RAMAN BIOSENSORS

ALI MOMENPOUR

A thesis submitted in partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Electrical Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

This PhD thesis focuses on improving the limit of detection (LOD) of Raman biosensors by using surface enhanced Raman scattering (SERS) and/or hollow core photonic crystal fibers (HC-PCF), in conjunction with statistical methods. Raman spectroscopy is a multivariate phenomenon that requires statistical analysis to identify the relationship between recorded spectra and the property of interest. The objective of this research is to improve the performance of Raman biosensors using SERS techniques and/or HC-PCF, by applying partial least squares (PLS) regression and principal component analysis (PCA).

I began my research using Raman spectroscopy, PLS analysis and two different validation methods to monitor heparin, an important blood anti-coagulant, in serum at clinical levels. I achieved lower LOD of heparin in serum using the Test Set Validation (TSV) method. The PLS analysis allowed me to distinguish between weak Raman signals of heparin in serum and background noise.

I then focused on using SERS to further improve the LOD of analytes, and accomplished simultaneous detection of GLU-GABA in serum at clinical levels using the SERS and PLS models. This work demonstrated the applicability of using SERS in conjunction with PLS to measure properties of samples in blood serum. I also used SERS with HC-PCF configuration to detect leukemia cells, one of the most recurrent types of pediatric cancers. This was achieved by applying PLS regression and PCA techniques.

Improving LOD was the next objective, and I was able to achieve this by improving the PLS model to decrease errors and remove outliers or unnecessary variables. The results of the final optimized models were evaluated by comparing them with the results of previous models of Heparin and Leukemia cell detection in previous sections.

Finally, as a clinical application of Raman biosensors, I applied the enhanced Raman technique to detect polycystic ovary syndrome (PCOS) disease, and to determine the role of chemerin in this disease. I used SERS in conjunction with PCA to differentiate between PCOS and non-PCOS

patients. I also confirmed the role of chemerin in PCOS disease, measured the level of chemerin, a chemoattractant protein, in PCOS and non-PCOS patients using PLS, and further improved LOD with the PLS regression model, as proposed in previous section.

Acknowledgments

This PhD thesis proposal is the result of my own hard work, and contributions, guidance and encouragement from other researchers.

First and foremost, I would like to thank my advisor, Professor Hanan Anis. She supported me throughout this research, and helped me develop scientific skills and overcome challenges. I am very grateful for her ideas and invaluable guidance that helped me succeed in this endeavour.

I also want to thank Dr. Vidhu S. Tiwari, who shared his valuable experiences with me. His interest, suggestions and encouragement were a great help.

In addition, I would like to pay tribute to my committee members: Dr. Karin Hinzer, Dr. Jean-Michel Ménard, Dr. Costel Flueraru, and Dr. Arkady Major, for their time, questions, insightful comments and feedback.

My sincere thanks also goes out to Louis Tremblay from the Chemical Engineering Department, for allowing me to use the laboratory facilities to prepare samples.

This research would not have been possible without the financial assistance of NSERC, the School of Electrical Engineering and Computer Science at the University of Ottawa and Ottawa Hospital Research Institute. I would like to thank those agencies. My time at the University of Ottawa was very enjoyable, due to the help and support I received from members of the biophotonics lab. I thank them all, particularly Dr. Altaf Khetani, Dr. Majid Naji, Dr. Mohammed Abosafe, Dr. Hussein Kotb, Brett Smith, Shiyu Gao, Huda Yusif and Robert Hunter.

Lastly, I want to thank my family, especially my wife Maryam. Without their spiritual support I would not have completed this work.

Table of Contents

List of Figures

Number Page

List of Tables

Glossary

ACT: Activated clotting time

AgNP: Silver nanoparticles

AML: Acute myeloid leukemia

ANOVA: Analysis of variance

APTT: Activated partial thromboplastin time

BP: Bandpass filter

BVSPLS: Backward variable selection for PLS

CELIF: Capillary electrophoresis-laser-induced fluorescence

CARS: Coherent anti-stokes Raman scattering

CCD: Charge-coupled device

CFS: Correlation-based feature selection

COVPROC: Covariance procedure

CNS: Central nervous system

CSF: Cerebrospinal fluid

CT: Charge transfer

CW: Continuous wavelength

DI: De-ionized

EF: Enhancement factor

ELISA: Enzyme-linked immunosorbent assay

EM: Electromagnetic

FCV: Full cross validation

FF: Follicular fluid

FOM: Figure of merit

FT-IR: Fourier transform infrared

GA: Genetic algorithm

GABA: γ-aminobutyric

GC: Gas chromatography

GC/ECD: Gas chromatography with an electron capture detector

GC/FID: Gas chromatography with a flame ionization detector

GC/MSD: Gas chromatography with a mass spectrometry detector

GC/TED: Gas chromatography with a thermionic emission detector

GC/FT-IR: Gas chromatography with a Fourier transform infrared detector

GLU: Glutamate

HCA: Hierarchical cluster analysis

HC-PCF: Hollow core photonic crystal fiber

HOMO: Highest occupied molecular orbit

HPLC: High performance liquid chromatography

HPLC-ECD: High performance liquid chromatography with electrochemical detection

HPLC-FD: High performance liquid chromatography with fluorescence detection

IPLS: Interval PLS

IPW-PLS: Interactive predictor weighted PLS

IR: Infrared

KMCA: K-means cluster analysis

LC–MS: Liquid chromatography mass spectrometry

LC–MS/MS: Liquid chromatography/tandem mass spectrometry

LOD: Limit of detection

LUMO: Lowest unoccupied molecular orbit

LSPR: Localized surface plasmon resonance

MGITC: Malachite green isothiocyanate

MLR: Multiple linear regression

MRD: Minimal residual diseases

MSC: Multiplicative scatter correction

MVA: Multivariate analysis

NA: Numerical aperture

NIR: Near Infrared

NIPALS: Non-linear iterative partial least squares

NP: Nanoparticle

PD: Parkinson's disease

PBG: Photonic bandgap

PBS: Phosphate buffer solution

PC: Principal components

PCA: Principal component analysis

PCF: Photonic crystal fiber

PCOS: Polycystic ovary syndrome

PLS: Partial least squares

PMT: Photomultiplier Tube

R6G: Rhodamine 6G

RMSE: Root mean square error

RMSEC: Root mean square error of calibration

RMSEP: Root mean square error of prediction

ROA: Raman optical activity

RRS: Resonance Raman spectroscopy

SERS: Surface enhance Raman scattering

SPM: Scanning probe microscopy

SIMPLS: Statistically inspired modification of PLS

SVM: Support vector machines

SwPA-PLS: Sub-window permutation analysis coupled with PLS

TEM: Transmission electron microscopy

TERS: Tip-enhanced Raman spectroscopy

TIR: Total internal reflection

TRS: Transmission Raman spectroscopy

TSV: Test set validation

UFS: Ultra filtered serum

USP: United states pharmacopeia

UV: Ultraviolet

UVE-PLS: Uninformative variable elimination in PLS

VIS: Visible

Symbols

d: Diameter of the nanoparticles

E: Electric field

L_{int}: Effective constant interaction length

n_{air} **:** Refractive index of air

n_{sil} **:** Refractive index of silica

n_{liq} **:** Refractive index of the liquid or gas

n_{core} **:** Refractive index of the HC-PCF core

n_{clad} **:** Refractive index of the HC-PCF cladding channels

P: Dipole moment

α: Linear polarizability

β: 2nd order nonlinear coefficient

γ: 3rd order nonlinear coefficient

λ: Wavelength of the laser light in the vacuum

λ'**:** Wavelength of the shifted bandgap

λ_0: Wavelength of the fiber when empty

NA: Numerical aperture

ϵ: Complex dielectric constant

ϵ_r: Real dielectric constant

ϵ_i: Imaginary dielectric constant

m: Complex refractive index

n: Refractive index

k: Absorption coefficient

χ: Shape factor of nanoparticles

x: Average spectrum

a: Intercept of verage spectrum

b: Slope of verage spectrum

e: Residual of verage spectrum

S_{Dev}: Standard deviation

Y: Dependent variable

X: Independent variable

B: Regression coeffiecient

E: Error

SS_{res}: Residual sum of squares

SS_{tot}: Total sum of squares

R^2: Coefficient of determination

y_{ref}: Reference value for test sample

y_{pred}: Prediction value for test sample

y_{ave}: Average of reference values for test sample

Chapter 1. Introduction

A biosensor is an analytical device, used for the detection of an analyte that combines a biological component with a physicochemical detector.

