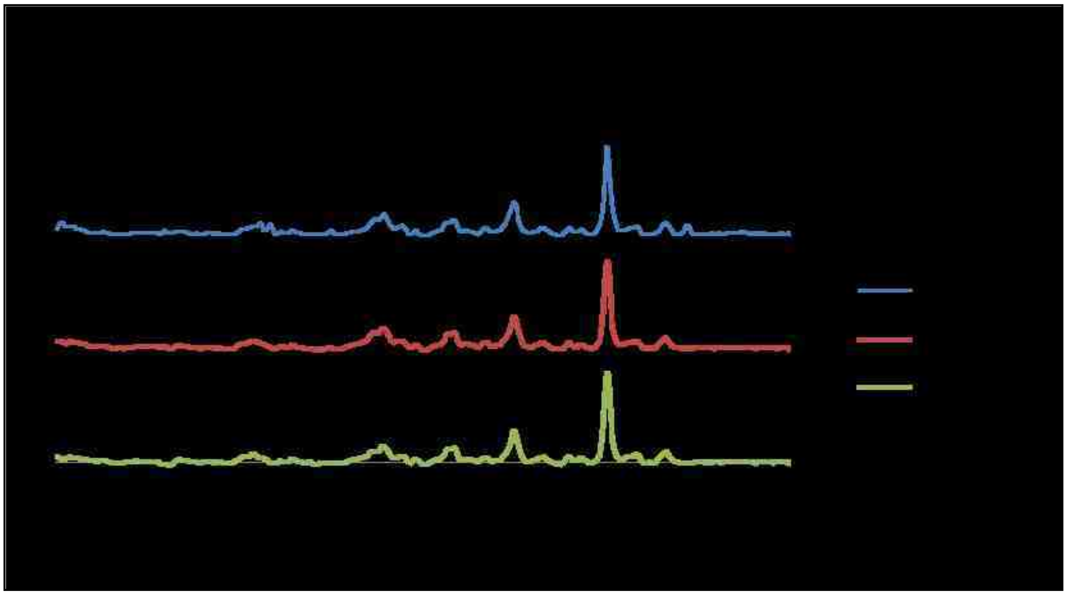


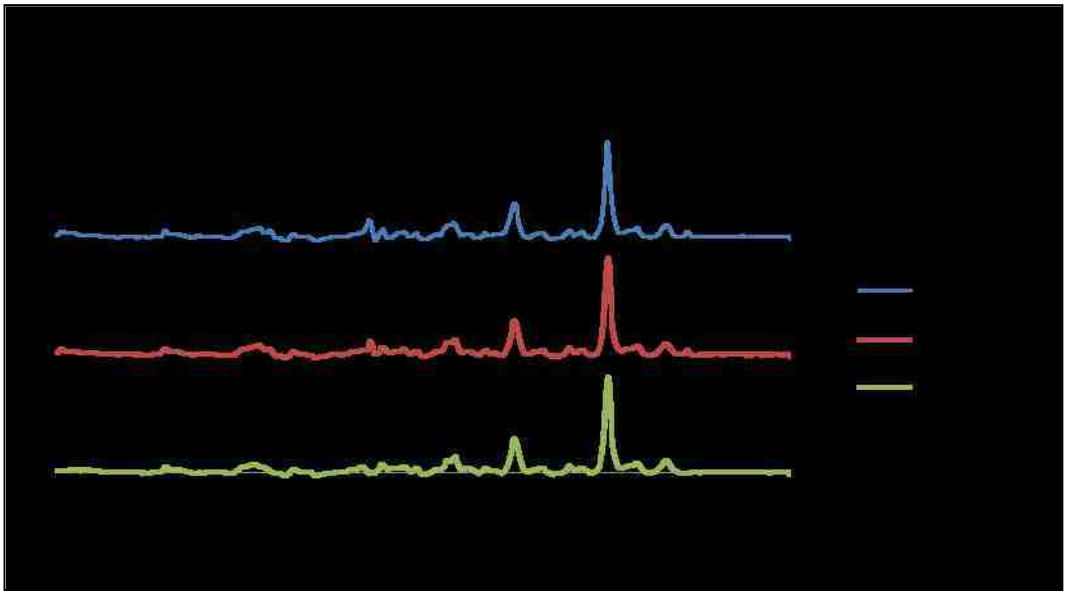


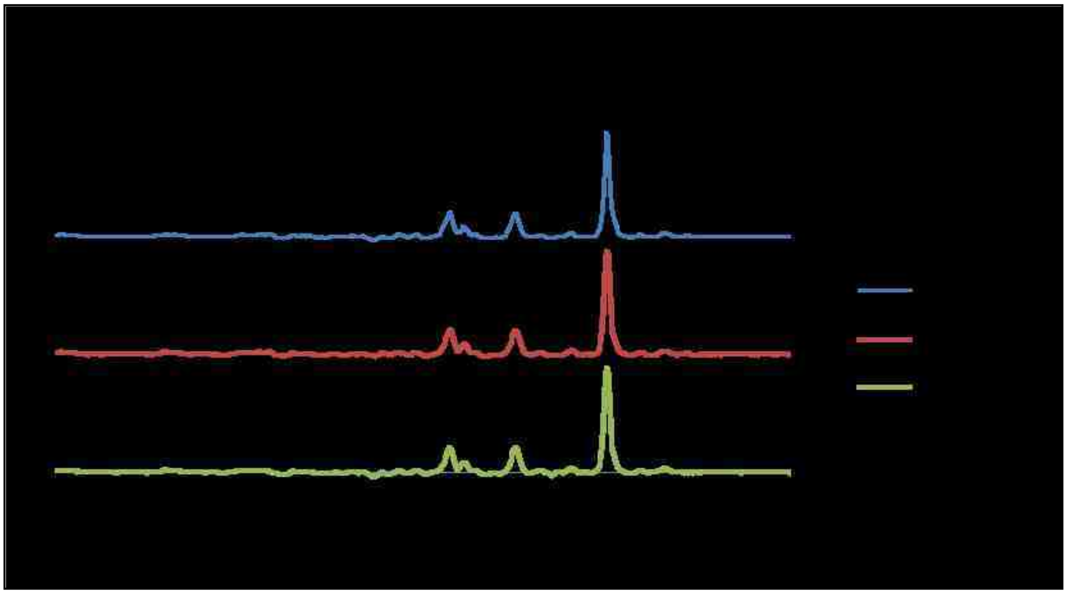


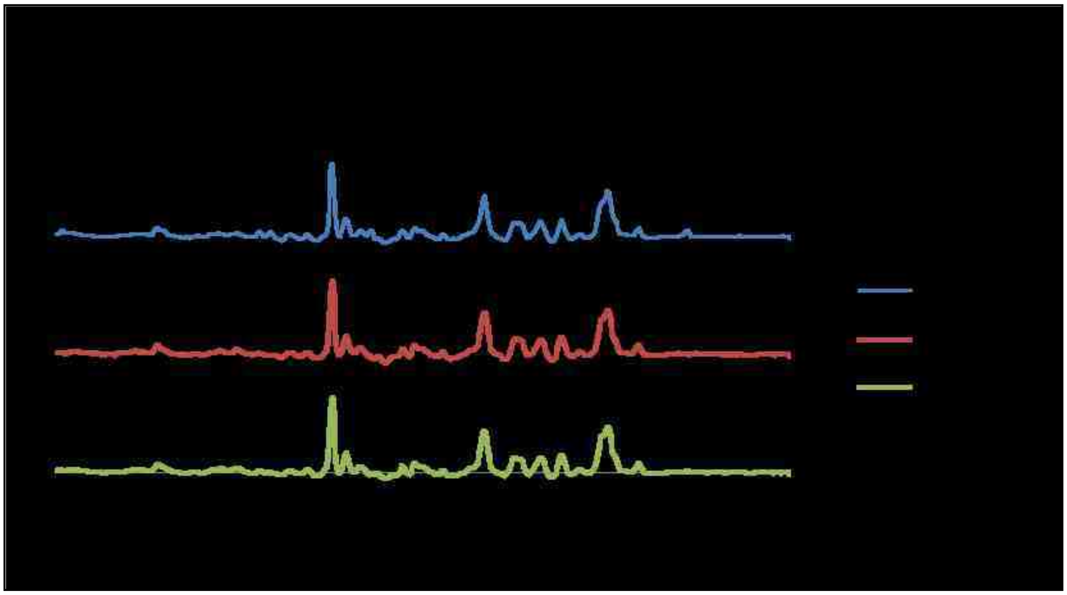


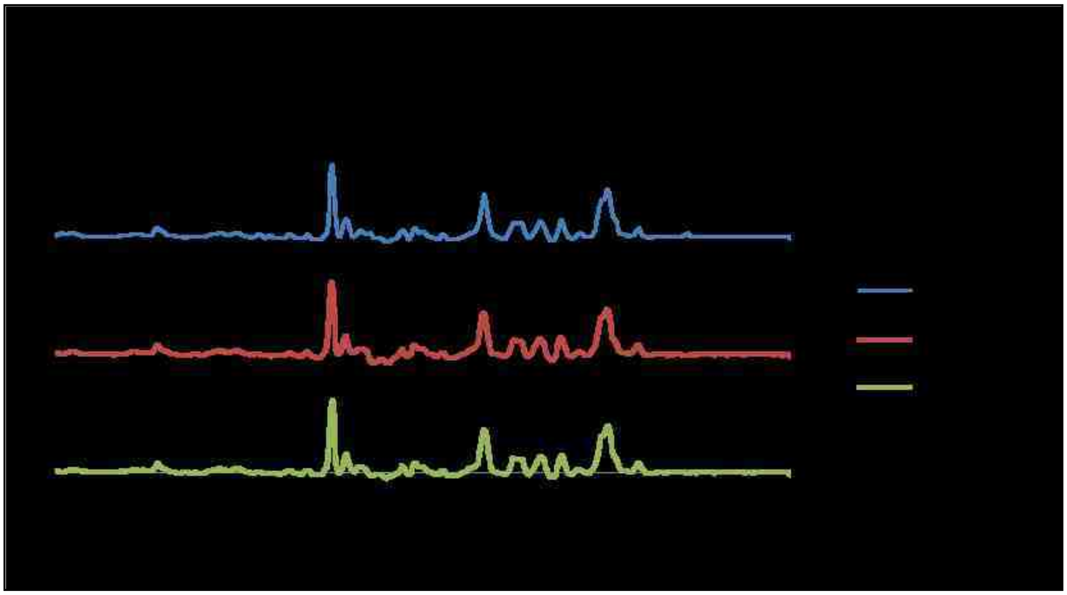


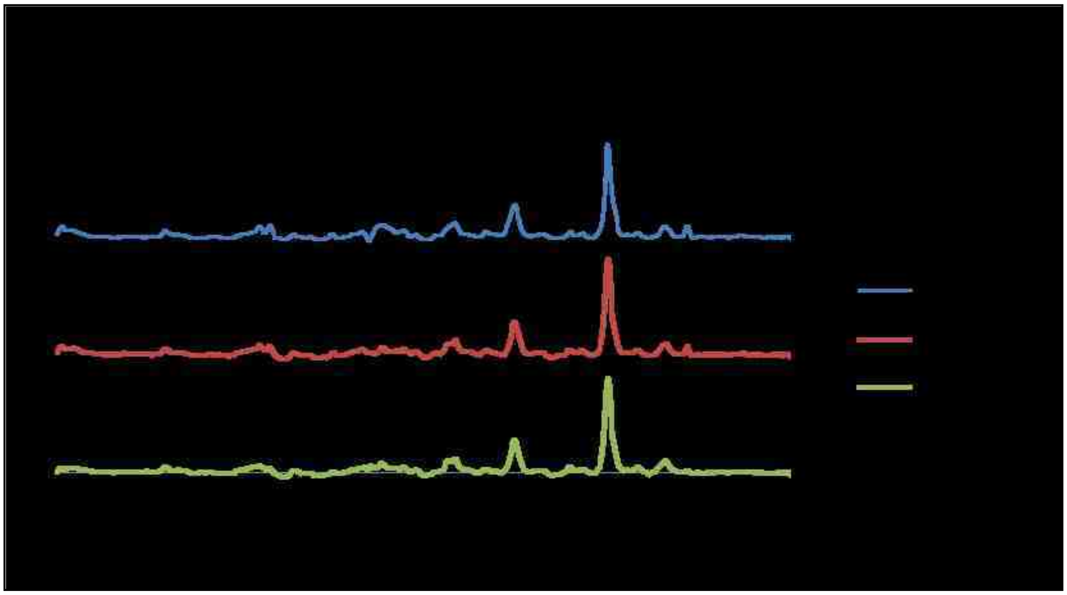


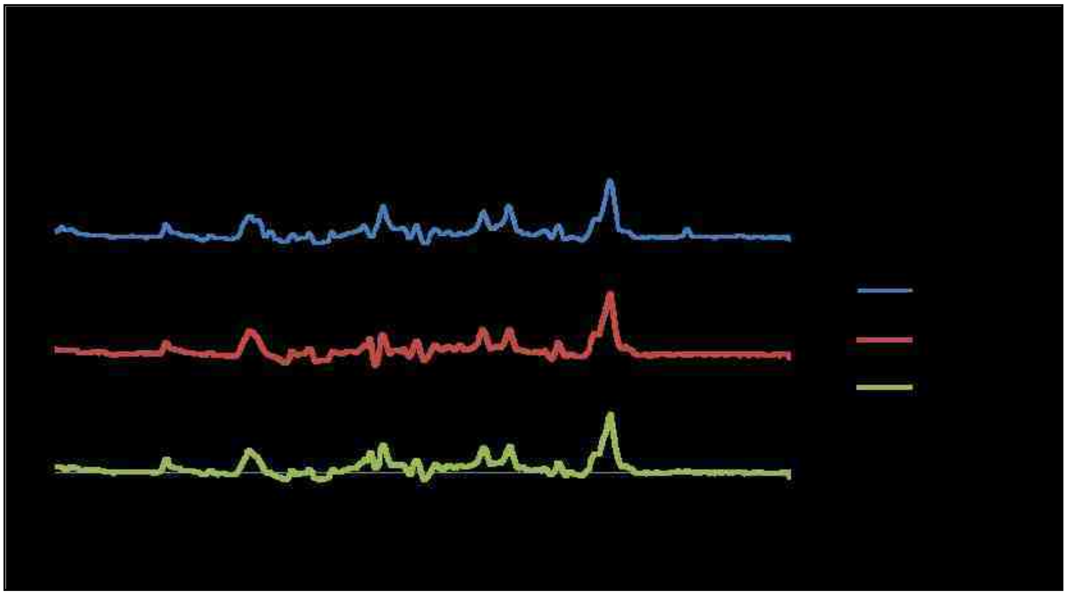


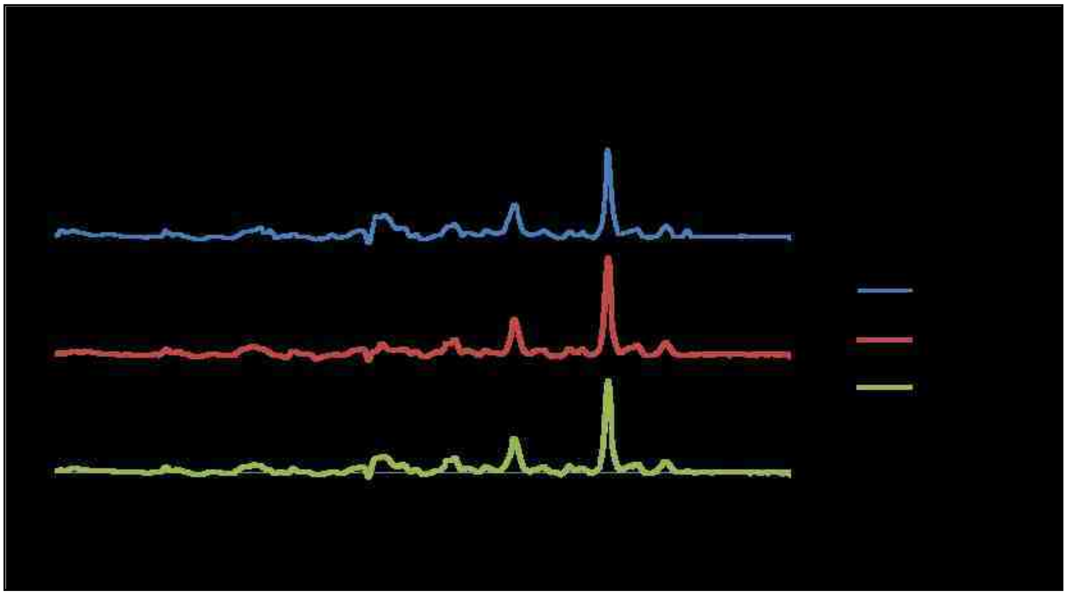


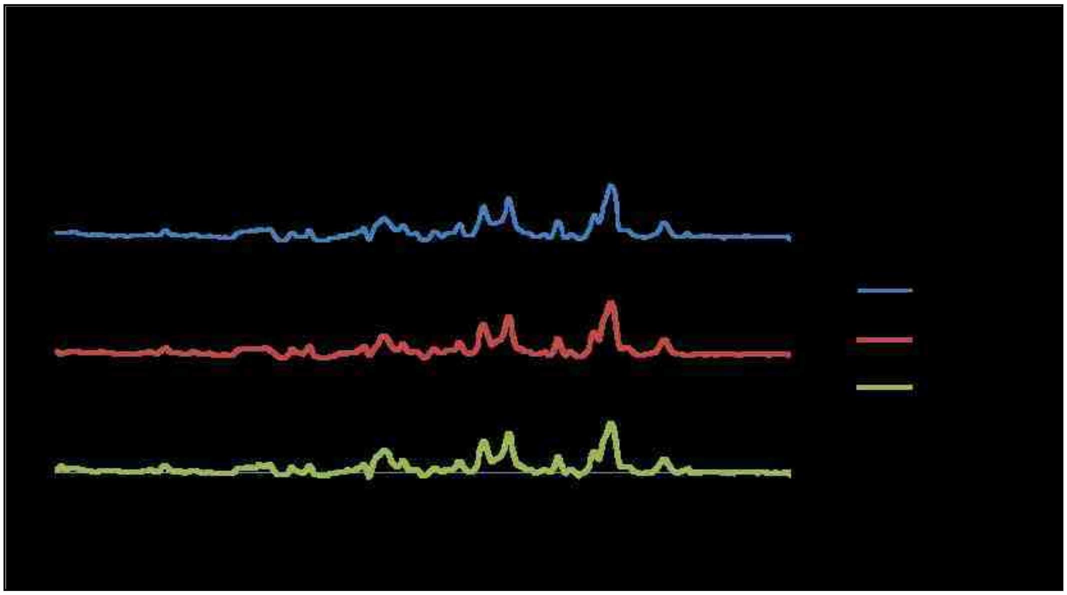


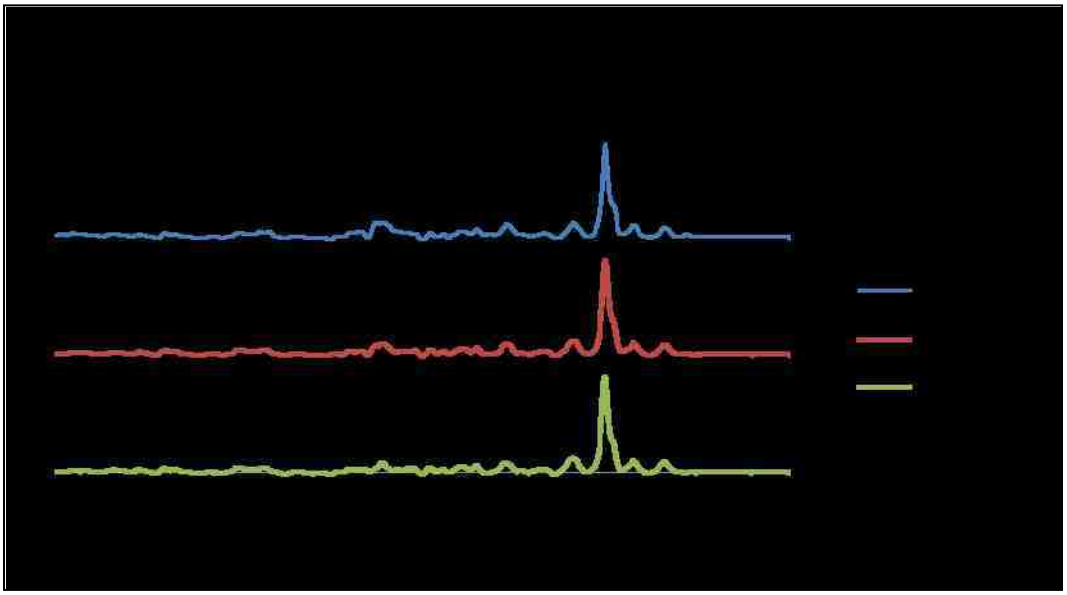


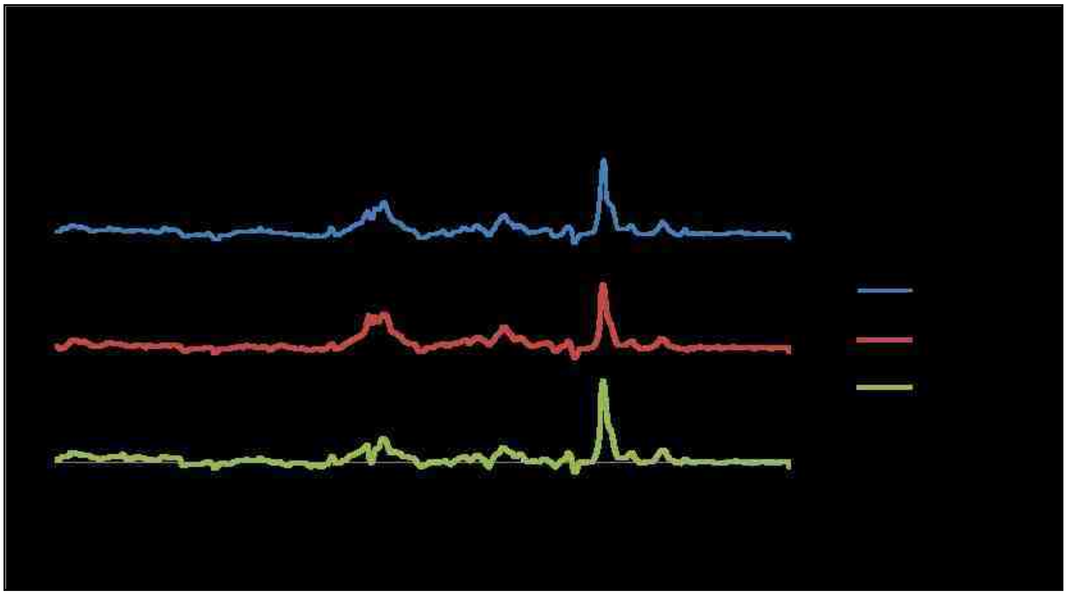


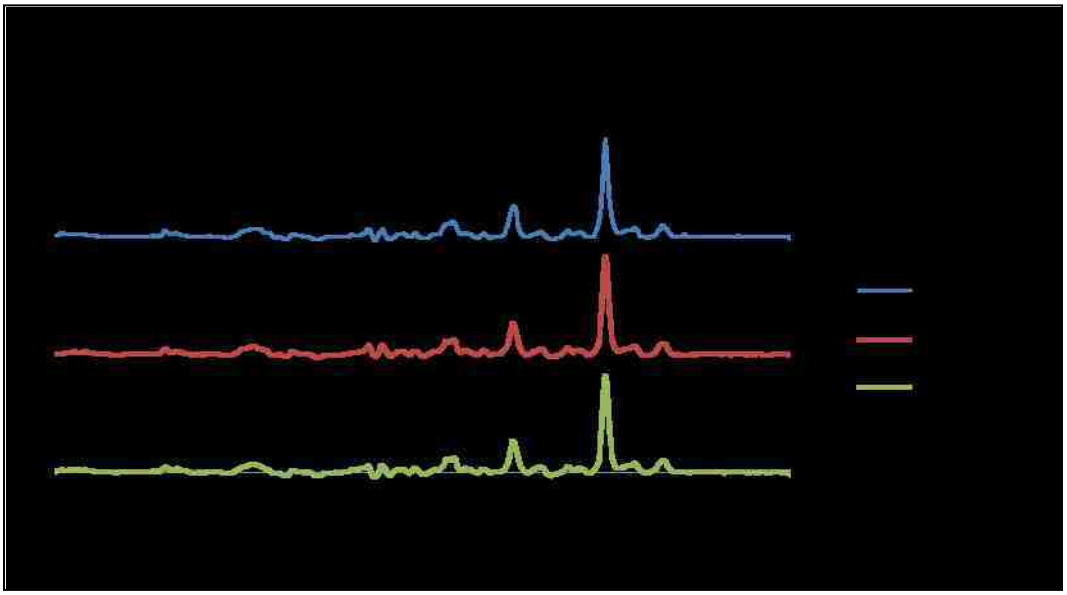


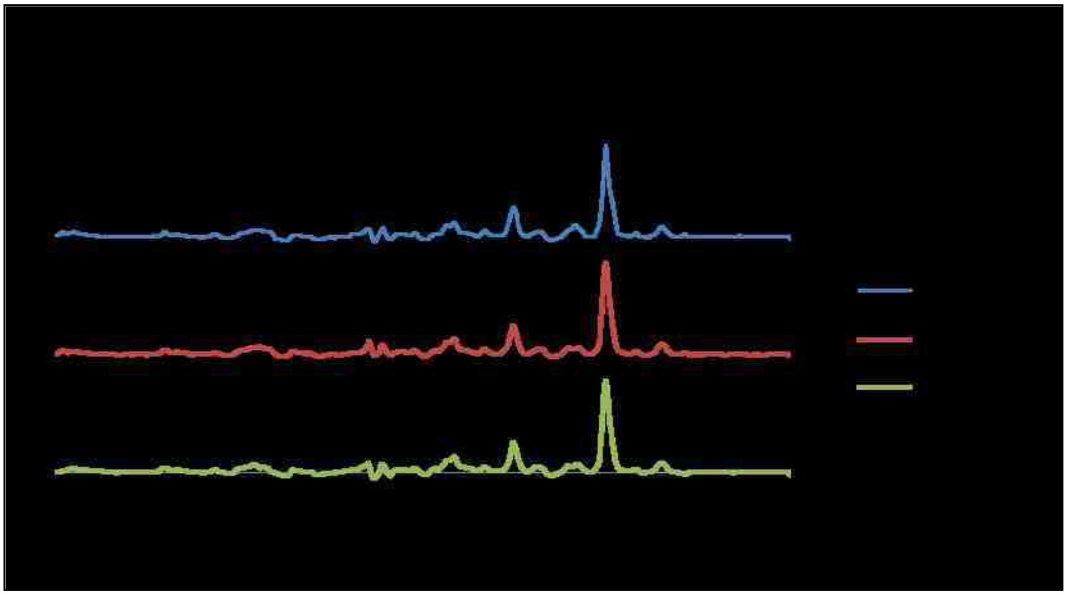


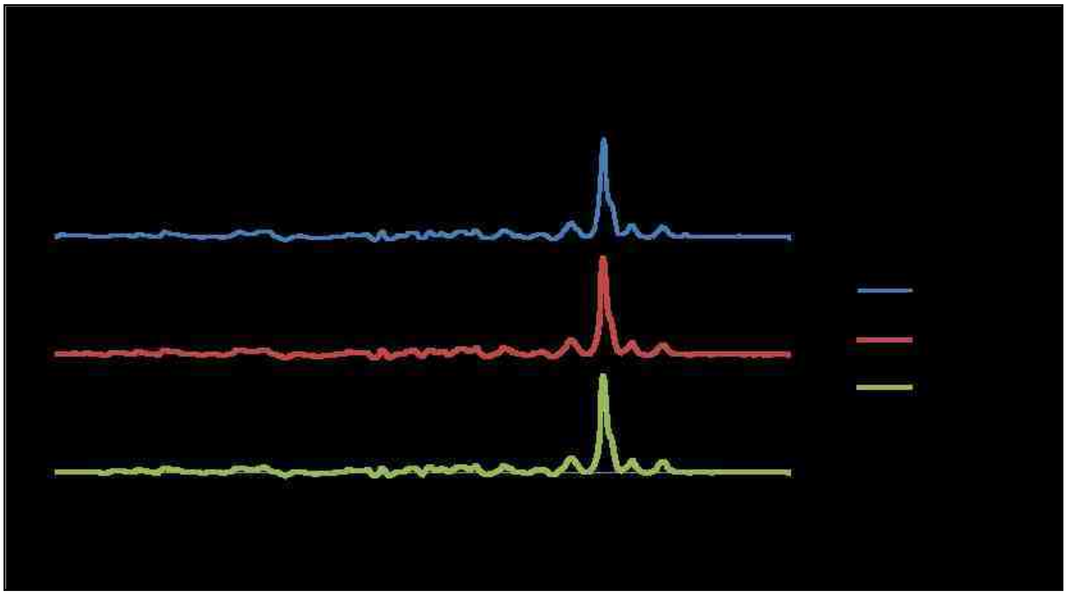


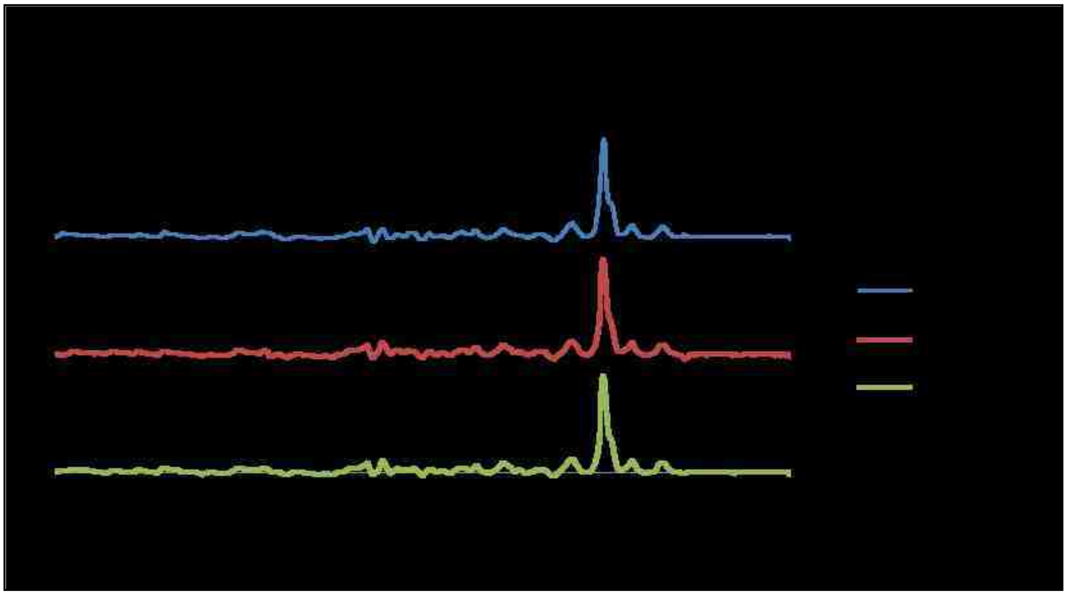


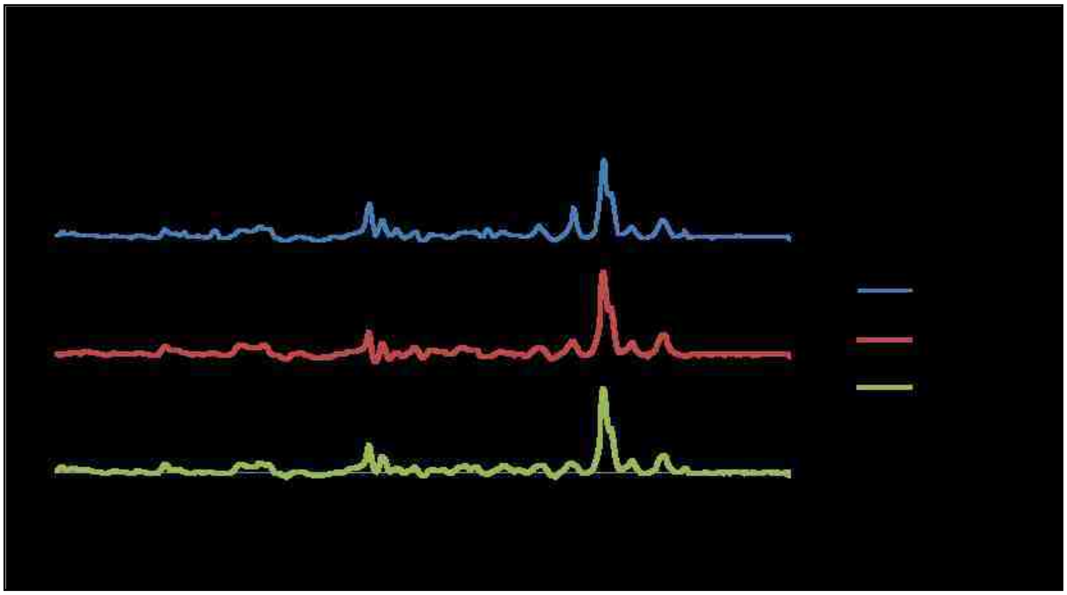


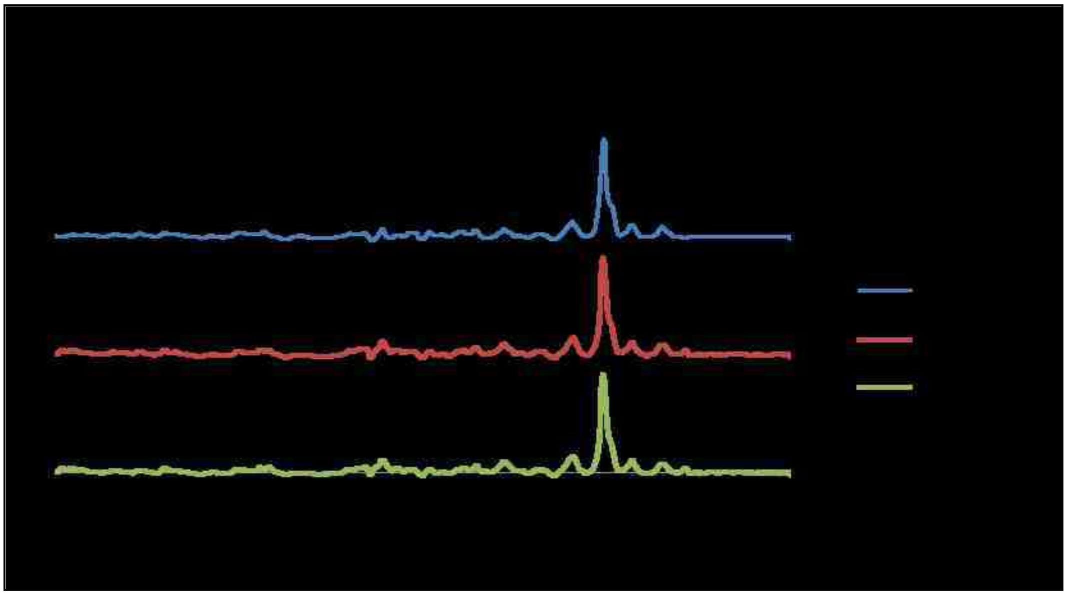


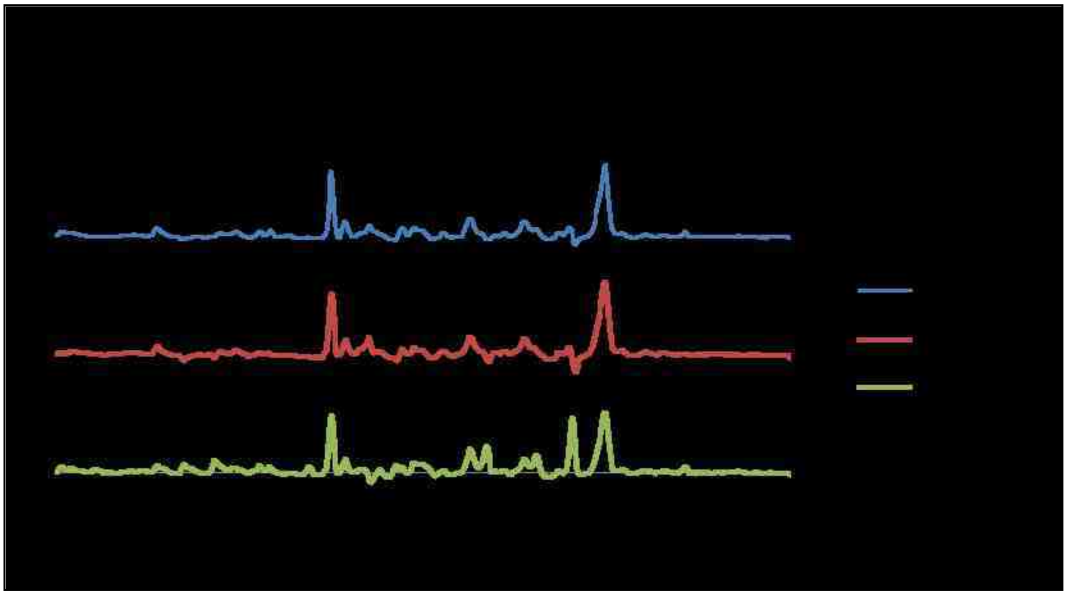


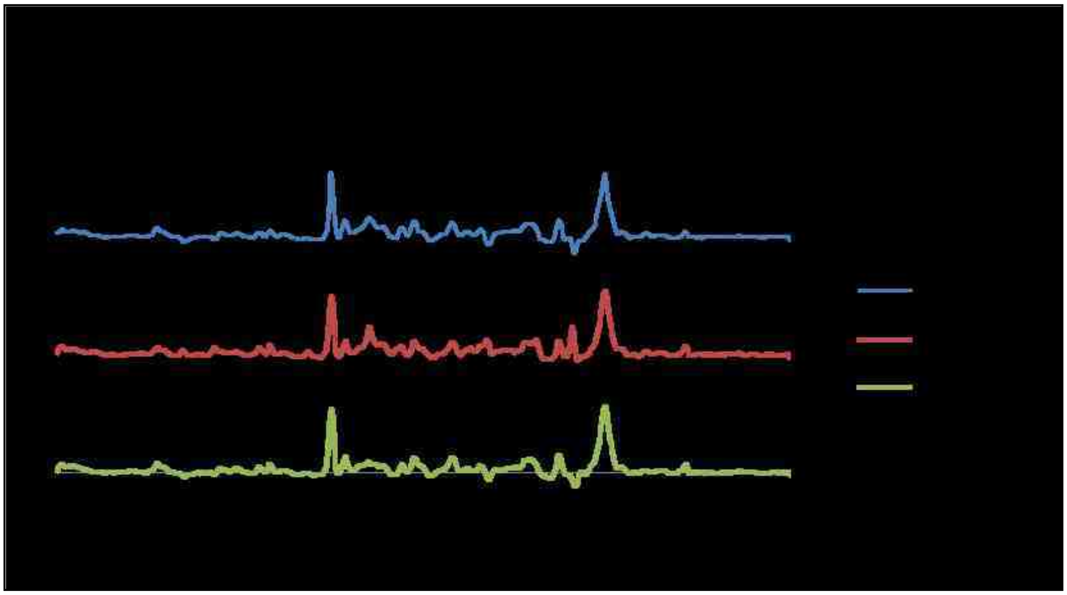


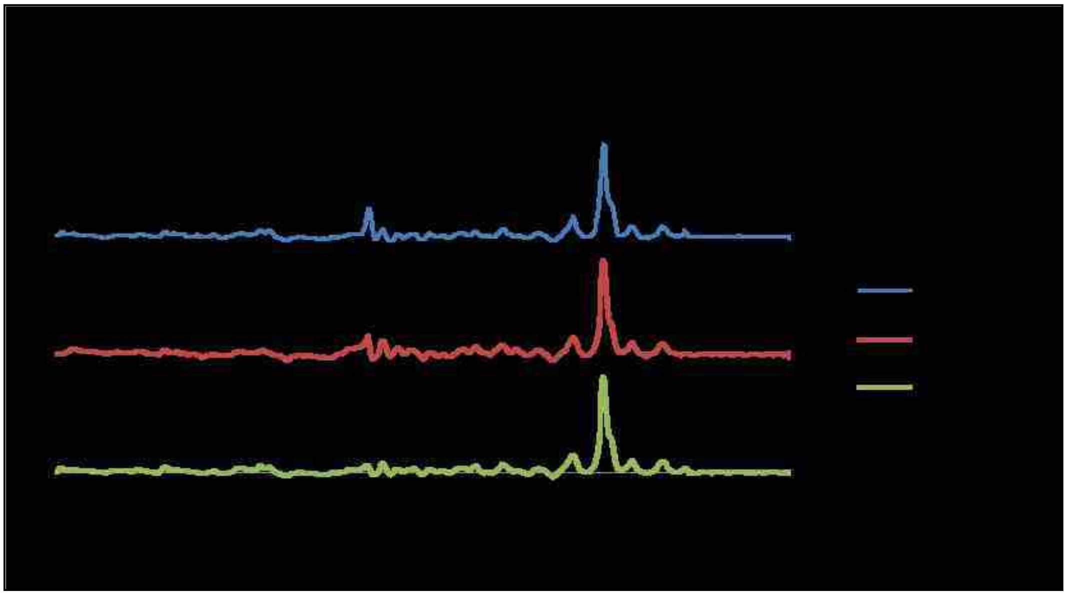


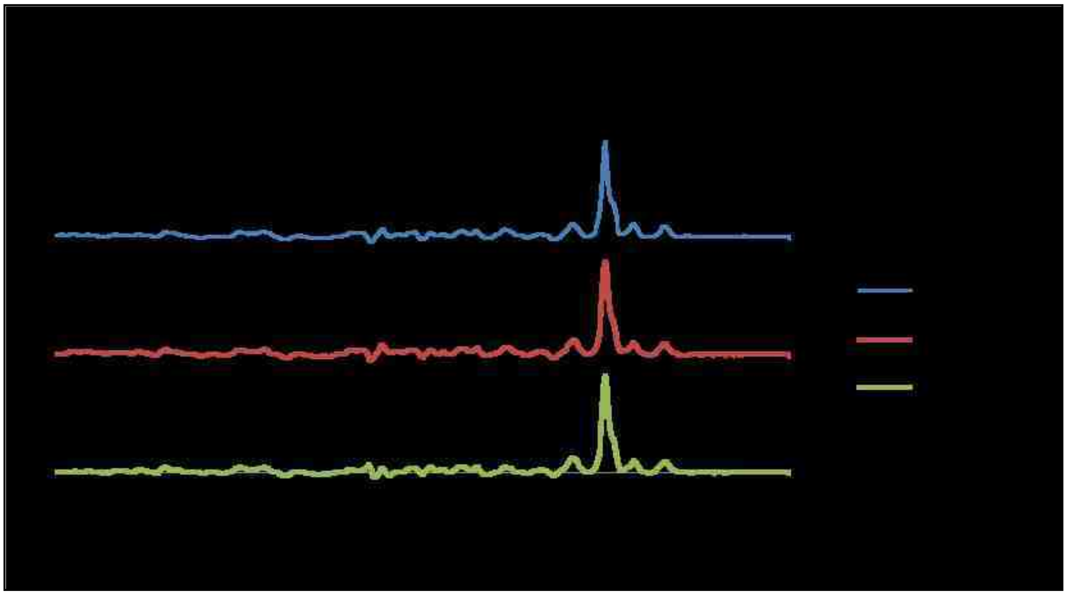


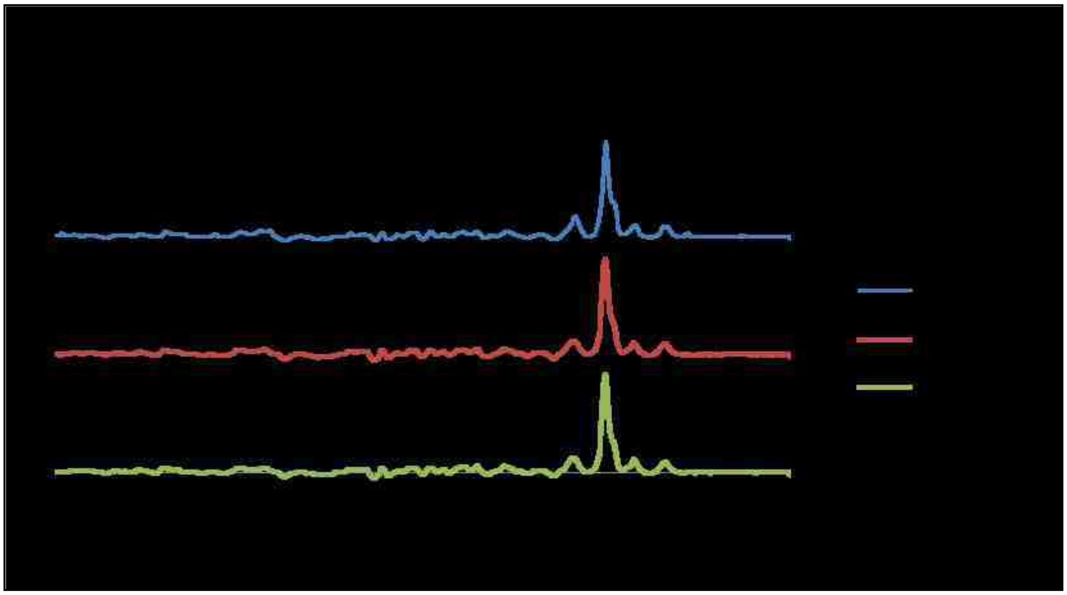