Optical biosensors are based on the detection of a change in optical phenomena such as absorption, polarization, scattering, surface plasmon resonance, or photo acoustic effects. And light-molecule interaction is the basis of the biosensor sensitivity. Optical biosensors are powerful alternative to conventional analytical techniques, for their particularly high specification, sensitivity, small size, and cost effectiveness. Raman biosensors are based on the detection of Raman scattering due to the vibrational bands of analyte. Detecting the Raman signal of an analyte is a promising method to discriminate between different species. While Raman is considered to be the 'fingerprint of an analyte', the intensity of this signal is weak. Two important ways to increase the intensity of a Raman signal are investigated in this thesis. The first is to use nanoparticles; this is known as Surface Enhanced Raman Scattering (SERS). The second is to use hollow core photonic crystal fiber (HC-PCF) as a sampling container. Although they are both powerful analytical techniques that enable the detection of extremely low concentrations of molecules, the limit of detection is totally dependent on the statistical methods used to extract the Raman signal from the background noise and improve the limit of detection.

SERS, in conjunction with partial least squares (PLS) analysis, has been used by many researchers to improve the limit of detection. PLS is a statistical method for spectral analysis of data that produces a linear regression model to describe the relationship between response variables and predictor variables. The PLS model is based on a dataset collected from different samples of one or more analytes in a solution such as serum. Prior to creating a PLS model, pre-processing, such as baseline correction and multiplicative scatter correction (MSC), is applied to correct the variability of the dataset. Validation of a PLS model, an important aspect of data processing, is done by full cross validation (FCV) or test set validation (TSV), the choice depending on the dataset. PLS models are typically evaluated against various statistical

parameters, including root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and the coefficient of determination (R^2). The outliers, extreme values in the dataset that are not close to other observation points, are eliminated to reduce the deviation and error in the model. As well as outlier elimination, variable selection can also improve PLS results.

Principal component analysis (PCA) is another statistical model used to classify different types of samples. With PCA analysis, the variations in the dataset are used for identification and interpretation. The projection of an X-variable (spectral wavelength) and a Y-variable (analytical data) to a new space of principal components (PCs) is the basis of PCA. The first PC is defined as the direction of the most variance in the Y-variables, and the second PC, which is orthogonal to the first PC, is defined as the direction of the second-most variance that was not described by the first PC, and so on. The scores show the similarities or differences among the samples, and similar samples with the same PC have close scores. The plotting of one PC against another could be used to interpret the structure of observation, and reveal the hidden structure of spectra.

1.1 Novelty and contribution

The motivation of this research is to develop Raman biosensors that have an improved limit of detection (LOD). This is done using surface enhanced Raman and/or hollow-core photonic crystal fibers in conjunction with statistical methods.

In this thesis, I start by showing that the weak Raman signal can be used to detect the clinical level of heparin in serum if I use PLS analysis. In this study, two validation methods, FCV and TSV, were implemented and the LOD of heparin was calculated to show that PLS analysis can extract the Raman signal from the background noise.

I then move to using SERS to further improve the LOD and apply that to the detection of GLU-GABA in serum at clinical level. This work explores the feasibility of employing surface-enhanced Raman scattering spectroscopy, in conjunction with PLS, for simultaneous

measurement of physiological concentrations of GLU and GABA in their aqueous solution and blood serum.

I was able to achieve further improvement in detection sensitivity by using HC-PCF as a sampling device in conjunction with nanoparticles and statistical analysis. This allowed me to detect leukemia cells using PLS and discriminate between normal and cancer cells using PCA statistical methods.

In order to obtain a lower limit of detection, I then focused on improving the PLS model to obtain a lower RMSEP. Using the TSV method of calibration and removing the outliers enabled me to reach a lower LOD. I have focused on optimizing the PLS regression model by removing unnecessary variables to improve the LOD of Raman biosensors.

Finally, for a clinical application I used the SERS technique in conjunction with PCA to detect polycystic ovary syndrome (PCOS) in patient samples. I also investigated the role of chemerin in PCOS patients, and measured its level in patient samples. The LOD decreased when using the improved PLS technique proposed in this thesis.

This research has generated the following journal papers, book chapter and conference proceedings;

- Journal papers
    - **A. Momenpour** and H. Anis: "An improved partial least-squares regression method for Raman spectroscopy", Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 185, 98–103 (2017).
    - **A. Momenpour**, V. S. Tiwari, V. L. Trudeau, and H. Anis: "Surface enhanced Raman scattering spectroscopy for the detection of glutamate and γ-aminobutyric acid in serum by partial least squares analysis", IEEE Photonics Journal, Vol. 7, No. 3, (2015).

- A. Khetani, **A. Momenpour**, E. I. Alarcon, and H. Anis: "Hollow core photonic crystal fiber for monitoring leukemia cells using surface enhanced Raman scattering (SERS)" Biomedical Optics Express, Vol. 23, No. 22, (2015).

- **A. Momenpour**, V. S. Tiwari, M. M. Tripathi, and H. Anis: “Raman spectroscopy for clinical-level detection of heparin in serum by partial least-squares analysis”, Journal of Biomedical Optics, 18(2), 027010 (2013).

- A. Khetani, J. Riordon, V. S. Tiwari, **A. Momenpour**, Michel Godin, H. Anis: “Hollow core photonic crystal fiber as a reusable Raman biosensor”, Optics Express, 21, 10, 12340 (2013).

- V. S. Tiwari, A. Khetani, **A. Momenpour**, and H. Anis: “Optimum size and volume of nanoparticles within hollow core photonic crystal fiber”, IEEE Journal of Selected Topics in Quantum Electronics, 20, 3 (2013).

- **A. Momenpour**, P. D. A. Lima, Y. Chen, C. Tzeng, B. K. Tsang, and H. Anis: “Use of surface-enhanced Raman scattering to detect polycystic ovary syndrome” (to be submitted).

- Book chapter

  - A. Khetani, **A. Momenpour,** V. Tiwari, and H. Anis: “Silver nanoparticles in SERS based (micro) devices” in “Silver nanoparticles from Surface resonance to biomedical applications”, edited by K. Udekwu, M. Griffith and E. Alarcon, published by Springer (2015).

- Conference proceedings

  - A. Khetani**, A. Momenpour**, J. Riordon, V. S. Tiwari, M. Godin, and H. Anis: “Hollow core photonic crystal fiber as a robust Raman biosensor”, SPIE Photonics West 8576-14 (2013).

- V. S. Tiwari, A. Khetani, **A. Momenpour**, B. Smith, V. Trudeau, and H. Anis: "Detection of amino acid neurotransmitters by surface enhanced Raman scattering and hollow-core photonic crystal fiber", SPIE Photonics West 8233 (2012).

- A. Khetani, V. S. Tiwari, **A. Momenpour**, and H. Anis "Monitoring of adenosine within hollow core photonic crystal fiber by surface enhanced Raman scattering (SERS)", IEEE Nano conference (2011).

1.2 Outline of the thesis

In Chapter 2, the background of Raman, SERS, and MVA is explained.

In Chapter 3, Raman and SERS, in conjunction with PLS analysis, using a cuvette as a sample holder is discussed. In the first part of the chapter, using Raman as an alternative method to measure heparin concentration in serum at a clinical level is studied, and a simple procedure to monitor the heparin concentration in serum by measuring the Raman spectrum of heparin-serum mixture in cuvette is discussed. In this part, we showed that PLS and PCA analytical methods can help to extract the weak Raman spectrum of heparin-serum mixture. In the second part of Chapter 3, the feasibility of employing SERS spectroscopy, in conjunction with PLS, is explored for simultaneous measurement of physiological concentrations of Glutamate (GLU) and γ-aminobutyric acid (GABA) in aqueous solution, and blood serum in cuvette. We established that SERS and PLS can be used to simultaneously monitor the concentration of GLU and GABA in serum.