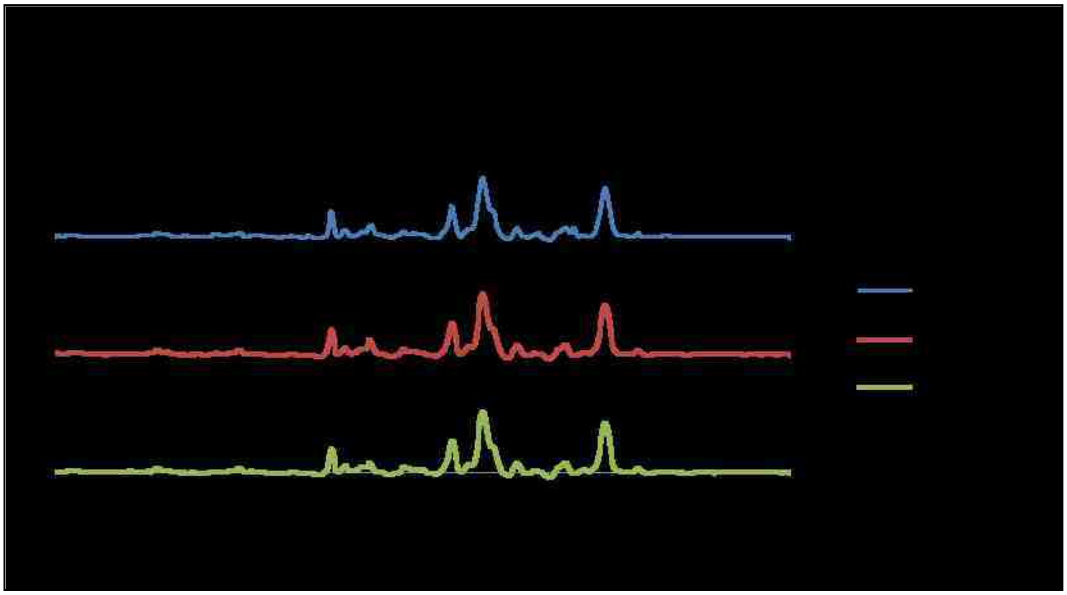


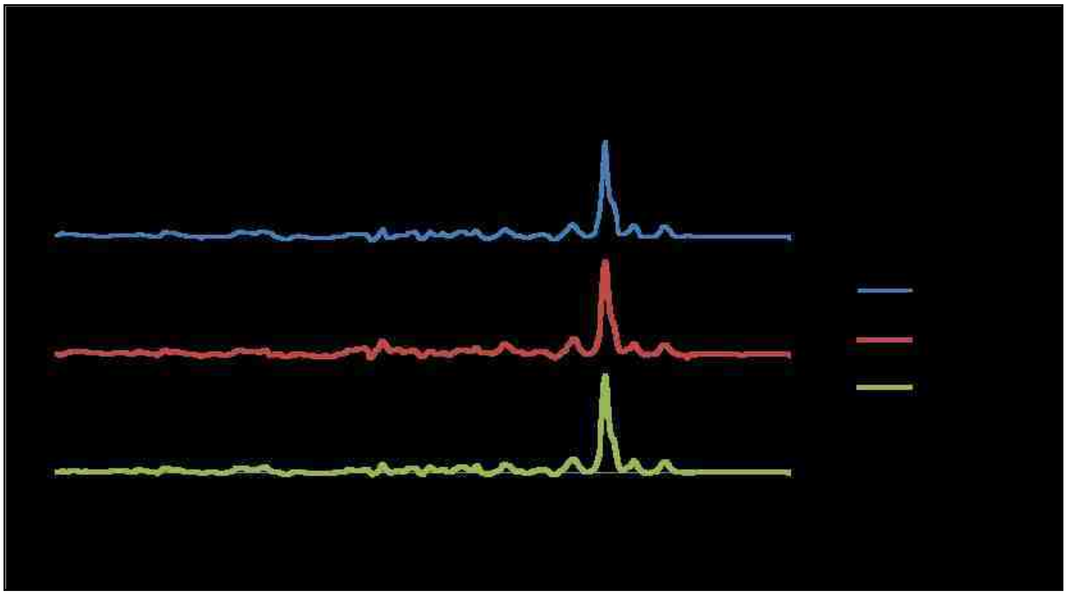


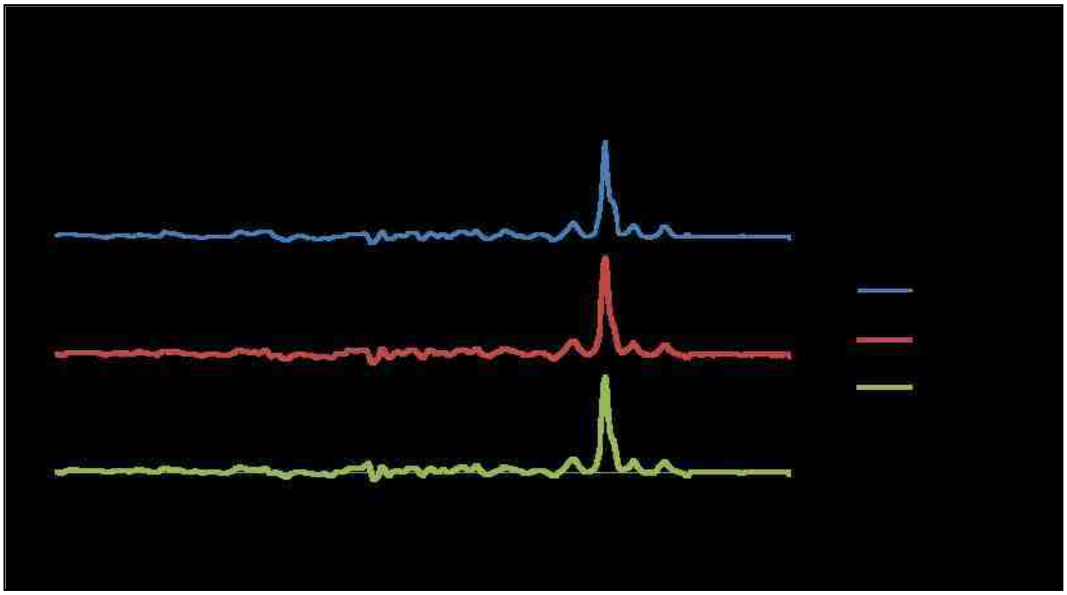


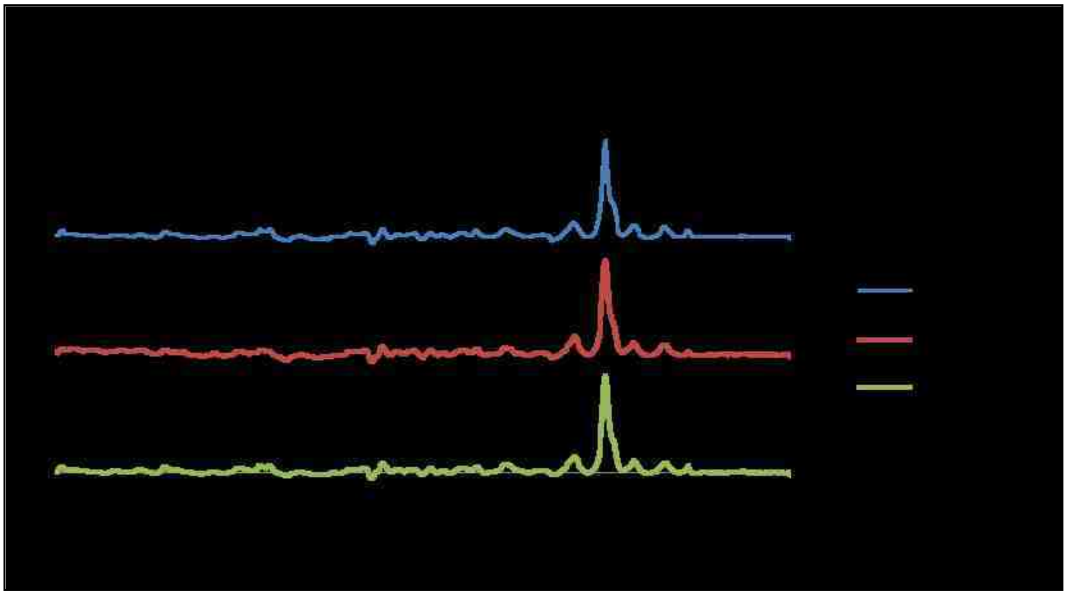


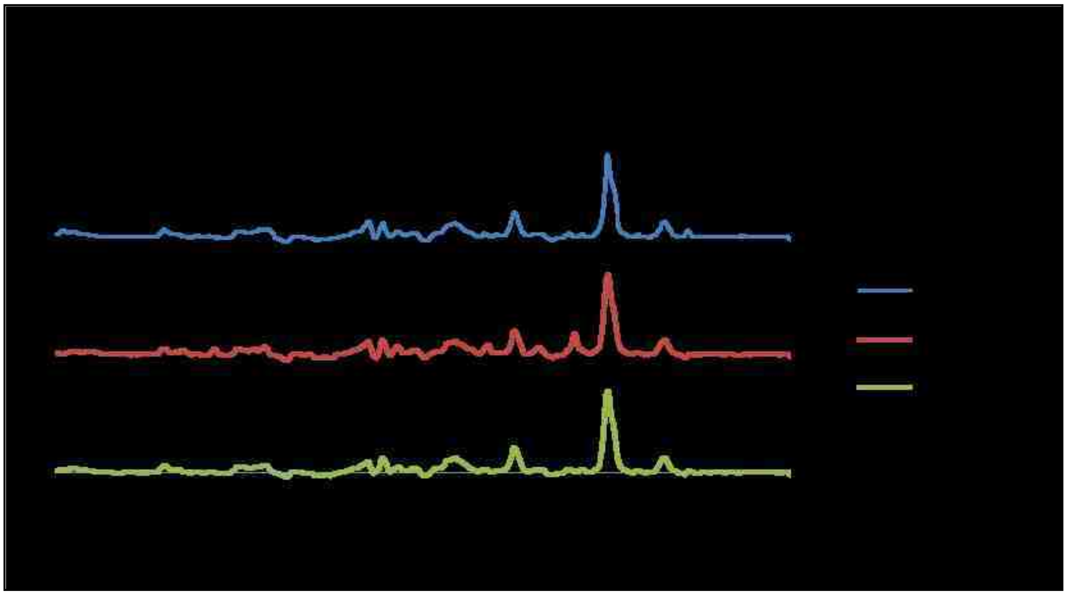


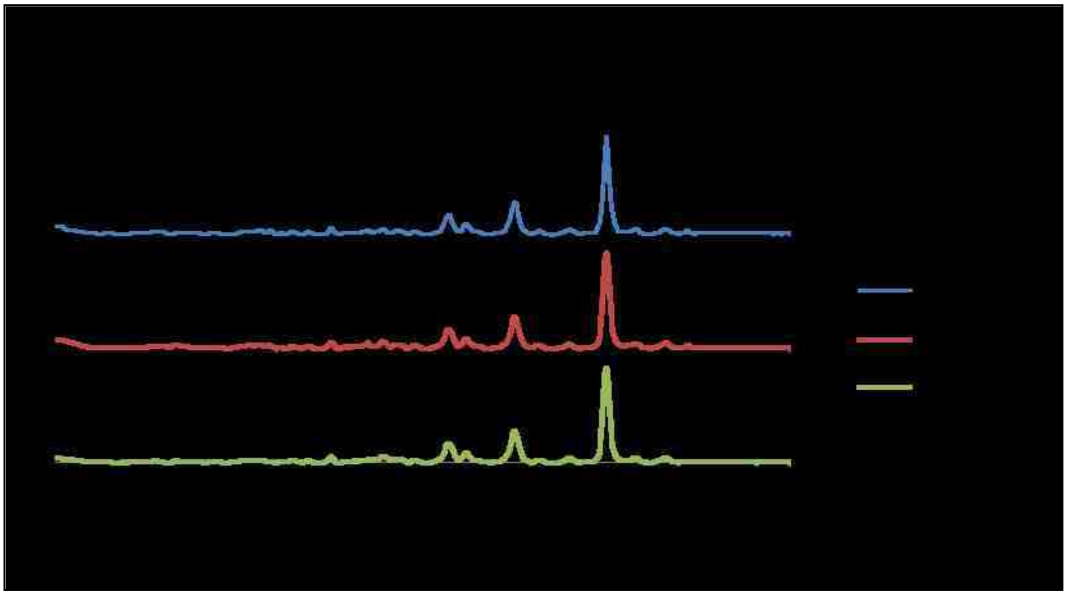


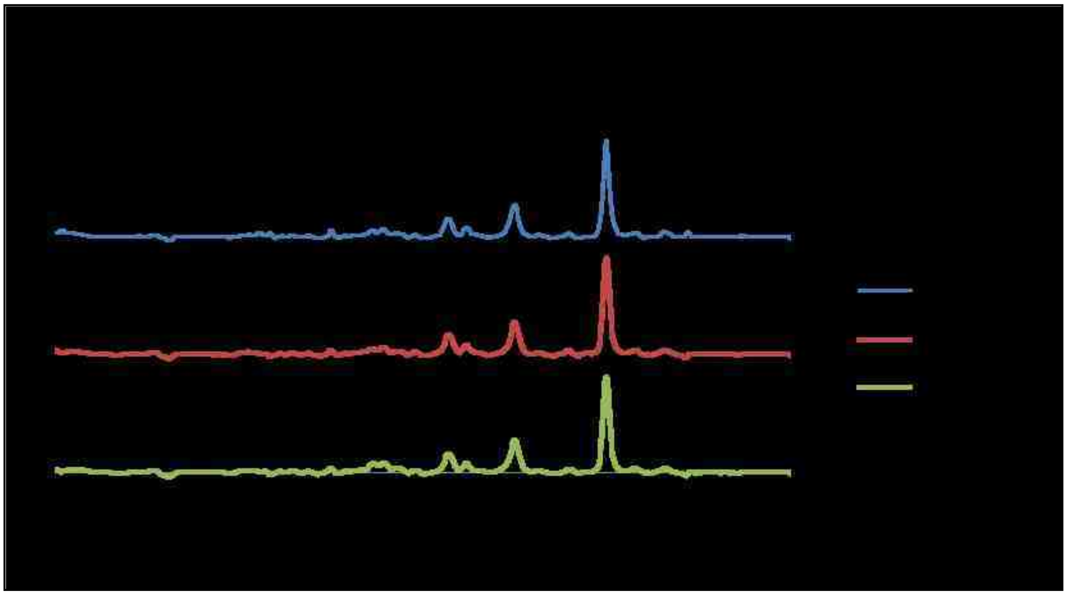


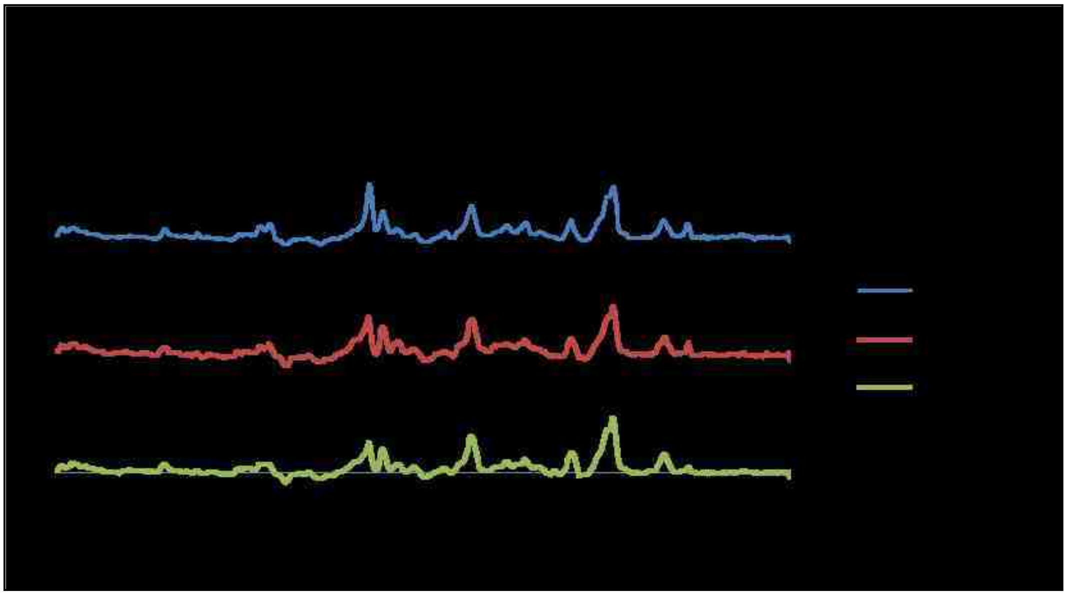


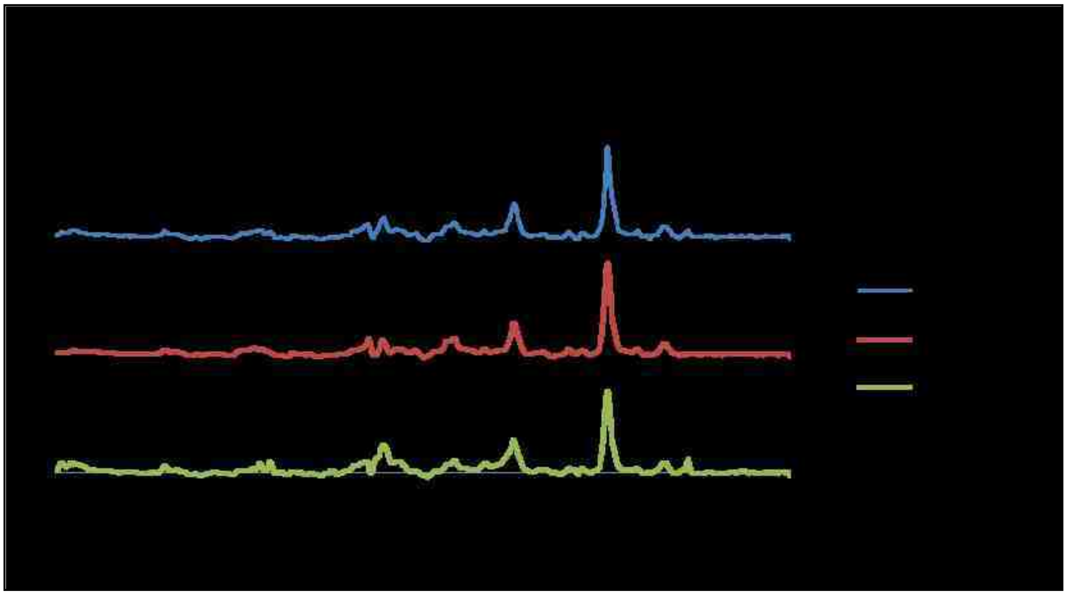


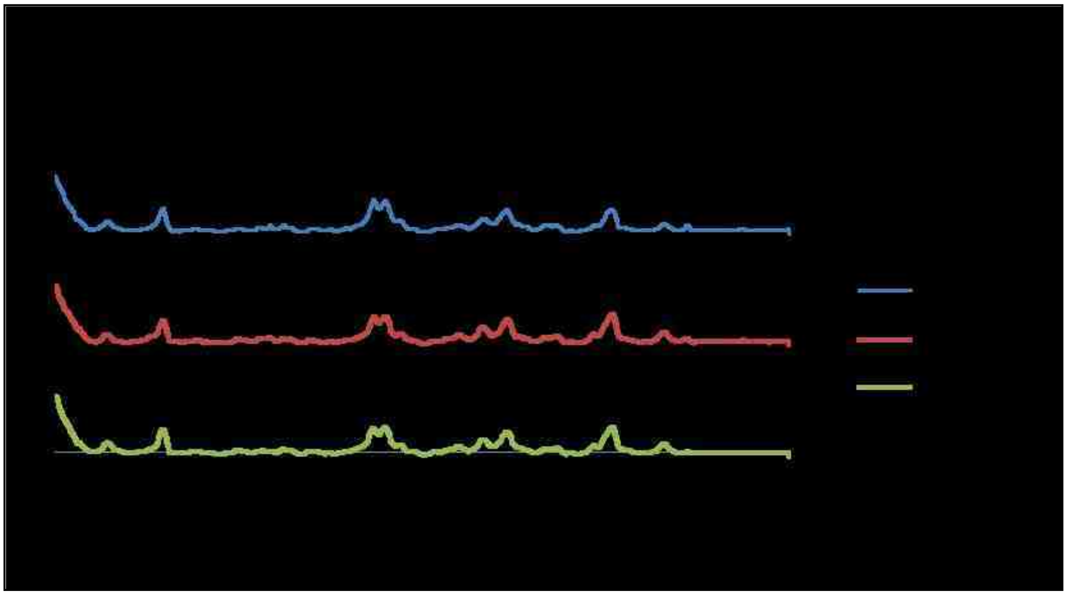


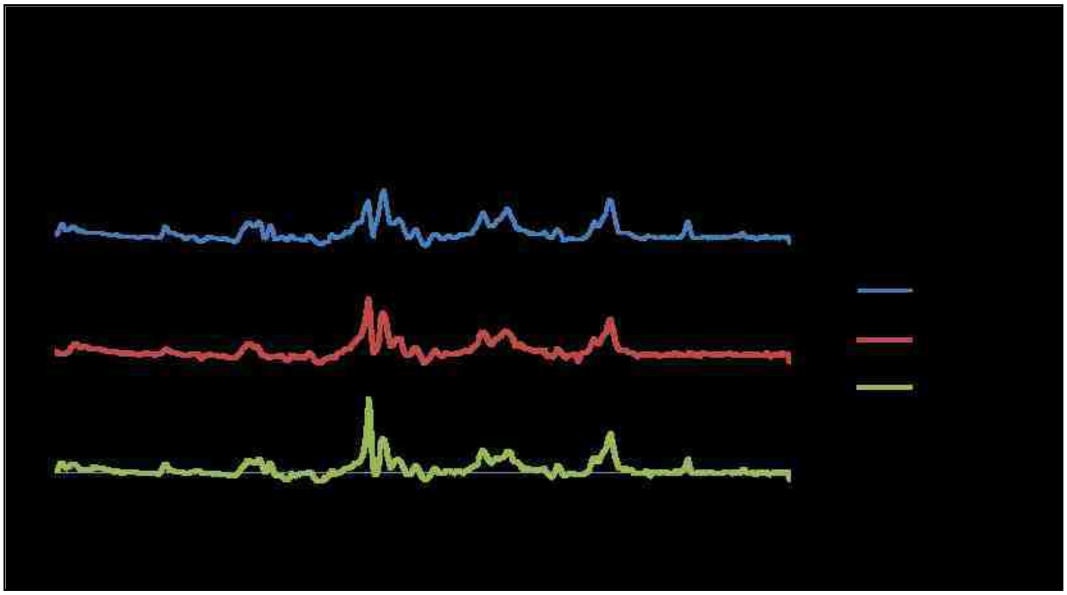