In Chapter 4, we show that HC-PCF enhances light-matter interaction via photonic band gap property. The SERS and HC-PCF based sensing platform enables enhancement of the Raman signal of important molecules, and has ability to clinically diagnose at the physiological level. Using an integrating pressure-driven flow with HC-PCF improves the stability, the filling speed and the reusability of HC-PCFs. As a clinical application the HC-PCF and SERS platform with PLS and PCA is used to detect and discriminate leukemia cells.

Chapter 5, this chapter shows how the LOD could be improved by decreasing the RMSEP. The selection of the range of variables improves the PLS model, and we show how this can enhance the regression model and reduce the LOD. The efficiency of proposed method is verified using datasets of heparin-serum mixture (Chapter3) and leukemia cells (Chapter4). We found that improved regression model can decrease the LOD of heparin and leukemia cells.

Chapter 6, this chapter discusses using SERS in conjunction with PCA to detect PCOS in patient samples. The role of chemerin, a chemoattractant protein, in PCOS patients is reviewed, and the levels of chemerin in phosphate-buffered saline (PBS) and follicular fluid (FF) samples are measured using the PLS technique. The proposed PLS method in Chapter 5 is then evaluated by measuring the chemerin level of PCOS patients. We established that the SERS platform with PCA is a valuable alternative method to distinguish PCOS patients from non-PCOS patients.

Chapter 2. Background

2.1 Raman scattering

Raman scattering is an inelastic phenomenon that occurs when a photon interacts with a molecule, as first demonstrated by Raman and Krishnan in 1928 [1]. Most of the photons emitted after scattering have the same frequency as the incident photons (Rayleigh scattering), while a relatively small number (approximately 1 in 10^6 to 10^{10} photons [2-3]) have shifted frequencies. This decrease or increase in frequency is known as stokes or anti-stokes Raman scattering, and provides information about vibrational transition in molecules. This specification of the Raman effect is an effective analytical technique to obtain optical 'fingerprints' of molecules. Raman spectra usually have numerous sharp peaks that correspond to specific molecular vibrational frequencies, and these can provide a clear signature defining the presence of specific molecules in a sample. Accordingly, Raman spectra can be used to qualitatively and quantitatively discriminate between chemical species in materials [4]. Despite these capabilities, a key limitation of the Raman effect is its extremely weak signal due to the small number of photons that are Raman scattered. This means that recording an intense Raman spectrum requires high power lasers and long acquisition time, which can lead to damaged samples and limited clinical application. This shortcoming can be addressed by SERS [5], which is explained later.

A schematic diagram of a Raman setup is shown in Figure 2-1. In the setup, the laser beam of a 785 nm continuous wavelength (CW) multimode diode laser (B&W Tek Inc.) with a maximum output power of 450 mW, is collimated by a plano-convex lens (L1), and passed through a bandpass filter (BP) to remove other wavelength components. It is then directed through a dichroic filter (DM), which reflects the laser light at an angle of 45 degrees. The dichroic filter also acts as a reflector for the laser beam, which is further focused onto the sample by a microscopic objective lens (L2). As well, the dichroic filter acts as a long-pass filter for the light scattered backward from the sample, thus allowing only the stokes Raman wavelengths through. The filtered Raman light is then imaged onto a fiber bundle (including 26 multimode fibers with core diameter of 100 micron) by another microscopic objective lens (L3), and the

output of collection fiber (CF) is transformed into a Kaiser f⁄1.8i spectrograph (SP) using a thermoelectrically cooled charge-coupled device camera. This spectrograph has a low frequency stokes grating with spectral range 34 to 1894 cm^{-1} with a resolution of ~2.05 cm^{-1}/pixel. Finally, the spectrum is monitored on a data acquisition computer, and the spectral information can determine the composition of the illuminated sample.

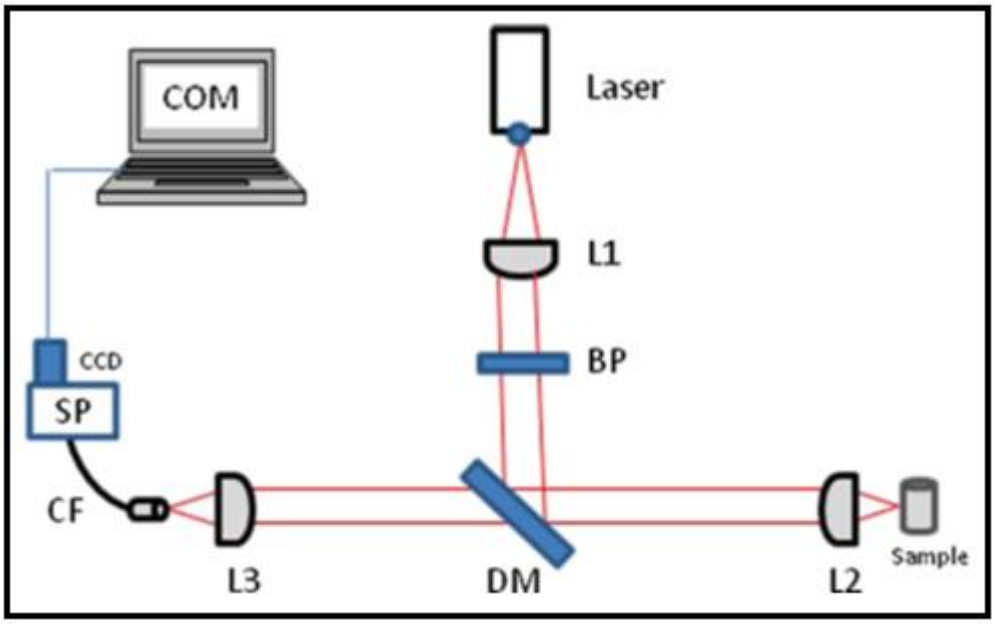

Figure 2-1 Schematic of experimental Raman setup

2.1.1 Raman instrumentation

The main components of any Raman setup are a laser, a sample illumination and light collection system, a spectrometer and a detector. The geometrical configuration can be forward or backward axial, or at 90° to the Raman collection. A schematic diagram of our backward axial configuration is shown in Figure 2-1. The sample is usually illuminated by a UV, VIS, or NIR CW laser. The fluorescence background problem in Raman spectroscopy, particularly in complex organic molecules, can be rectified with using ultrafast pulse laser, or decreased using an NIR laser rather than a shorter wavelength laser.

A number of lenses, filters and fiber bundles are used to collimate the laser beam, guide it toward the sample and collect the scattered light from the sample. These optical components must be chosen carefully, to just collect the Raman scattered light.

The spectrometer is used as a wavelength selector, and employs diffraction gratings to separate different wavelengths and prepare them for detection. The optical resolution of the spectrometer (the minimum difference in wavelengths that the spectrometer can distinguish them) depends on its grating, which is determined by the number of grooves and the blaze angle. The grating determines the range of wavelengths that the spectrometer can use. The optical resolution of the spectrometer also depends on its slit (8 mm height and 25 μm width in our setup), which is the entrance point of the beam, and the shape of the slit should correspond to the fiber bundle that carries the scattered light. As the width of the slit decreases, more paraxial rays enter the spectrometer which significantly improves the optical resolution of the spectrometer. Sharper and more rectangular images of the slit increase the optical resolution. The optical resolution of the spectrometer is also improved when the image width of the entrance slit is greater than the pixel width of the detector array..

The detector of a Raman setup could be a charge-coupled device (CCD) or photomultiplier tube (PMT), and its role is to convert the entrance photon intensity to an electronic signal that can be processed using the software. Finally, the spectrum of Raman intensity is plotted as a function of the wavelength or wavenumber, and can be used for further data analysis. This Raman spectrum shows the vibrational mode of the constitution of the sample. According to these assignments, and using data analysis tools, the presence or breakdown of the chemical bonds in the analyte is traced, and a model to interpret the correlation between the Raman peaks and the constitution of analyte is achieved.

2.1.2 Raman spectroscopy techniques

Although Raman spectroscopy introduced the fingerprinting technique that can be used to reveal unique and important information about materials, it does have some limitations that should be considered carefully. One of the main limitations of the Raman effect is its intensity, which is too low and has an impact on its monitoring and diagnostic applications. Another limitation of Raman spectroscopy is the intense fluorescent background of some analytes. This is usually strong, which makes observing the weak Raman peaks of many solutions present in the fluorescent background very difficult. Some attempts to overcome these types of issues

have led to new ways to change Raman instrumentation,. The main reason for changing Raman instrumentation is to improve the quality of Raman signal, in order to acquire more accurate information. Some specific Raman spectroscopic techniques that have been developed and introduced are based on linear Raman spectroscopy, while others are obtained through a nonlinear approach. To understand these phenomena, we need to focus on light-sample interactions. When a sample is illuminated by the laser beam, the electric field of laser light (E) induces a dipole moment (P) in the molecule. As long as the incident laser light is weak, the linear relationship between E and P is given by the following linear equations;

$$P = \alpha E \qquad (1)$$

$$E = E_0 \sin \omega t \qquad (2)$$

In these equations, α is the linear polarizability of the molecule and ω is the angular frequency of incident light. The polarizability of molecule, due to electric field of light, will oscillate at molecule's characteristic angular frequency ω_{vib}:

$$\alpha = \alpha_0 + \alpha_1 \sin \omega_{vib} t \quad (3)$$

By substituting equations (2) and (3) in equation (1), we can drive another expression for P that shows three chance of oscillations for induced dipole moment:

$$P = \alpha_0 E_0 \sin \omega t + \frac{1}{2}\alpha_1 E_0 \cos(\omega - \omega_{vib})t - \frac{1}{2}\alpha_1 E_0 \cos(\omega + \omega_{vib})t \quad (4)$$

These oscillations at unshifted frequency (ω), down shifted frequency ($\omega-\omega_{vib}$), and upshifted frequency ($\omega+\omega_{vib}$) correspond to Rayleigh scattering, stokes scattering, and anti-stokes scattering respectively.

The development of lasers, spectrometers and detectors has encouraged Raman spectroscopy, to find new areas of research and development, and ultimately provide improved Raman techniques.

Linear Raman techniques include SERS, Tip-enhanced Raman Spectroscopy (TERS), Raman Optical Activity (ROA) and Resonance Raman Spectroscopy (RRS), which are all reviewed in many papers [6-8].

SERS is a technique used to enhance Raman signals using nanoparticles. This is described in the following section.

TERS, another method to enhance Raman signals, is very similar to SERS except that it uses a Scanning Probe Microscopy (SPM) system as the sampling device. The spatial resolution of Raman spectroscopy goes down to nm with the TERS technique [6].

In contrast to TERS, Transmission Raman Spectroscopy (TRS) is another technique to collect data from the bulk sample. With this technique the light is sent to a translucent sample, then moves along the entire thickness of the sample [6].

ROA is another Raman method that is sensitive to chiral molecules. The intensity of Raman scattered light from a chiral molecule depends on the degree of circular polarization of the incident beam. This provides a way to measure the vibrational optical activity of the molecule [7].

RRS is considered an effective method to enhance specific bands. The resonance effect is related to exciting the electronic bands by a laser light with energy that is equal to the energy of the electronic bands. RRS can be used to identify specific bands of some biological molecules, such as proteins and large polyatomic molecules, show the sites of a molecule and differentiate between complicated molecules [8].

Nonlinear Raman Spectroscopy is another method used to resolve technical issues. It applies a few techniques such as coherent anti-stokes Raman spectroscopy (CARS), which provides a technical way to avoid the fluorescent background issue. The technique is based on the third order nonlinear coefficient, where two laser beams, one of which is stronger than the other, are combined. The wavelength of one laser is constant, and the other is adjustable. The interaction of the two laser beams, only if the frequency difference between two lasers fields

coincides with the frequency of a molecular vibration, provides a strong Raman signal at the sample. This method has been used as a non-invasive imaging technique, and implemented as a novel approach for microscopic measurement in some biological applications [8].

2.1.3 Raman sampling arrangements

Raman spectroscopy can be used to analyse most solid, liquid and gaseous materials. Clear materials are preferred, as low density gases with typically very low concentration of molecules, and highly reflective solids have issues when collecting scattered light. The volume of the sample that is illuminated by the incident laser beam should be as large as possible, to maximize the chances of interaction between the light and the molecules and improve the efficiency of the Raman scattering. Thus, the geometry of the excitation/collection system must be carefully selected.

2.1.3.1 Cuvette sampling

A cuvette is a simple, standard type of liquid sample holder that can be used for Raman spectroscopy. The cuvette is usually in square or rectangular form with transparent sides, and is made of non-fluorescence glass that has the lowest interference with the Raman signal. The main advantage of using a cuvette as a sample holder is its simplicity. In cuvette based geometry, the laser beam is tightly focused on the sample by an objective lens, with high intensity at the focal point within the sample in the cuvette (Figure 2-2). The figure of merit (FOM) of laser beam focusing system is a parameter that shows the performance of sampling. This parameter is defined as:

$$\mathrm{FOM} = \frac{\mathrm{L_{int}}\lambda_{\mathrm{exc}}}{\mathrm{A_{eff}}} \quad (5)$$

where L_{int} is the interaction length, λ_{exc} is the excitation light wavelength, and A_{eff} is the effective cross-section area of laser beam and analyte. As the effective cross-sectional area is smaller (or has a higher focused intensity), a shorter effective interaction length L_{int} is achieved, and the two properties counterbalance each other. Accordingly, FOM parameter of cuvette geometry, which is diffraction-limited for tighter focus of laser beam, is almost 2 and it is inefficient for increasing the effect of light-matter interaction [9].

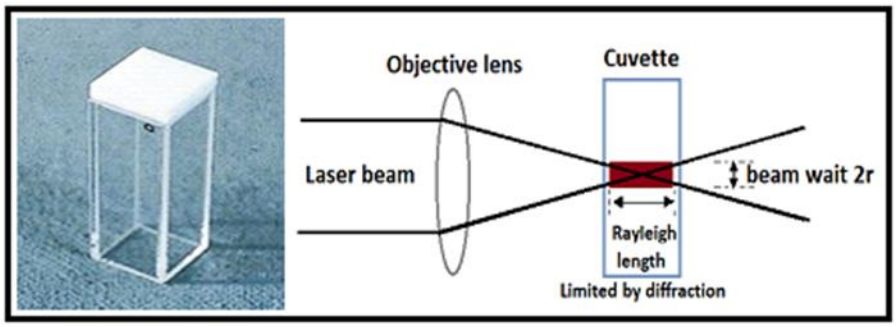

Figure 2-2 Cuvette based geometry with Lint limited by the Rayleigh length in a focused free-space laser
Reprinted with permission Ref. [9]. Copyright (2015) Springer

2.1.3.2 Fiber optic (waveguide) sampling

Marcatili et al. showed that one way to increase the light-matter interaction is to use dielectric capillaries (i.e. metal-coated tubes) [10]. Using capillaries increases the effective interaction length, but has the disadvantage of high propagation losses, as shown in Figure 2-3 [11]. A fiber optic approach is an interesting alternative to the capillary technique as it has higher intensity, and can decrease the propagation loss of the system.

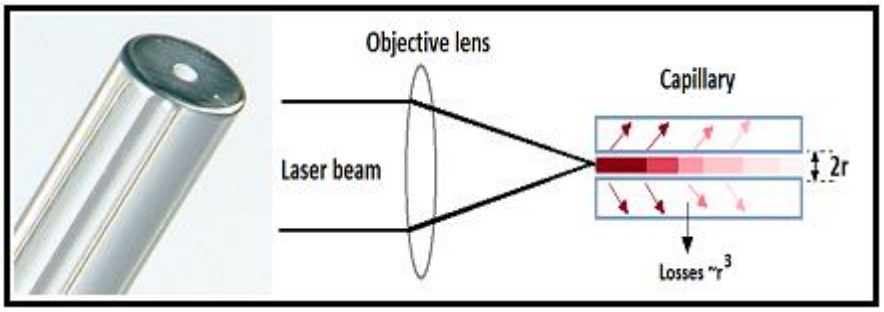

Figure 2-3 Capillary based geometry with high propagation losses.
Reprinted with permission Ref. [9]. Copyright (2015) Springer