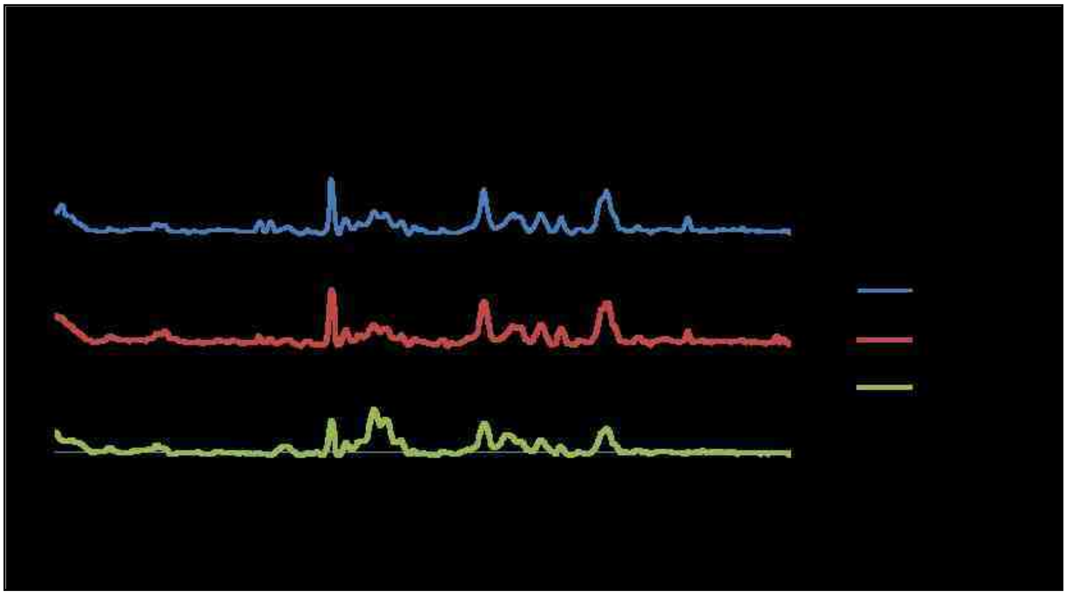


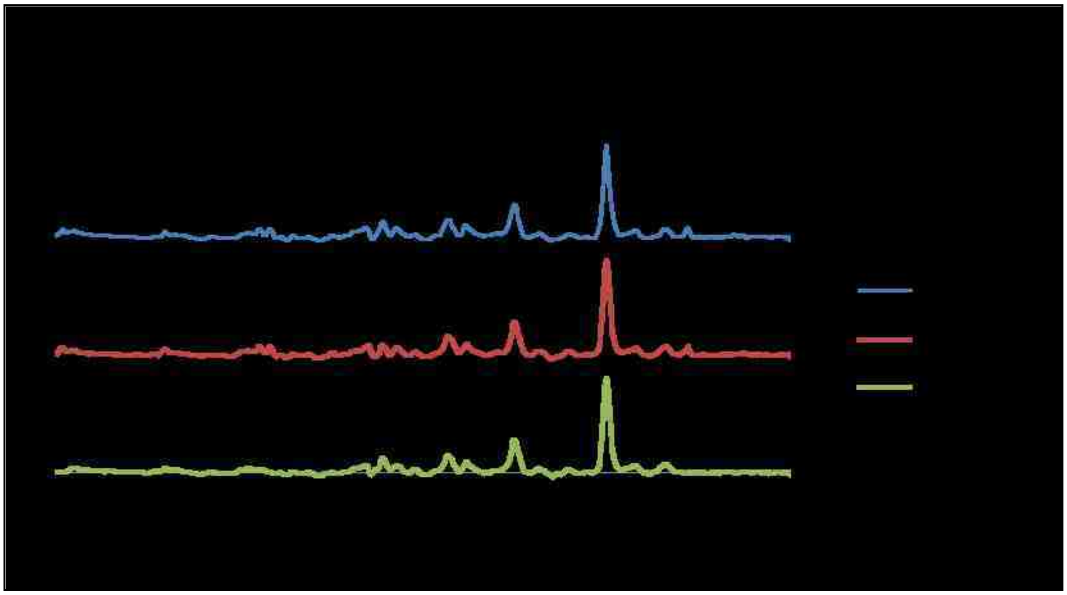


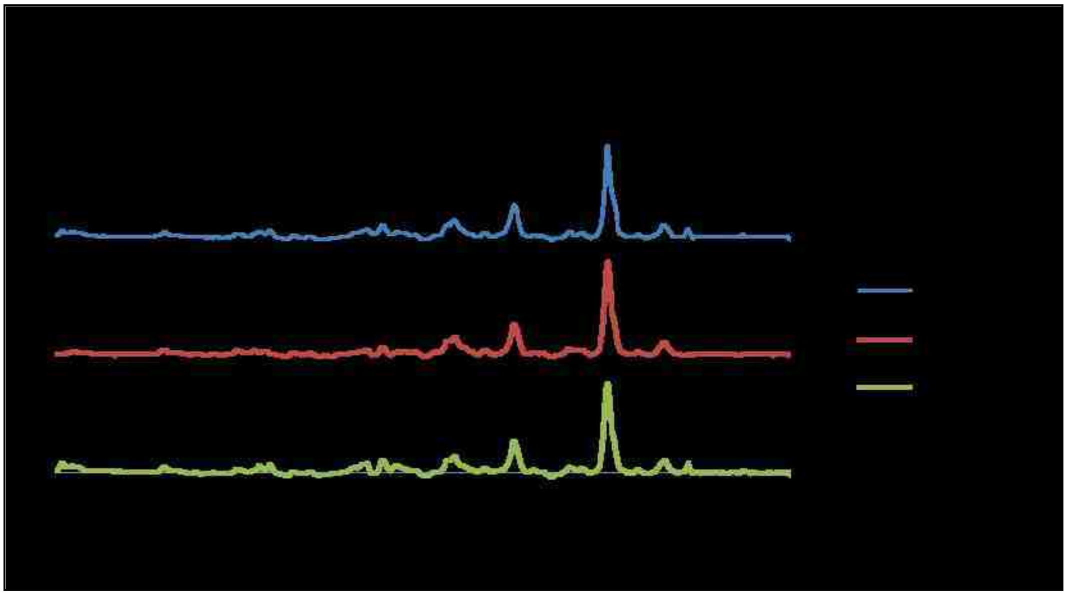


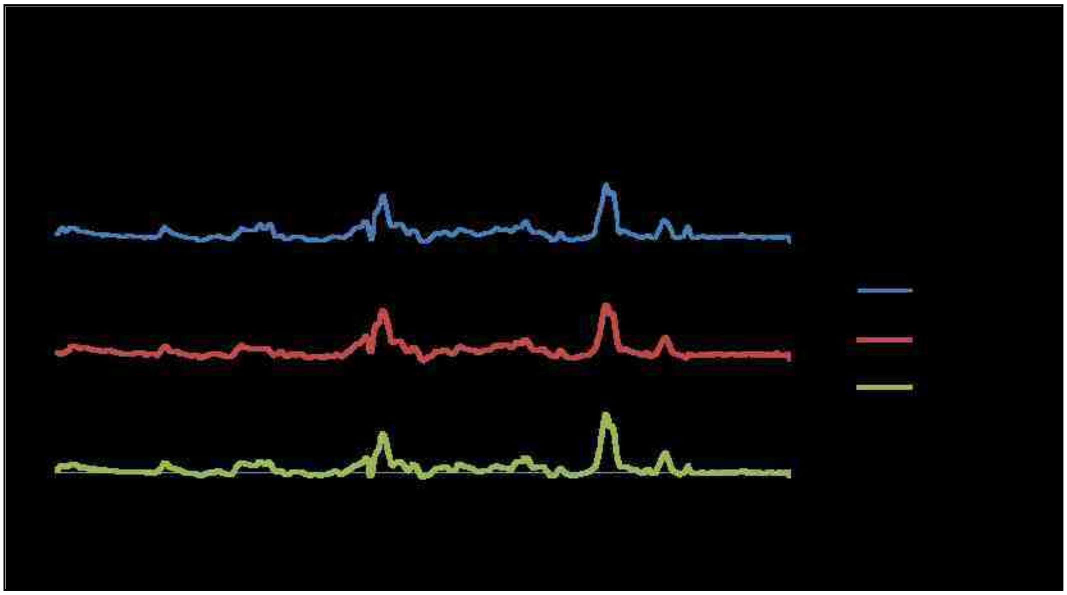


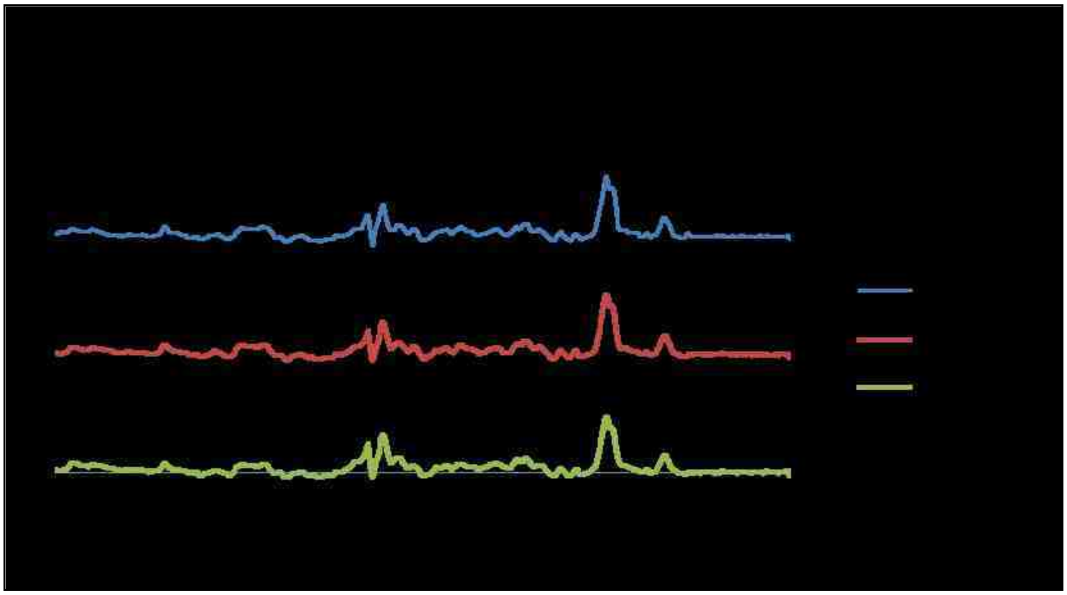


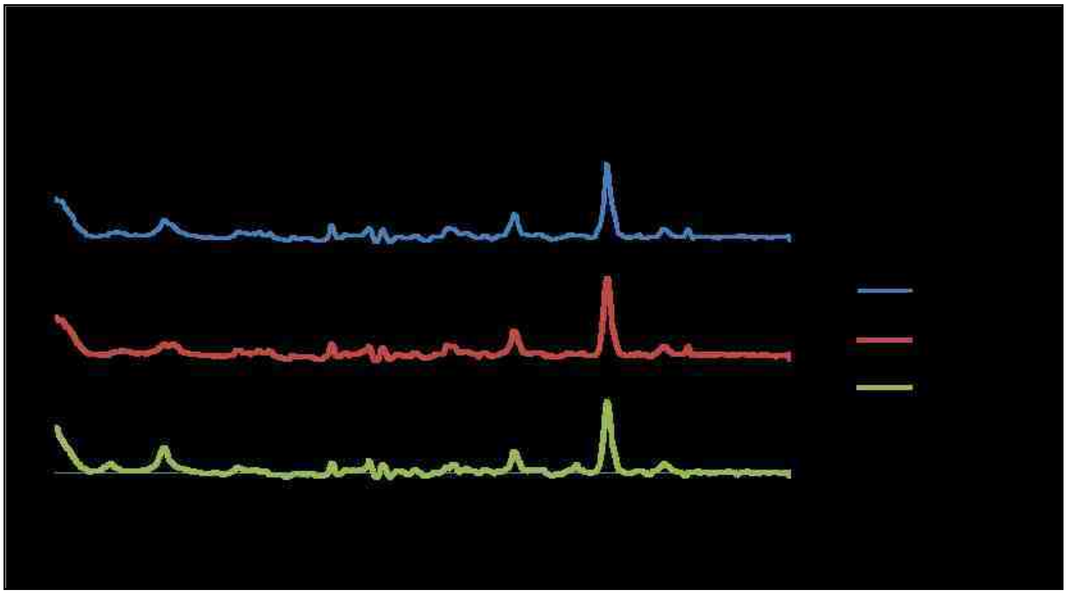


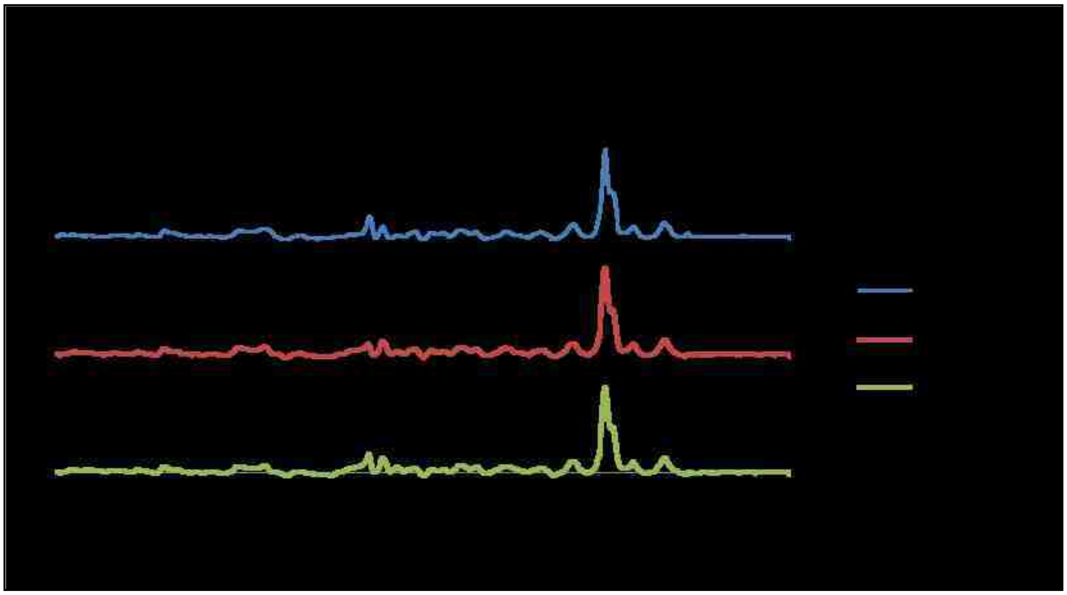


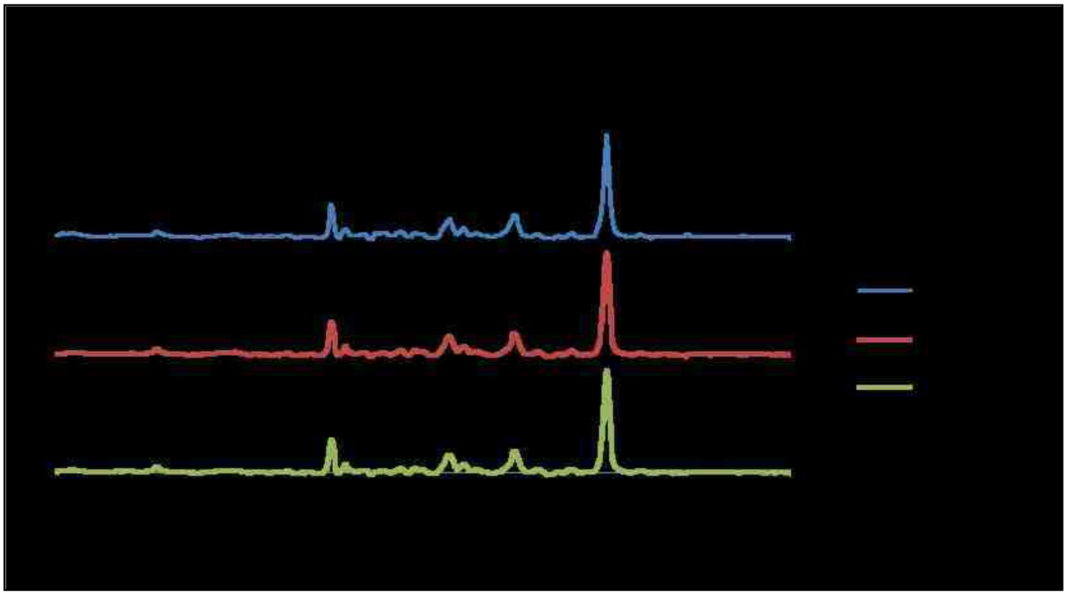


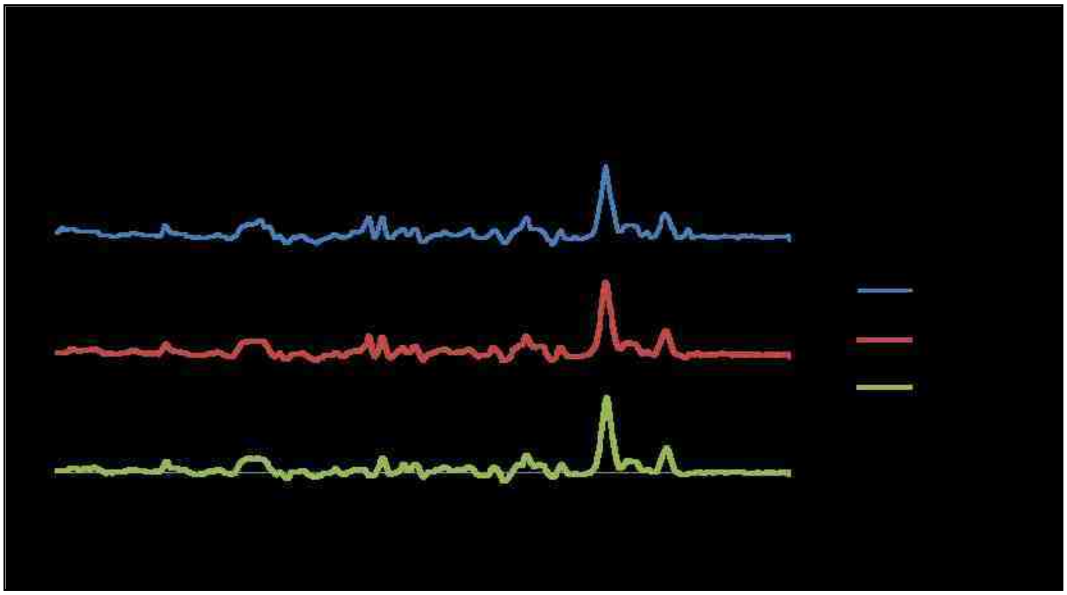


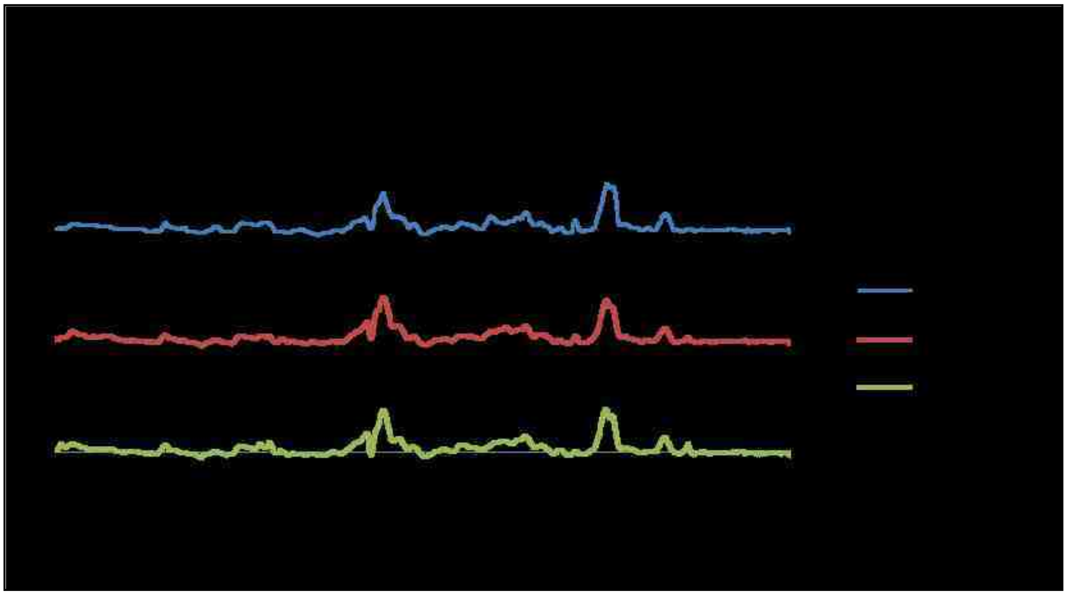


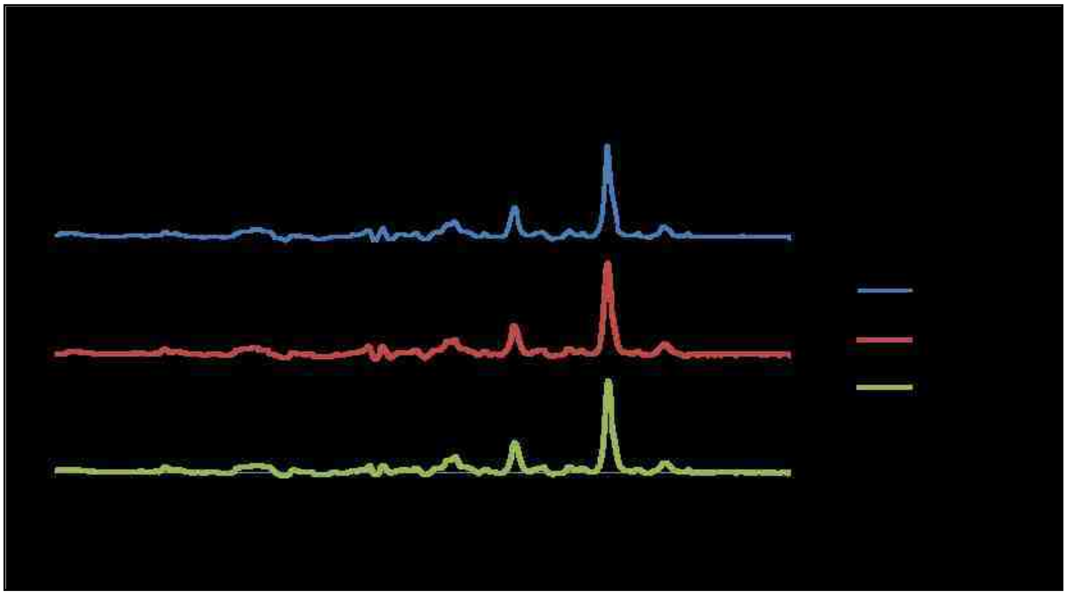