A fiber optic typically consists of two parts: a core, and the surrounding medium (known as the cladding). It can be used as a waveguide if the refractive index of the core is higher than the refractive index of the cladding. In such conditions, the input light could be confined within the core and guided along the fiber axis, which would make the propagation loss with fiber optics less than with the capillary type. The guidance power in the optical fiber is expressed by its

numerical aperture (NA), which is defined in terms of the refractive index of the core and cladding, or the sine of the maximum acceptance angle of the input beam (Figure 2-4).

$$\mathrm{NA} = \sin(\theta) = \frac{1}{n_0}\sqrt{n_{\mathrm{core}}^2 - n_{\mathrm{cladding}}^2} \qquad (6)$$

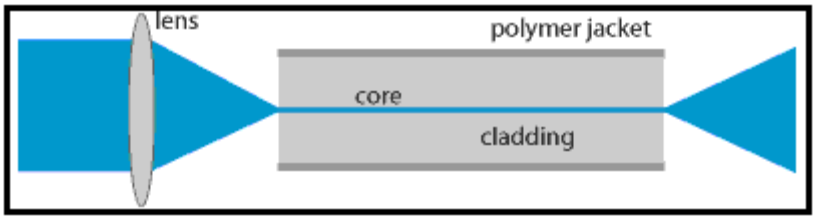

Figure 2-4 Fiber based geometry with less propagation losses

In order to use fiber optics as a sampling device a special kind of fiber, known as a Photonic Crystal Fiber (PCF), is required. This fiber consists of one hollow core in the center and more smaller core in the cladding, as shown in Figure 2-5. HC-PCF is an ideal configuration for effective interactions that require a diffraction-free, single-mode waveguide with a core diameter equal to the waist of the focused laser beam. This type of PCF provides an increase in interaction length, with less loss than cuvettes or capillaries. A comparison between Raman intensity of heparin using cuvette or HC-PCF is shown in Figure 2-6. According this comparison, Raman intensity of heparin with HC-PCF configuration is 90 times higher than cuvette configuration [12].

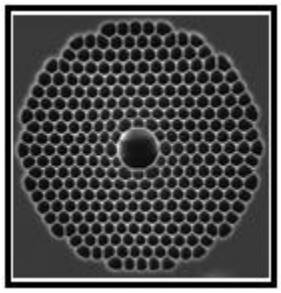

Figure 2-5 Microscope picture of a HC-PCF.
Reprinted with permission Ref. [9]. Copyright (2015) Springer

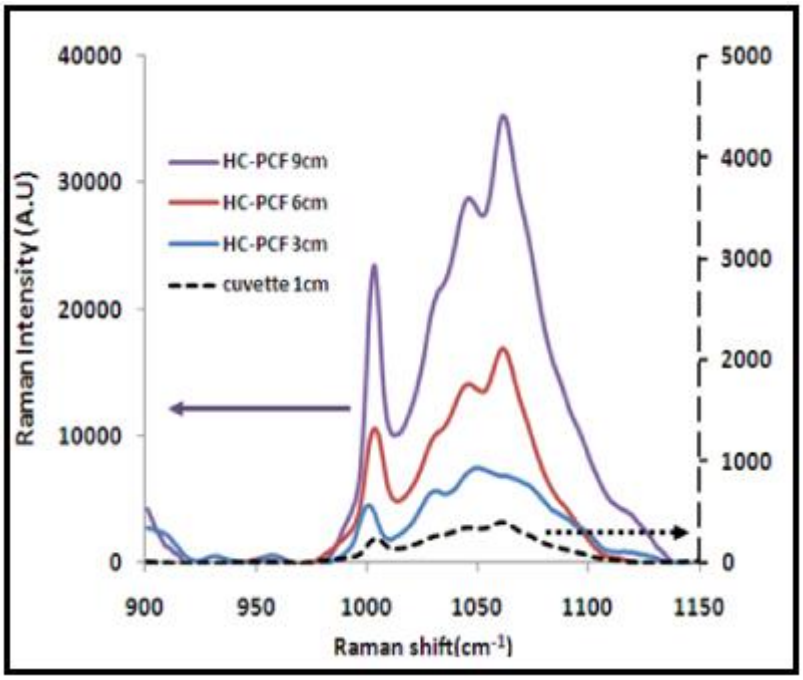

Figure 2-6 Comparison between Raman spectra of heparin using cuvette and HC-PCF.
Reprinted with permission Ref. [12]. Copyright (2011) OSA

The propagation of light through an HC-PCF is based on the Photonic Bandgap (PBG) effect. The light in a microstructure such as HC-PCF propagates along the axis of the fiber within the core(s), and does not propagate through the cladding area [13], as shown in Figure 2-7.

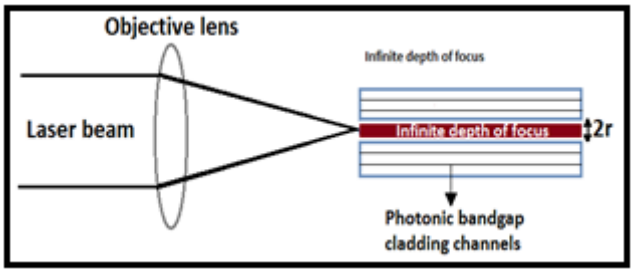

Figure 2-7 HC-PCF guiding mechanism in HC-PCF.
Reprinted with permission Ref. [9]. Copyright (2015) Springer

An HC-PCF can be filled by a selective or non-selective filling approach. With selective filling, the light is guided by the Total Internal Reflection (TIR) rather than the photonic bandgap effect, which does not use the full light guiding potential of photonic bandgap property of HC-PCF. This

method that was used by Irizar et al. to enhance the Raman signal obtained from nanoparticles [14] is quite cumbersome, however this method can be used to achieve asymmetric coupling, birefringence, hybrid guiding, tunable beam diffraction, self-defocusing, and high nonlinearity. The second method of using HC-PCF for characterizing samples is to fill the holes non-selectively. An empty HC-PCF guided at 800 nm (when empty) was used by Yang et al. to guide a laser at 785 nm wavelength when filled with the sample [15]. In this case, the photonic bandgap (or transmission wavelength) of the sample filled HC-PCF shifts to another transmission wavelength which it does not match the excitation wavelength. The shift in the transmission wavelength of HC-PCF can be determined from the equation given by Antonopoulos et al. [16], as follows:

$$\lambda' = \lambda_0 \left[\frac{1 - \left(\frac{n_{liq}}{n_{sil}}\right)^2}{1 - \left(\frac{n_{air}}{n_{sil}}\right)^2} \right]^{1/2} \qquad (7)$$

where λ_0 is the wavelength of the fiber when empty, λ' is the shifted wavelength of the fiber when filled with the sample, n_{liq} is the refractive index of the sample, n_{air} is the refractive index of the air, and n_{sil} is the refractive index of the HC-PCF, which is made of silica. Thus, depending on the excitation wavelength, the fiber that could guide this wavelength effectively when filled can be determined.

According to equation (7), light guiding property changes depend on the refractive index of the filled sample. This means that with non-selective filling the light is weakly guided into the HC-PCF, and not tightly confined in the sample-filled core region. This results in weak light-matter interaction, and once again the photonic band gap is not preserved.

HC-PCF has several major advantages over conventional sample cells [17-20], including:

- Low waveguide losses of a few dB/m enables the use of long optical-path lengths, and greatly enhances the effective light-matter interaction, as shown in Figure 2-7.

- The sample under analysis typically uses at least a milliliter of volume when examined with a test tube or cuvette. In contrast, HC-PCF uses samples in the nano to picoliter range, significantly decreasing the sample consumption rate.

- The small required sample volume and large overlap of the propagating laser mode field with the sample in HC-PCF provide the potential to develop simple, compact and sensitive biosensors. At comparable input power, the intensity in the hollow core is five orders of magnitude higher than with cuvette-based approaches.

- PCFs are fabricated with available technology from chemically inert, high-quality silica glass, with negligible scattering, absorbance or fluorescence.

- Finally, PCF in conjunction with suitable data analysis can demonstrate the potential of non-invasive and label free detection of biomolecules.

2.2 Surface Enhanced Raman Scattering (SERS)

Surface enhanced Raman scattering (SERS), which was first observed by Fleischmann in 1974 [3], is a powerful analytical method for the detection and identification of extremely low concentrations of molecular species. It overcomes the barrier of low Raman cross-section by exploiting the 'local electromagnetic field' generated within the assembly of nano-structured material and creating large field enhancement due to the electromagnetic coupling between the nanoparticles. SERS relies on adsorption of the analyte onto the surface of metal structures: typically silver, gold or copper. Under these conditions, the Raman signal of the target molecule is enhanced by several orders of magnitude, which enables detection down to a single molecule [21-24].

2.2.1 SERS mechanism

As described in the previous section, the intensity of Raman signal is proportional to the square of the electric dipole moment of the molecule ($p = \alpha E$). The polarizability (α) or electric field (E) are possible reasons for this enhancement; if it is due to polarizability it is called chemical enhancement, and if caused by electric fields it is called electromagnetic (EM) enhancement.

With EM enhancement, the SERS mechanism is described by the electromagnetic model, which uses an illuminated metal in an electromagnetic field. According to EM enhancement, the electric field near a metal particle is enhanced due to excitation of the surface plasmon, which is a confined electron gas near the surface of the metal. The adsorbed molecules on the metal particles show an enhanced Raman signal due to this excitation of the local electric field [25]. In this model, it is assumed that the metal diameter is much smaller than the wavelength of the exciting light.

The other enhancement mechanism using a chemical approach can manifest in three ways [26-27]. Figure 2-8 shows the different types of enhancement mechanisms, with the highest occupied molecular orbit (HOMO) and the lowest unoccupied molecular orbit (LUMO) indicated. The first chemical mechanism, known as the ground state chemical enhancement, is the simplest. It can occur when the adsorbate does not bind covalently to the metal, and enhancement takes place due to ground state chemical interactions between the molecule and nanoparticles that are not associated with excitation of the nanoparticle-molecule system. The presence of the metal disturbs the electronic structure of the analyte, causing a 'mild' change in its electronic distribution and changes the polarizability of the analyte. In this case, the charge transfer (CT) is not required.

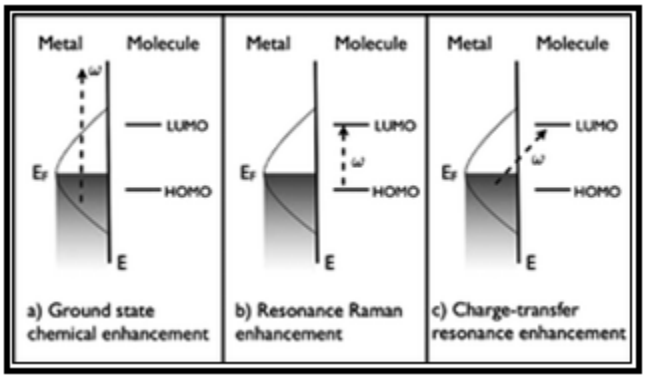

Figure 2-8 Illustration of the different types of enhancement mechanisms in SERS.
Reprinted with permission Ref. [27]. Copyright (2008) The Royal Society of Chemistry

The second resonance Raman enhancement involves the presence of nanoparticle surface-molecule complexes that either bind directly (covalent) to the metal, or bind indirectly with the

aid of an electrolyte ion (typically chloride). These surface conditions change the intrinsic polarizability of the molecule, and can also create a new electronic state that is explicitly, or close to being, in resonance with the laser, thus contributing to the enhancement of the resonant Raman type.

Charge-transfer resonance is the last of the chemical phenomena. It is basically a more sophisticated version of resonance Raman enhancement that involves charge transfer between the analyte and the metal. This can happen when the difference between the Fermi level (E_F) of the metal and the HOMO or LUMO energies of the molecule match the laser, and the excitation wavelength resonates with the nanoparticle–molecule charge transfer transitions.

Overall, the relative contribution of the different mechanisms depends on the experimental conditions. The chemical enhancement contributions to SERS are complex, and require highly accurate electronic structure calculations of the molecule-metal system.

The contribution of the chemical mechanism to the overall enhancement of SERS is much less than that of the EM mechanism. The enhancement factor (EF), a parameter used to evaluate SERS, shows that due to the EM mechanism SERS enhancement is approximately 10^6, compared to 10 to 100 for the chemical mechanism. Thus, the overall enhancement of SERS by the superposition of EM and CT enhancements can be in the 10^6 to 10^8 range [28]. It has been demonstrated that the use of stronger electromagnetic fields leads to even higher enhancement factors.

2.2.2 Localized surface plasmon resonance

As discussed in the previous section, enhancement from EM is due to surface plasmon resonance that is produced near the nanoparticle and the target molecule. Free electron charges on the metallic nanoparticles respond to external electromagnetic fields (laser light), and oscillate at resonance wavelengths.

Localized surface plasmon resonance (LSPR) depends strongly on the optical properties of metallic nanoparticles, which are described by a complex dielectric constant (ϵ) or a complex refractive index ($m = \sqrt{\epsilon}$):

$$\epsilon(\lambda) = \epsilon_r(\lambda) + i\epsilon_i(\lambda) \qquad (8)$$

$$m = n + ik \qquad (9)$$

where n (real part of ϵ) is the refractive index, and k is the absorption coefficient of the nanoparticles. LSPR depends on the wavelength of the incident light due to the wavelength-dependency of the dielectric constant of nanoparticles. For a metallic sphere in the presence of an external field, LSPR can be observed when the nanoparticle diameter (d) is much smaller than the laser light ($d \ll \lambda$). The electric field of the light can be considered uniform in this case, and it allows electrostatic equations to be solved [29]. The solution of Maxwell's equations for a spheroid metallic particle leads to an expression for the extinction $E(\lambda)$; that is, the sum of the absorption and scattering of a nanoparticle:

$$E(\lambda) \propto \frac{\epsilon_i(\lambda)}{(\epsilon_r(\lambda) + \chi\epsilon_{med})^2 + \epsilon_i(\lambda)^2} \qquad (10)$$

This relation shows that the electric field depends on the dielectric constants of nanoparticles. Silver, gold and copper nanoparticles, and their diverse optical constants, provide different enhancements. For example, AgNPs (UV to IR region) provide 10 to 100 times higher efficiency than gold nanoparticles (IR region) [29]. The other parameter, known as the shape factor (χ), describes the deviation from spherical particle geometrics into higher aspect ratio structures. This relation shows that the extinction of nanoparticles is strongly dependent on a particle's shape, and Figure 2-9 shows the dependency of plasmon resonance on particle shape. The major to minor axis ratio (r) varies from 1 to 10, and the red shift in the peak as the particle becomes more oblate is shown in Figure 2-9.

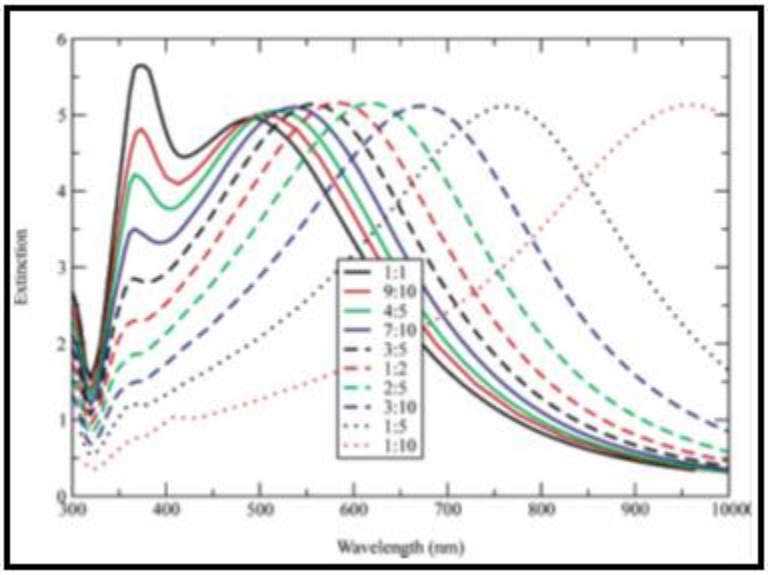

Figure 2-9 The extinction spectra of different spheroids with the same volume, corresponding to a sphere radius of 80 nm.
Reprinted with permission Ref. [34]. Copyright (2003) American Chemical Society

The red shift also appears when the nanoparticle size increases. SERS increases with particle size, because the enhancement of electromagnetic field intensity depends on the number of atoms that are excited, and the volume of the nanostructure [21]. This effect on surface plasmon is not limited to spherical particles. To illustrate, SERS of silver nanorods with an aspect ratio of 10, is 10 to 100 times higher than that of a sphere, simply due to shape [30]. Nanoparticles with sharp corners and edges produce higher SERS enhancement [31], as does metal with a rough surface due to more localized surface plasmon and stronger field gradients [32]. Certain roughness can cause activity in the surface plasmon and change the resonant frequency, which enables more scattering [33].