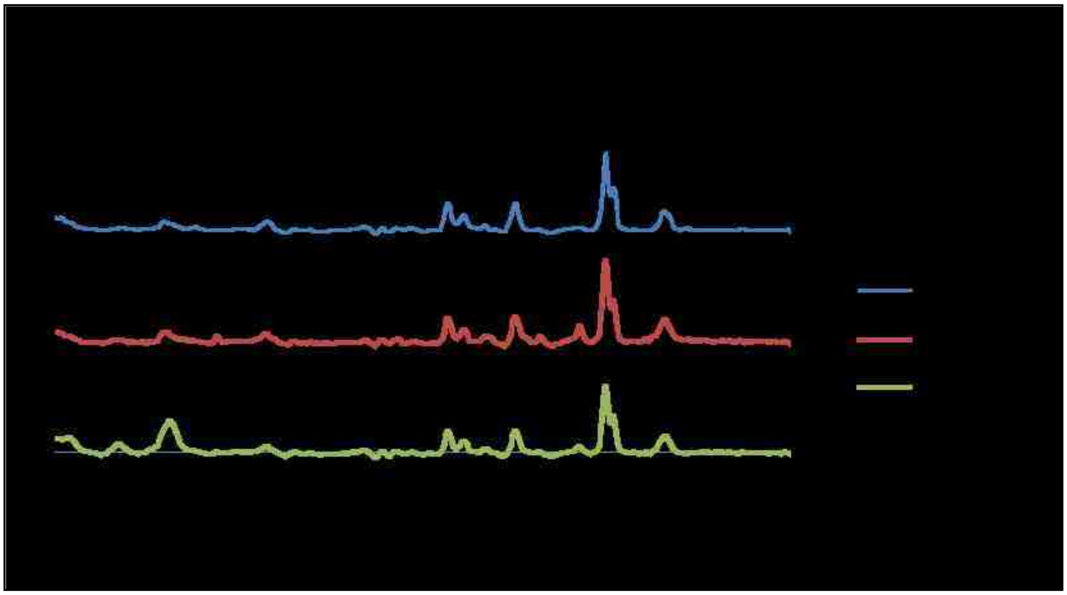


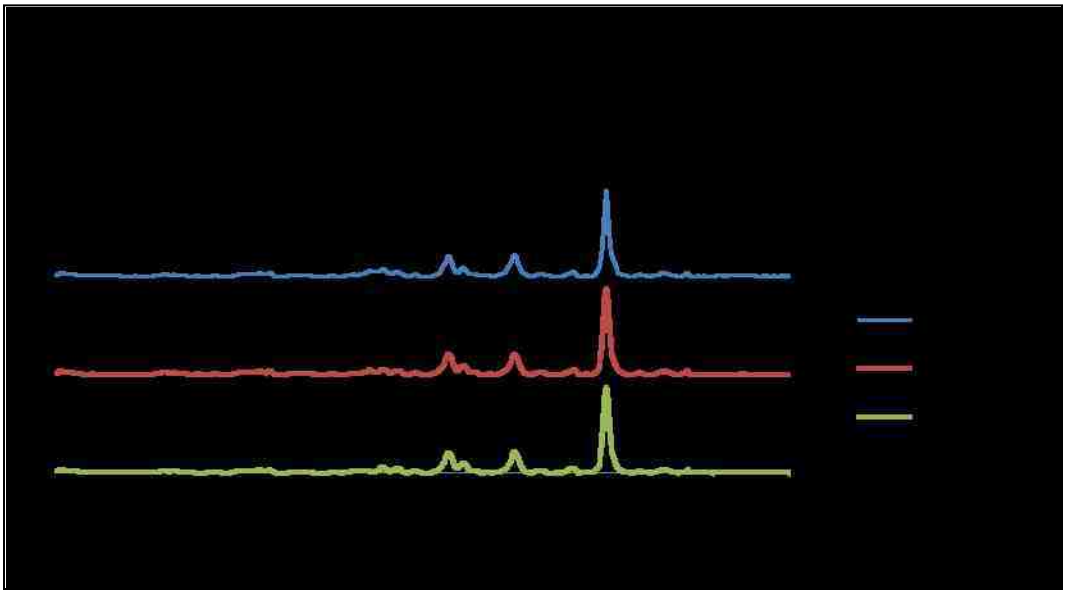


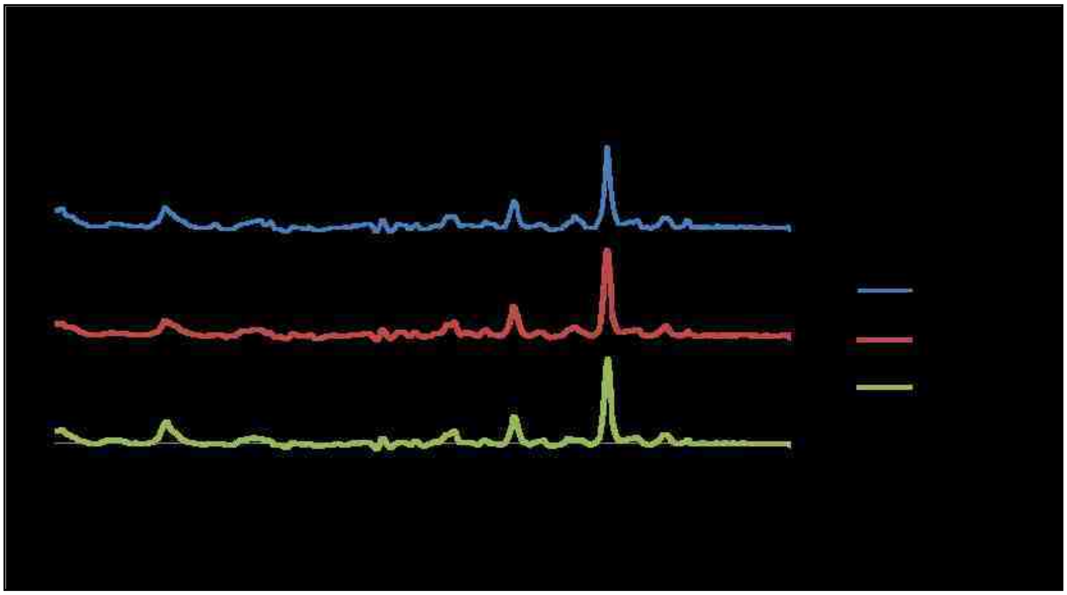




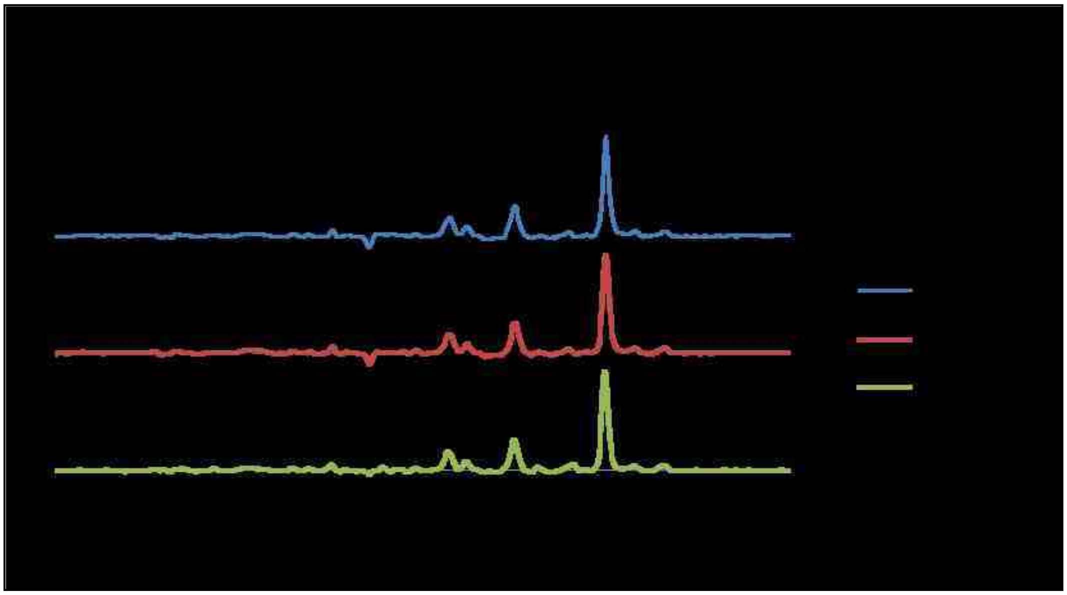


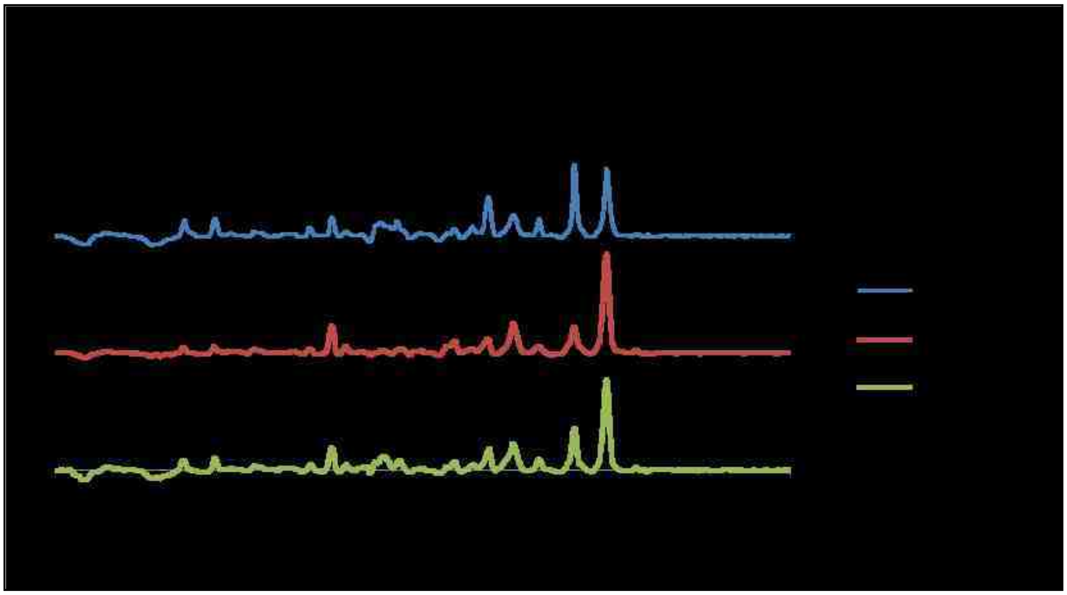


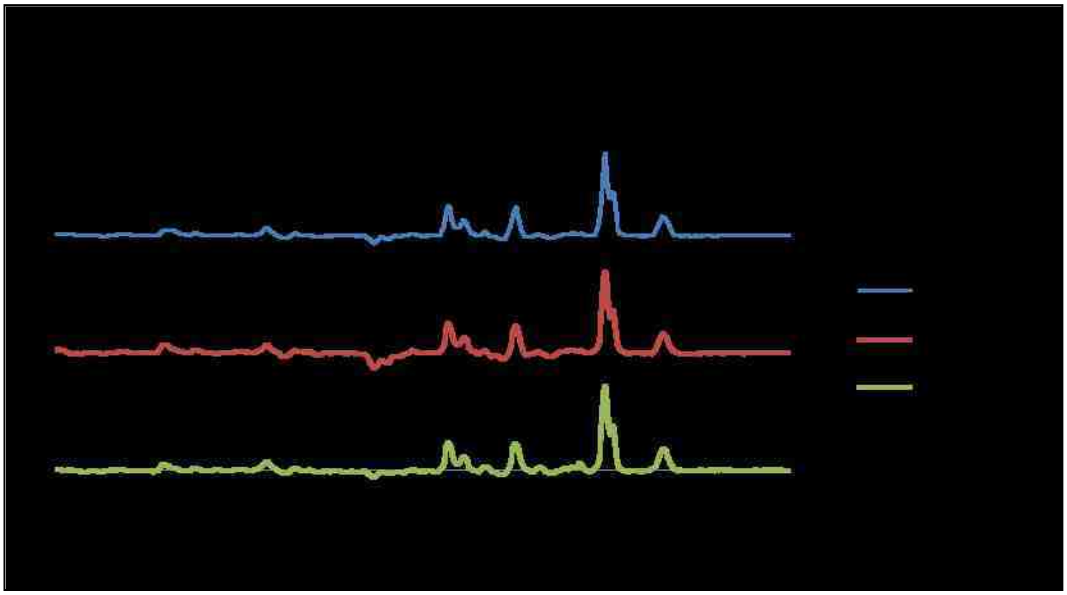


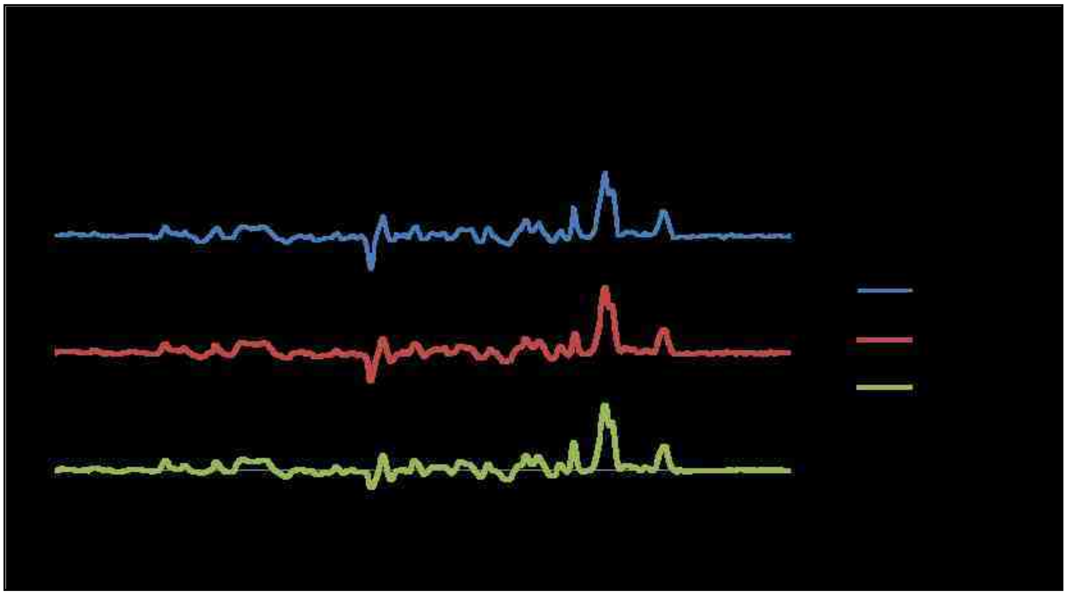


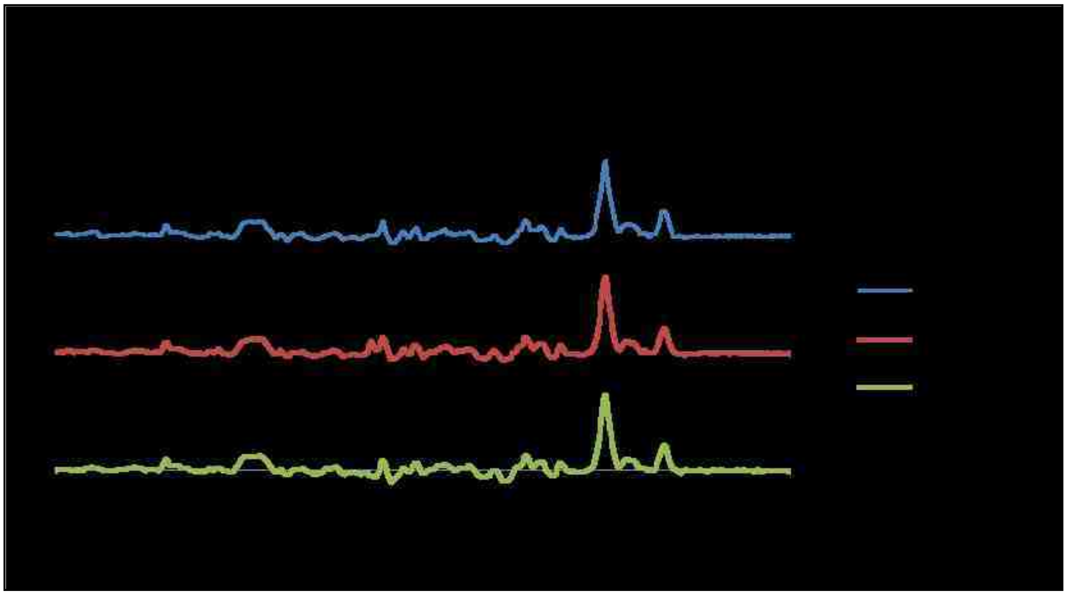
B.2 External Validation

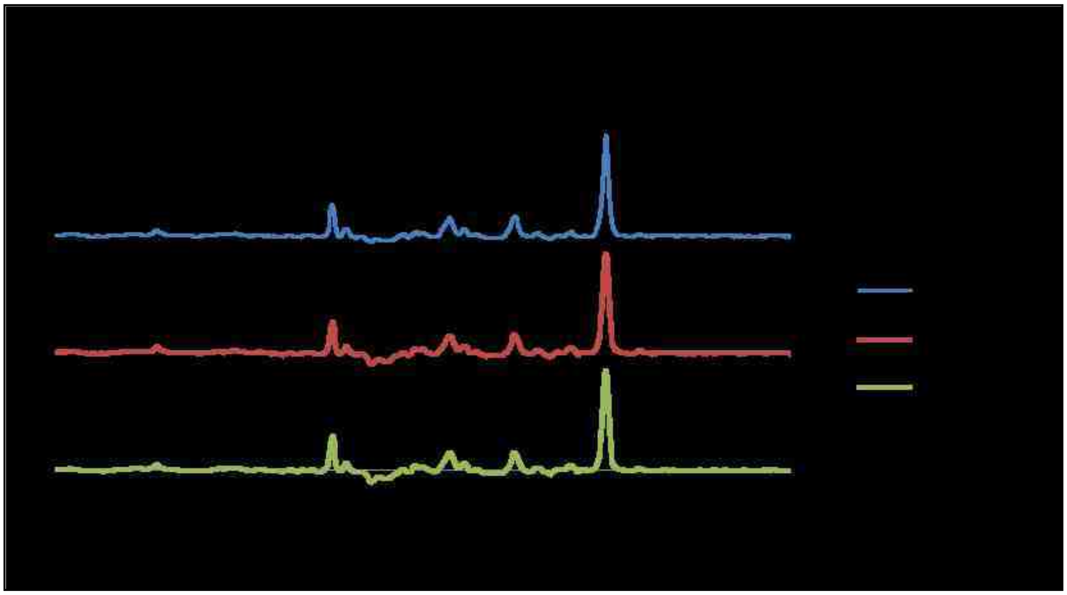


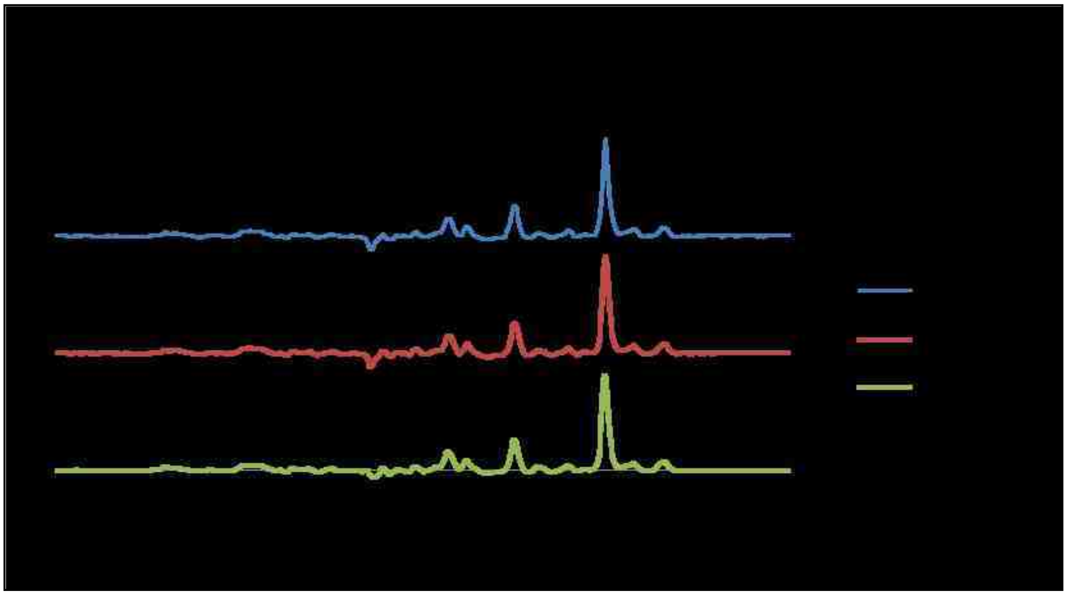