The laser wavelength is more critical when the surface plasmon absorption spectrum is relatively narrow. In this case, the wavelength dependency is very important because using SERS excites the surface plasmon by laser light. The broadening of surface plasmon absorption depends on the aggregation, and higher aggregation means broader absorption accompanied by a shift of the absorption maximum to a shorter wavelength [34]. Thus, if the plasmon peak is λ_1 when aggregation occurs, it shifts to λ_2, and $\lambda_2 > \lambda_1$. Making the nanoparticles less stable and forcing them to aggregate creates different absorption bands. However, enhancement depends

on the degree of overlap of the excitation wavelength with the plasmon frequency which, as discussed, is shifted during nanoparticle aggregation. Further, calculations show that the EM enhancement is strongly (i.e. to the inverse twelfth power) dependent on the metal-molecule distance; that is, as the distance increases the EF decreases due to the declining intensity of the dipole moment. However, this does not mean that enhancement requires direct contact of the metal surface and the molecule [35].

2.2.3 Recent progress in SERS

SERS can provide order of magnitude increases in Raman intensity, which overcomes the inherent weakness of Raman scattering. Over the past decade, various approaches to enhance the Raman signal have been attempted, with a wide range of analytes.

Sensitivity has attracted much attention lately, and some researchers have applied SERS to detect biologically relevant small molecules. Van Duyne et al. studied rapid and accurate detection of bioagents, and used SERS to detect anthrax biomarkers [36]. They described a procedure for rapid extraction of CaDPA from B.subtilis spores and simulants for B.anthracis spores, followed by SERS detection on reproducible and stable silver film over a nanosphere substrate (AgFON), and measured the spore concentration range down to 10^{-14} M. A group at the University of Georgia at Athens (UGA) placed rows of silver nanorods on a slide, and detected biological agents or pathogens at attomolar levels (10^{-18} M) [37]. Hongyan Liang et al. conducted other research related to highly sensitive SERS using monodispersed 'flower-like' AgNPs [38]. These nanoparticles have a rough surface, and were used for malachite green isothiocyanate (MGITC) molecule detection at concentrations down to 10^{-10} molar. Comparing the intensity of molecular peaks, they demonstrated that the sensitivity of this type of nanoparticle is 10^6 to 10^8 times higher than that of normal Raman scattering.

The main challenges of SERS are long-term stability and reproducibility, and the well-ordered metal nanostructure approach is a promising way to create a stable and reproducible Raman spectrum. In this ordered metallic nanostructure (as a SERS substrate), the periodic nanostructure is covered by the exciting laser, and because the spot of the exciting beam is typically in the micron range, the entire SERS substrate is excited homogenously. Baia et al.

reported using corrugated gold film on highly ordered polystyrene nanospheres as SERS-active substrates. SERS experiments were efficient enough to detect discrete molecules adsorbed onto the surface [39]. Zhang et al. investigated and developed a SERS substrate using physical vapour deposition of silver nanolayers onto different types of paper. They demonstrated the detection of analyte concentrations down to 10^{-10} M [40], which is an important step toward the development of a low cost SERS sensor.

2.3 Multivariate analysis

Raman spectroscopy has many applications in diverse areas. Its ability to detect and monitor different molecular mixtures creates a wide range of medical applications, including pharmaceutical field [41-42], pathology [43], microbiology studies [44], nutrition researches [45] and diagnostic applications (e.g. diagnosing breast cancer) [46], and agriculture [47]. In addition, numerous studies have illustrated the significant potential of Raman spectroscopy for forensic investigations [48-49], environmental engineering [50], archaeology [51], geoscience [52], and astrobiology applications [52].

All these applications deal with spectra that are collected and observed to determine certain properties, such as the concentration of a chemical in a solution for example, or the number of specific cells in a clinical sample. Once the spectra are recorded, they are analysed to identify the relationship between them and the property of interest. The recorded spectra must provide the information required to determine the need for further prediction or diagnosis.

As with many other phenomena, Raman spectroscopy has a multivariate nature that needs to be interpreted computationally [41-52]. As mentioned in Section 2.1, it can show peaks at different wavenumbers due to specific molecular vibrations of the chemical bonds in the materials. For example, different concentrations of ethanol show the same Raman peaks with different intensities at 433 cm^{-1} (C-C-O bend), 882 cm^{-1} (C-O-C symmetrical stretch mode), 1051 cm^{-1} (C-O anti-symmetrical stretch mode), 1097 cm^{-1} (C-H rock), 1276 cm^{-1} (CH_2 twist), and 1454 cm^{-1} (C-H assymetric deformation). As illustrated in Figure 2-10, all eight samples of ethanol water solutions have identical peaks at those wavenumbers, but the Raman intensities vary according to the ethanol concentration of the samples. The variations of the Raman intensity at

the same peaks in different samples reveals that the relationships between ethanol concentrations are dependent variables, while the Raman intensities at different wavenumbers are independent variables.

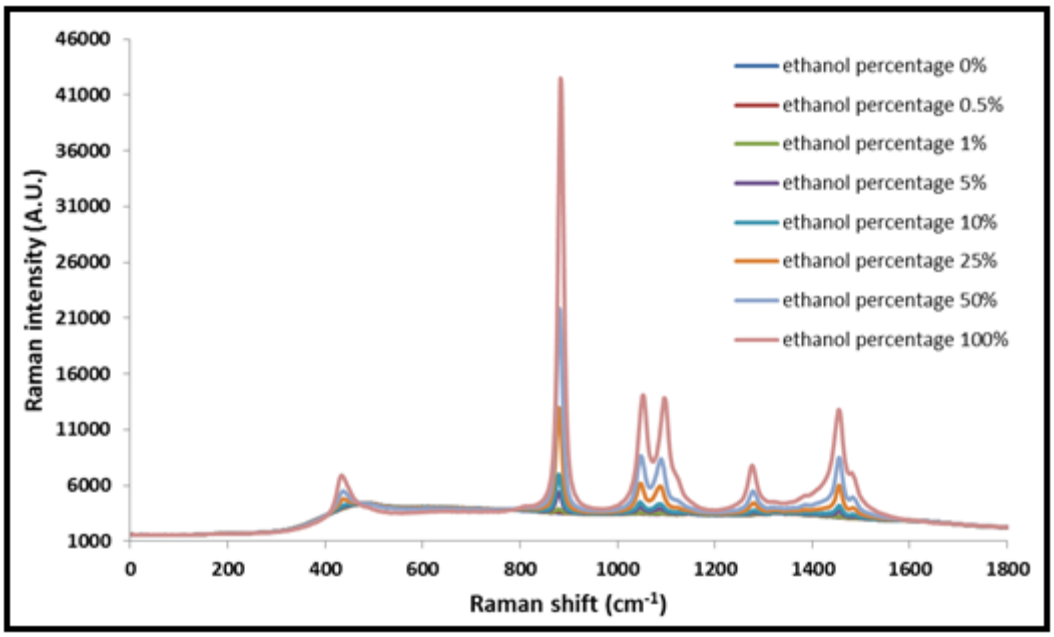

Figure 2-10 Raman spectra of eight samples

This is one of the reasons that Raman spectroscopy should be considered a multivariate effect, and why the data needs multivariate analysis (MVA). MVA is an effective way to transform raw spectra into the quantitative and qualitative information required. The main objectives of MVA are to develop a model that can identify and classify similarities and differences between samples, or find a model that can predict properties in future samples, such as the concentration of a component or thickness of a layer. Developing an effective model for classifying samples or predicting sample properties requires preprocessing of the dataset before applying MVA.

2.3.1 Preprocessing

Preprocessing plays an important role in developing an efficient model that can be used to interpret a dataset or estimate certain properties of new samples. Irrelevant or non-informative data has a major impact on the constructed model, so removing it improves the results.