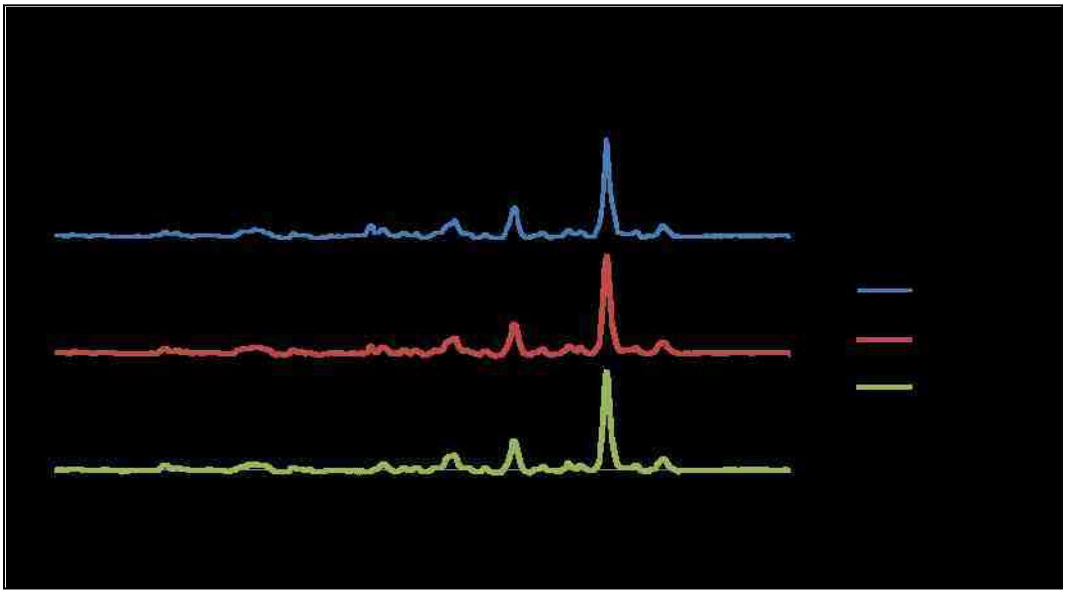


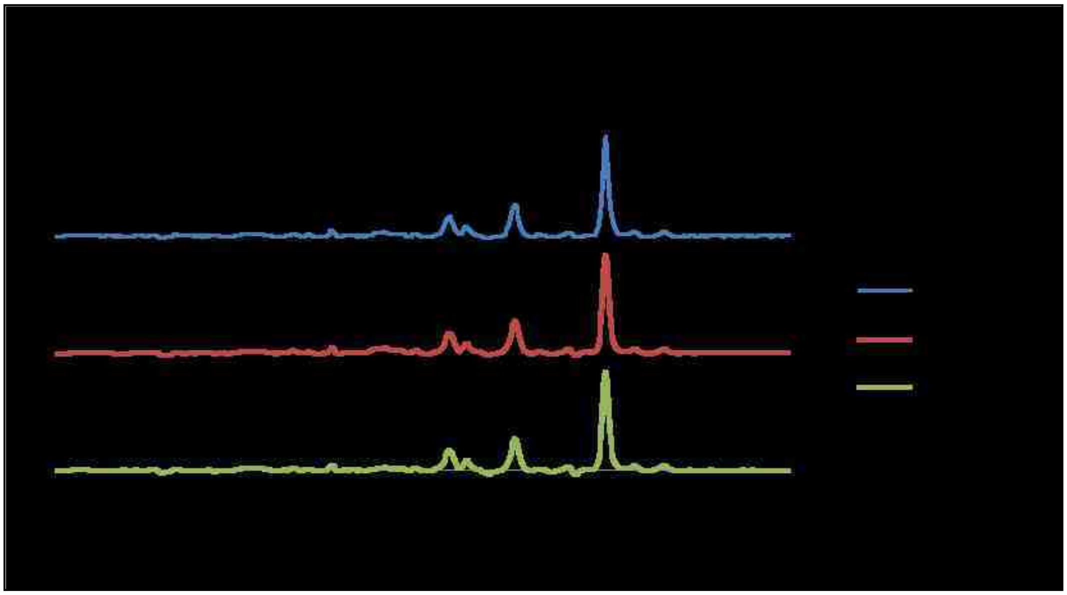


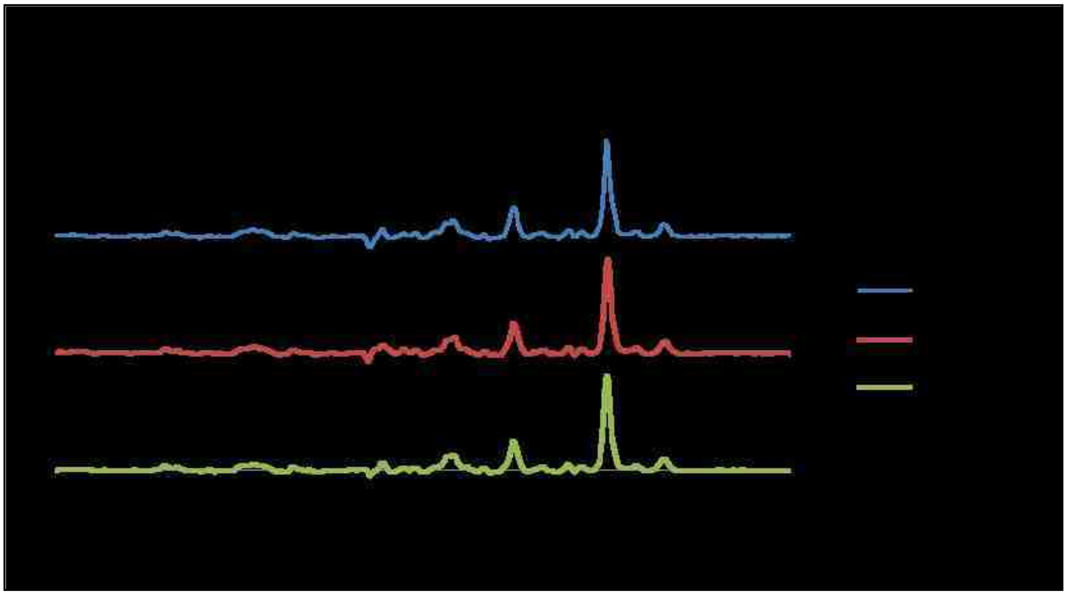


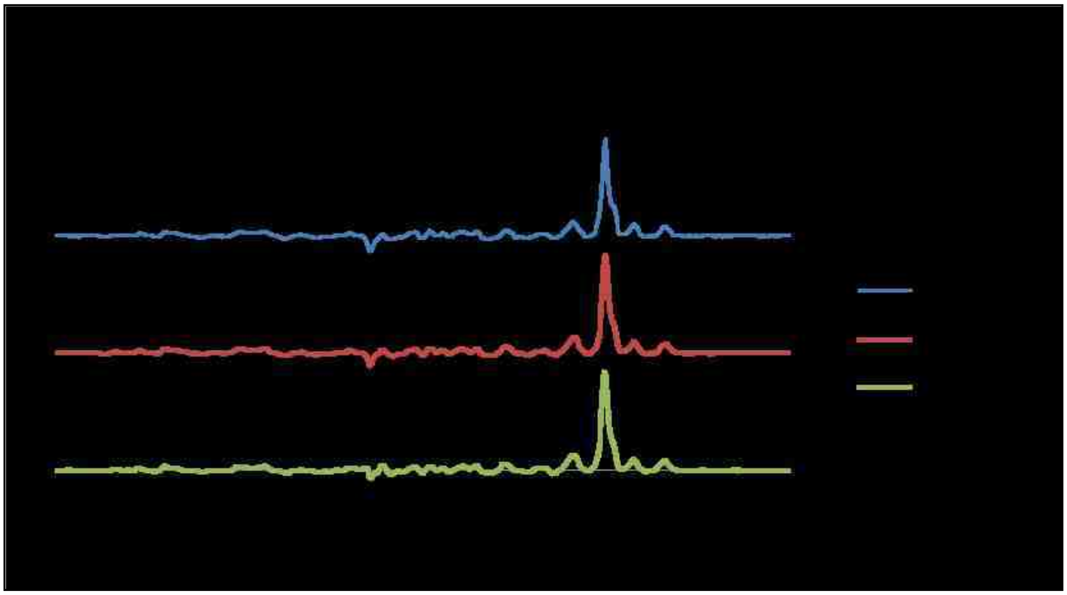


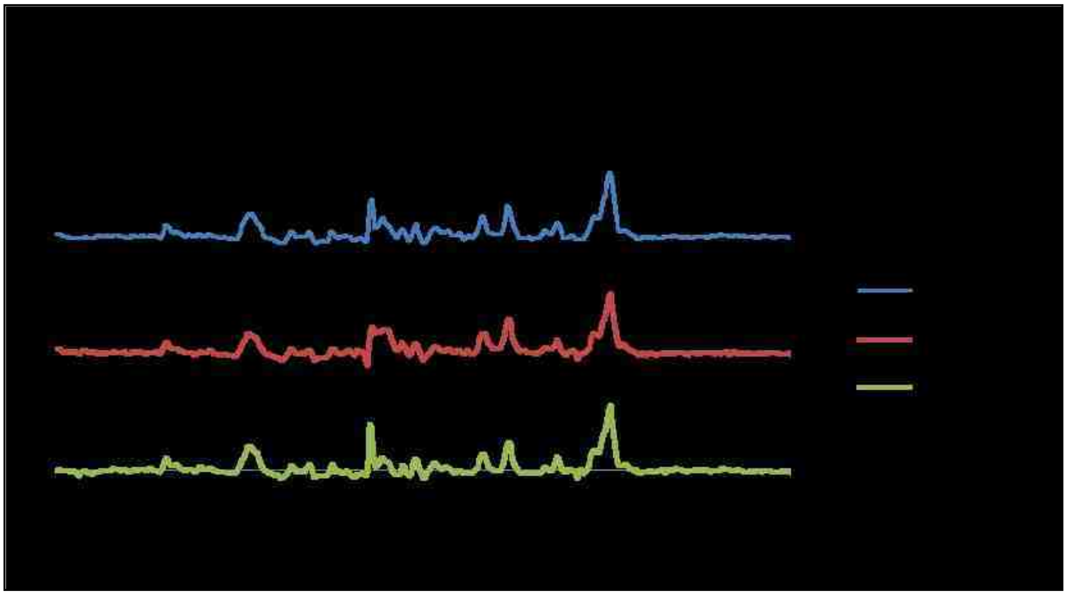


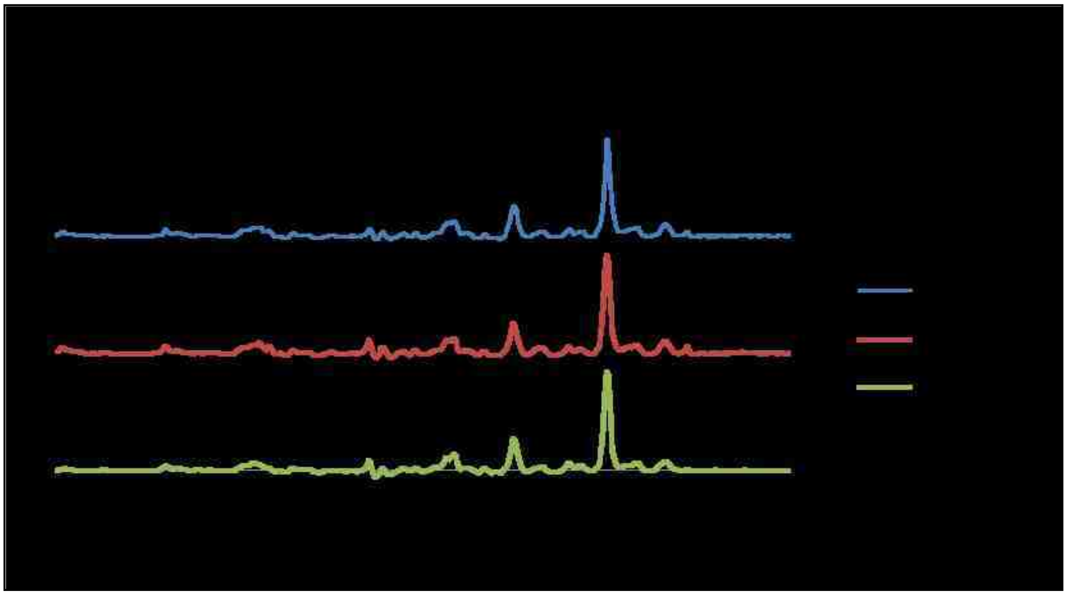


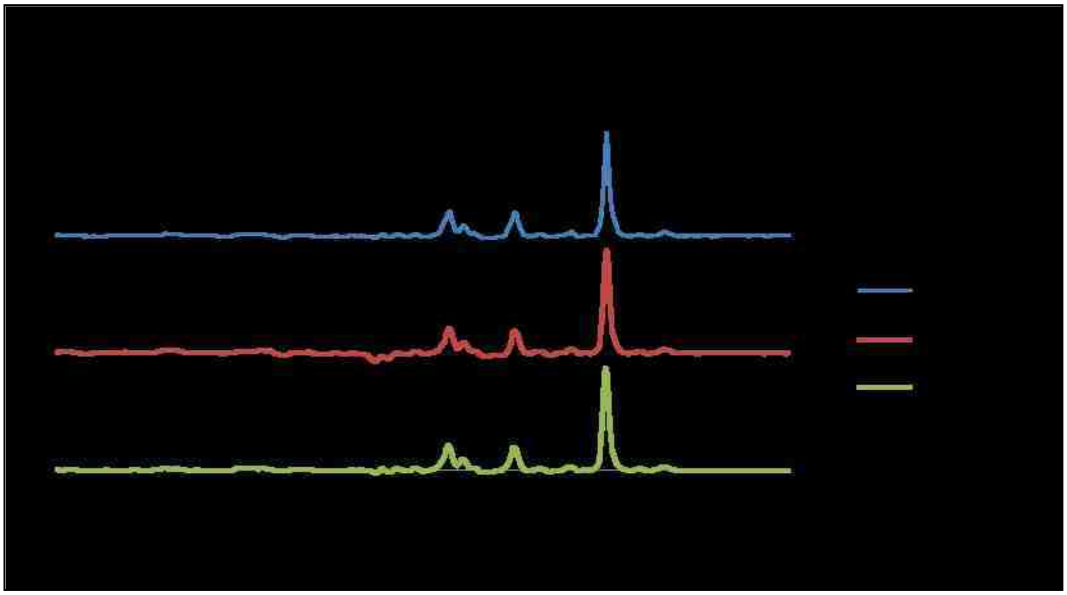


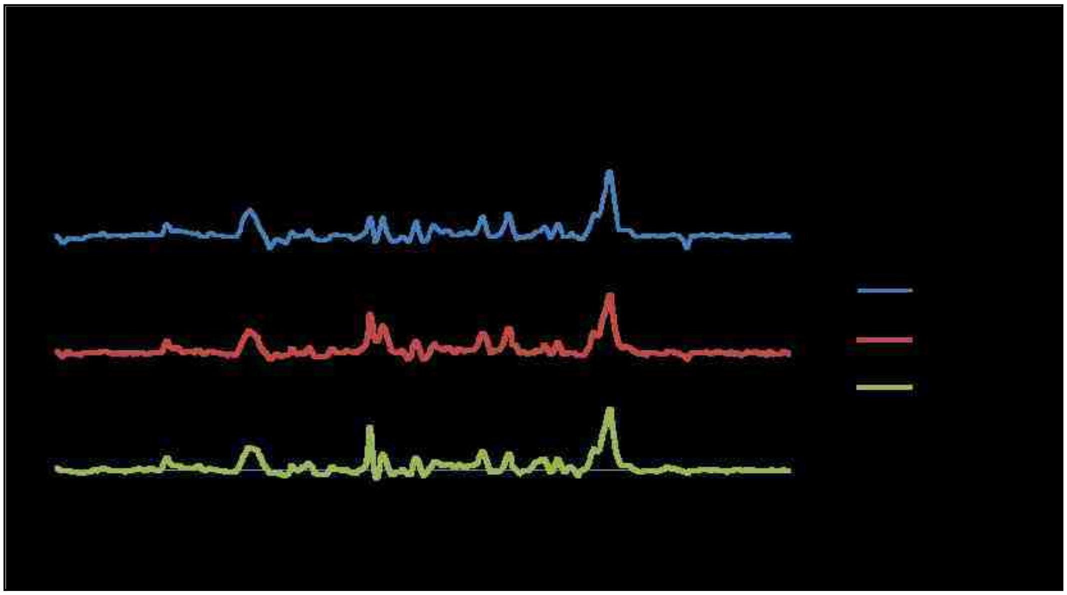


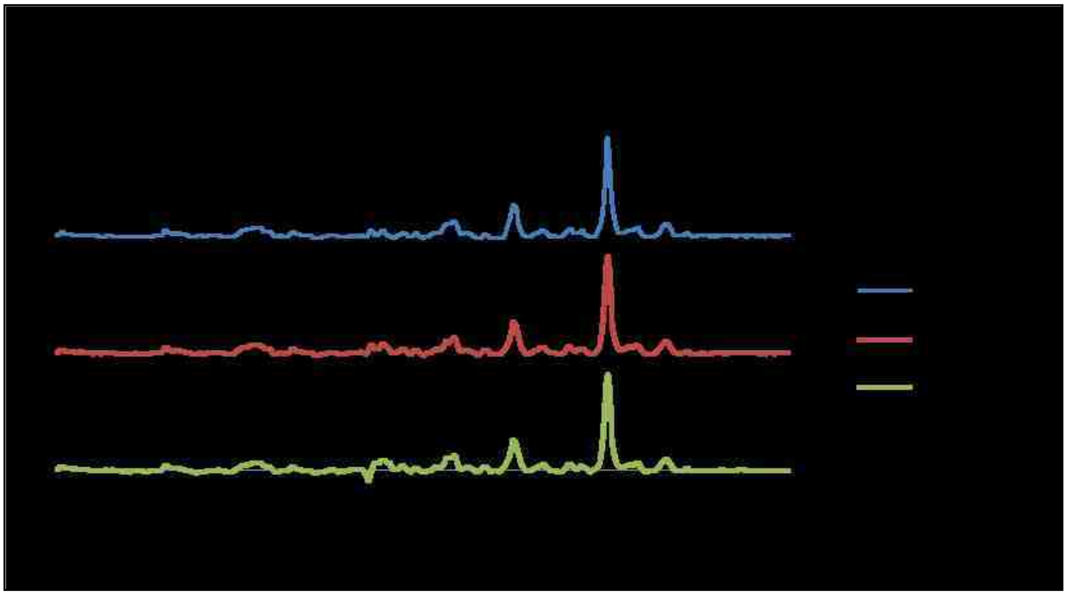


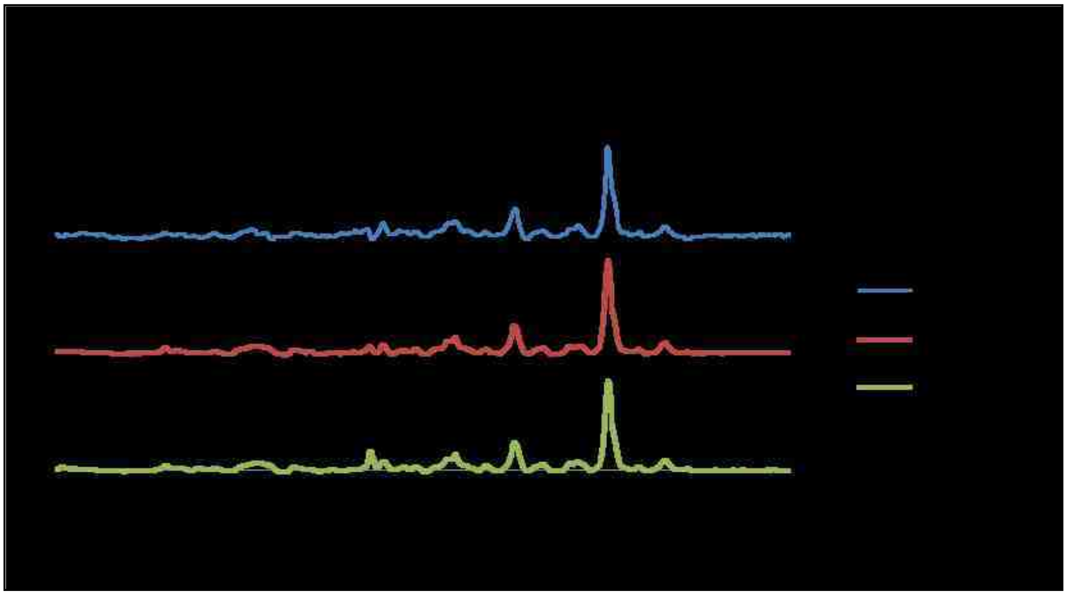


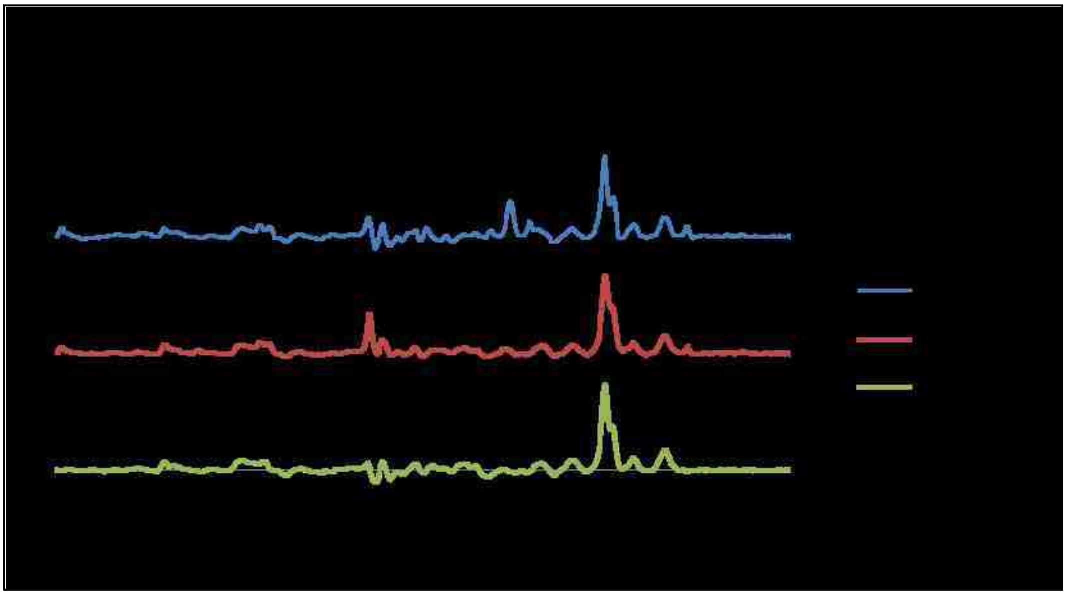


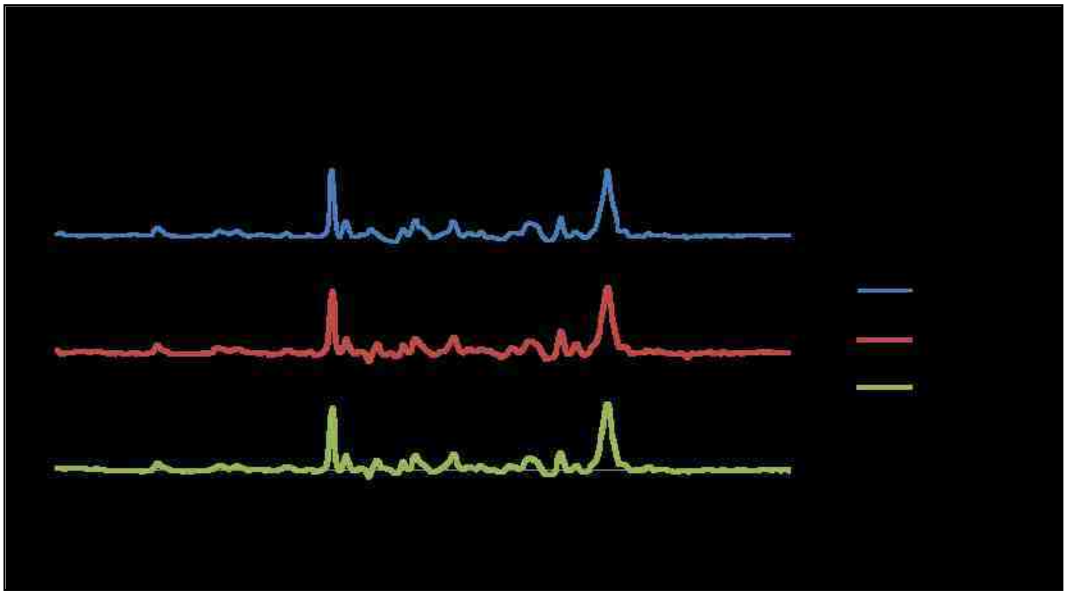


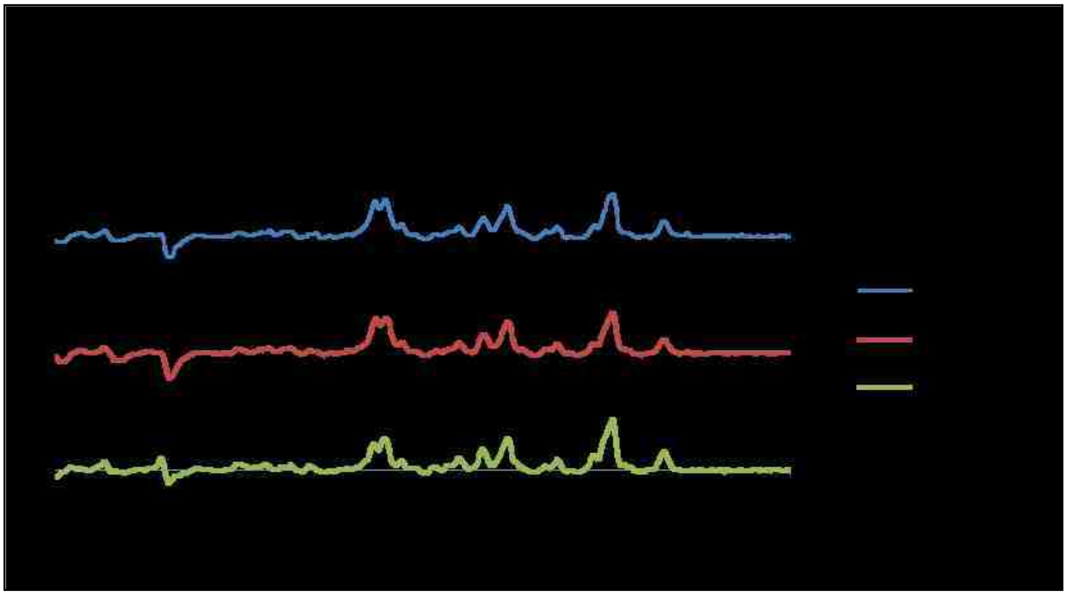


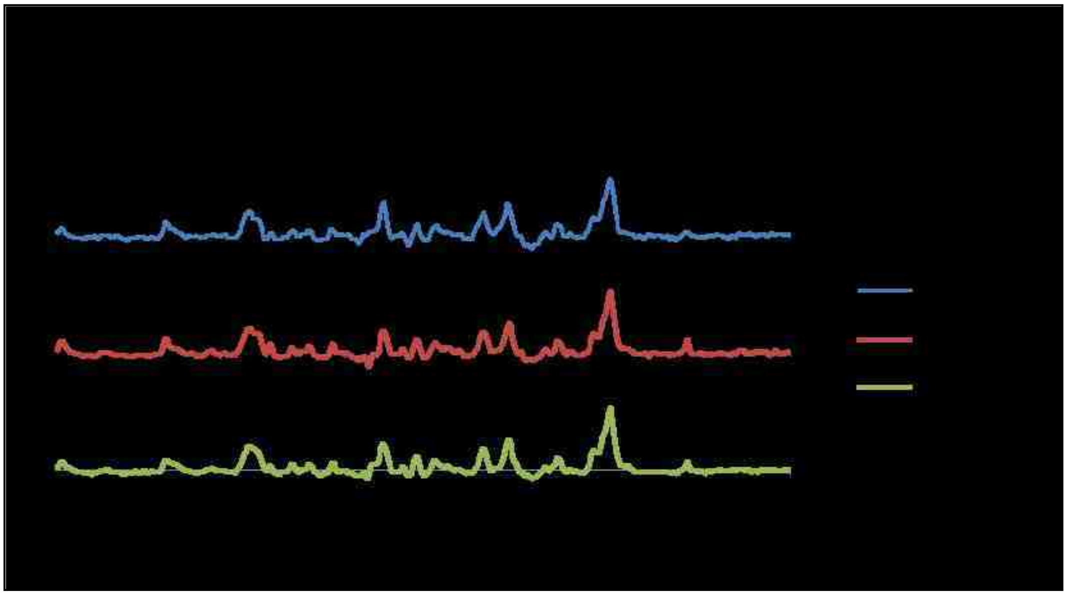


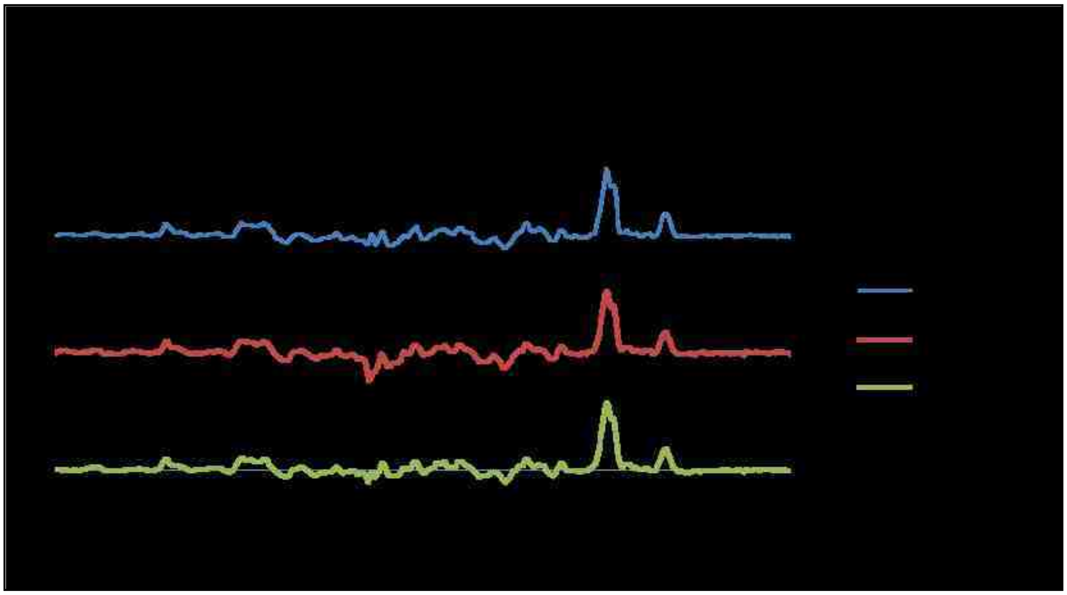


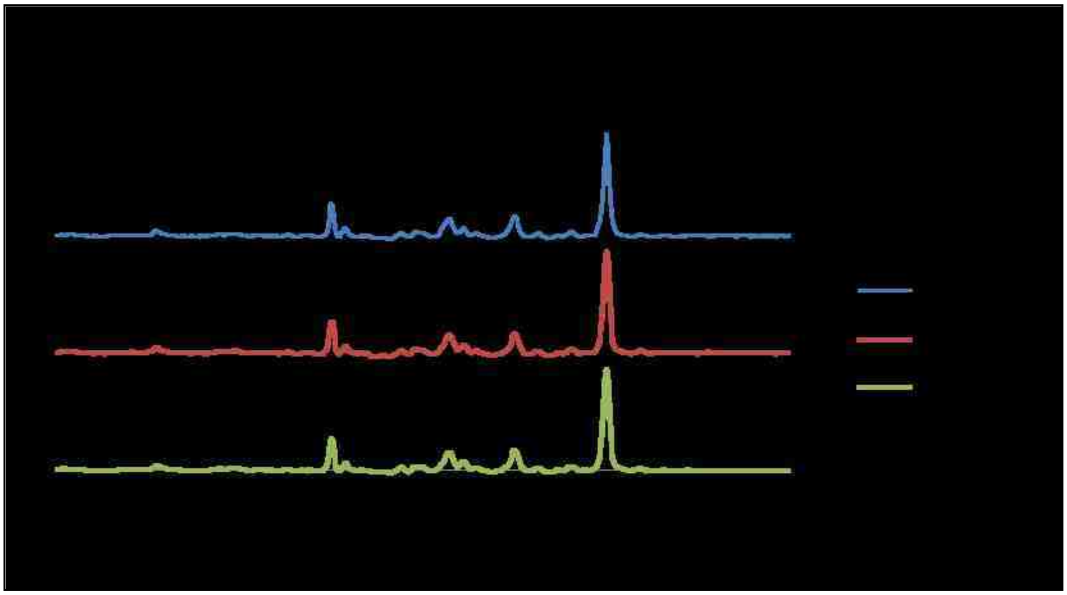


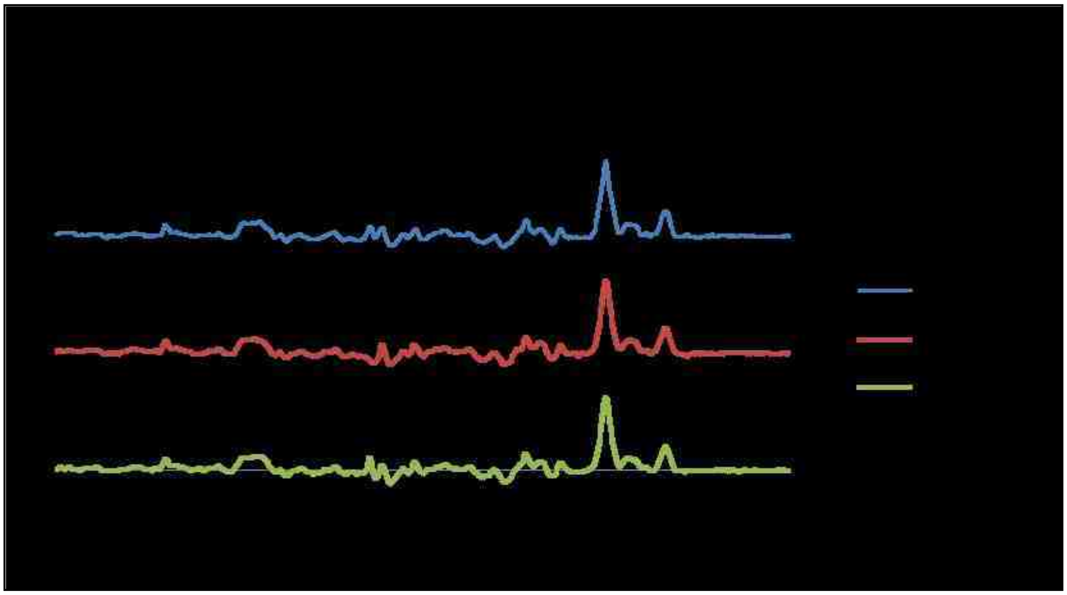


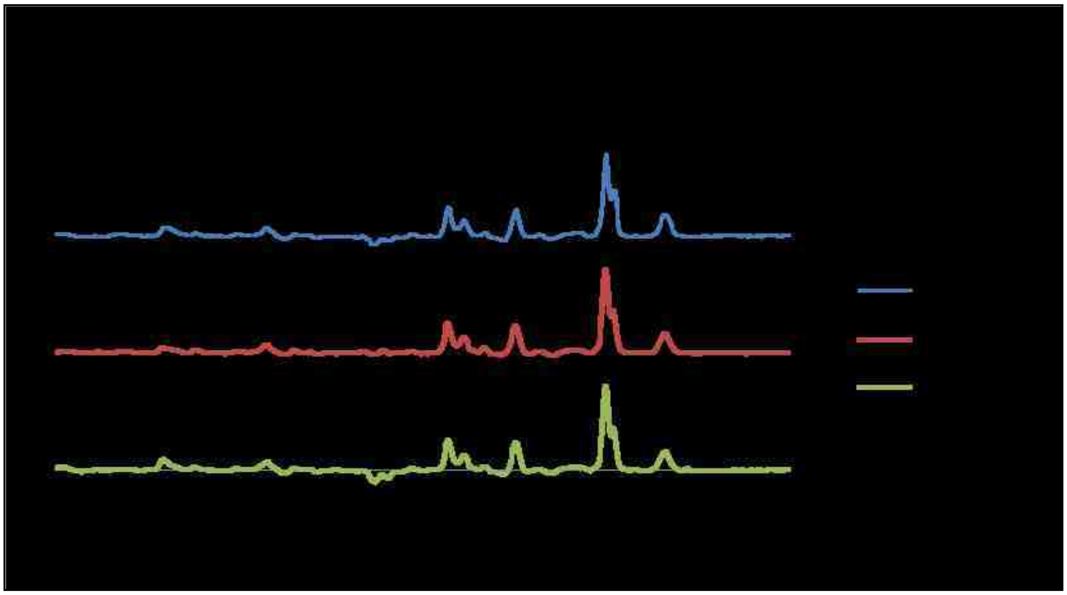












B.3 External Validation Key

External Validation Sample	**Original Sample**
EVY01	4-Sharp
EVY02	7-Hewlett Packard
EVY03	10-Kodak
EVY04	15-Mita
EVY05	21-QMS
EVY06	22-Panasonic
EVY07	30-Xerox
EVY08	33-Tektronix
EVY09	35-Sharp
EVY10	44-Mita
EVY11	46-Okidata
EVY12	48-Ricoh
EVY13	52-Kodak Color Edge 1560
EVY14	53-Konica 7728
EVY15	57-Ricoh Savin SC 106
EVY16	58-Canon CLC 950
EVY17	63-IBM Network Color Printer
EVY18	66-Okidata C711
EVY19	69-Hewlett Packard Hummingbird
EVY20	82-Tektronix Phaser 780
EVY21	84-Ricoh Aficio Color 5000 Series
EVY22	91-Mita Ci 7600
EVY23	94-Panasonic FP C1/FP C1XE
EVY24	95-QMS MagiColor LX
EVY25	98-Kodak Color Edge