Preprocessing is performed by one or more of following tools: baseline transformation, normalization of samples, sample weighting and smoothing. In my PhD research, MVA of all experimental datasets was performed with and without all the preprocessing types, to check if they were required. However, the objective was to reduce the limit of detection using optimized MVA, and preprocessing procedures such as normalization, weighting and smoothing can cause shifts in peak positions, which complicates understanding the raw dataset and degrades the resolution of the spectrum. Therefore, we focused on baseline transformation only, one of the most important preprocessing steps.

Baseline transformation:

The background effect is an important aspect of the Raman spectra that can negatively affect the results of data processing, so its removal is considered an essential preprocessing step. The baseline of the Raman can be related to strong fluorescence, blackbody radiation (hot samples), room light, samples highly diluted with water, or scattering from either the quartz window of sample holder or any parts, other than analyte [53]. The baseline issue can be addressed with different approaches, such as subtracting methods or multiplicative scatter correction (MSC) [54].

In baseline subtraction, it is assumed that X_B and X are the spectra of the sample before and after baseline correcting, respectively. The relation between X_B and X can be written as:

$$\mathrm{X_B} = \mathrm{X} + \mathrm{a_0} + \mathrm{a_1 X} + \mathrm{a_2 X^2} + \cdots \qquad (11)$$

By assuming a linear or nonlinear mathematical model in terms of an independent variable (e.g. wavenumber) for the baseline, subtracting it from the spectrum gives the spectrum of interest. The derivative method is a technique to calculate the coefficients of equation (11), which is based on taking the derivation of the spectrum with respect to the variable, and continuing the derivation in higher order according to the simple or complex form of the baseline. Polynomial fitting is another method to estimate the baseline and subtract it from the initial spectrum. There are numerous computer programs that can fit the polynomial to the data. Figure 2-11 demonstrates a Raman spectrum that is baseline corrected using the polynomial fitting method.

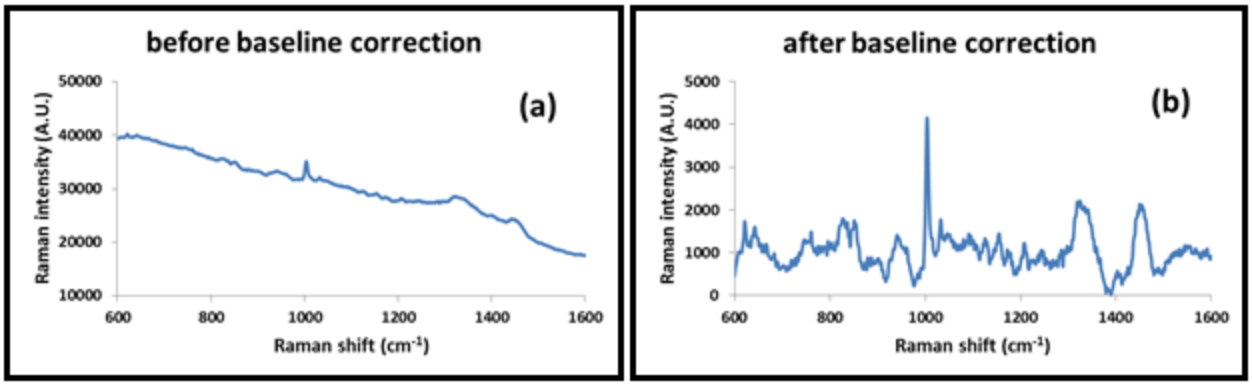

Figure 2-11 A Raman spectrum a) before and b) after baseline correction

Multiplicative scatter correction (MSC) is another technique to remove the baseline from the spectra, and reduce the scattering effect in diffuse reflection and transmission NIR spectra. In this method, the average spectrum of the dataset is assumed to be the best representative sample. The i^{th} spectrum can be fitted to the average spectrum using regression onto the average spectrum, and in terms of *a* (intercept), *b* (slope), and *e* (residual) which are constant for all wavenumbers:

$$x_i = a_i + b_i\bar{x} + e_i \qquad (12)$$

The corrected spectrum in terms of raw spectrum and constant coefficients is described as:

$$x_{i,MSC} = \frac{x_{i,raw} - a_i}{b_i} \qquad (13)$$

As described, this method cannot be applied on the spectrum of one sample as it requires a dataset, while the subtracting methods can be performed on an individual spectrum. This method is usually applied over the entire range of variables, but occasionally there is noise in the result. To avoid this, it should be used carefully in the selected range(s) of the variable [55].

2.3.2 Multivariate data analysis methods

MVA methods are generally categorized as unsupervised and supervised. The aim of the unsupervised methods is to find similar groups and reveal the hidden structure of unlabeled data. Principal Components Analysis (PCA), K-means Cluster Analysis (KMCA) and Hierarchical Cluster Analysis (HCA) are a few unsupervised methods. With supervised methods, the labeled data is used to find a model in the data that can predict the values of labels on future unlabeled

data. Multiple Linear Regression (MLR), Partial Least Squares (PLS) and Support Vector Machines (SVM) are supervised methods.

KMCA is a simple clustering method introduced by J. A. Hartigan [56] and used in many research areas. For example, M. Miljkovic et al. [57] used this to differentiate the nucleus, the nucleoli, and areas high in mitochondria in HeLa cells. Simplicity and speed are the major benefits KMCA, and a key weakness is that the number of clusters should be determined before clustering, and if the number of observations is relatively low the determination will affect the clustering. Another drawback of KMCA is that the outliers are included in the clusters, and could significantly change the clustering [58].

HCA is an unsupervised method to classify data that builds an arrangement of clusters in a tree structure. This powerful technique has been used frequently in Raman and IR imaging. T. P. Wrobel et al. used Fourier transform infrared (FT-IR) spectroscopy followed by HCA to study the content of free fatty acids, triglycerides, cholesteryl esters, and cholesterol in the aorta of mice with atherosclerosis [59]. One of the advantages of HCA is that there is no requirement to specify the number of clusters before beginning the procedure, while the main drawback is the high computational complexity of a large dataset, which increases the computation costs [58].

Regression analysis is typically applied to estimate relationships between variables, and MLR is the most common form of linear regression analysis for supervised methods. MLR is a supervised method to estimate the linear relationship between dependent variables (y) and independent variables (x) in labeled data, and it generates a model that could be used for predicting y in unlabeled data in the future. The main issue with MLR is when there are one or more relationships between independent variables. However, though this collinearity between variables, which is common in spectroscopic applications, causes a misleading result in the Least Square Criterion of the model, this can be managed by other methods, such as the PLS technique [60]. SVM is another supervised method that is used to classify labeled observations and identify the class of new unlabeled observations. Introduced by Cortes and Vapnik [61], this method considers input labeled data as two different classes, and a linear or non-linear model is generated to classify new unlabeled objects. SVM was successfully used by Fernandez Pierna

et al. to classify compound feeds by NIR spectroscopy [62]. The main drawback of SVM is its complexity, as it causes the algorithm to be slow [63].

There are many books and reviews about different types of MVA methods. In the following sections two unsupervised and supervised methods (i.e. PCA and PLS respectively) which are frequently used in the rest of this thesis are described in more detail.

2.3.2.1 PCA

The main function of MVA is to reveal hidden information such as similarities and differences in data, and to predict future observations. PCA is an unsupervised method that provides the best view of information and interpretation patterns in data, it represents the similarity of the observations, and it can also reduce the multidimensional space of a variable to a space with fewer dimensions. To understand how PCA works, we considered each sample as a point in a multidimensional variable space, with each point representing a different response value at different variables. For example, Figure 2-12 shows the Raman spectrum of a sample in the range of 901 cm^{-1} to 1200 cm^{-1}. In variable space, the sample is shown as a point with 400 values as coordinates. The first coordinate value specifies Raman intensity at 901 cm^{-1}, the second specifies Raman intensity at wavenumber 902 cm^{-1}, and so on. In this example, a data table of the Raman intensity of samples is shown as a group of points in a variable coordinate system (Figure 2-12). In the graphical view, more similar samples have similar coordinates, while different samples have very different coordinates and are located far away from other samples.

If we specify all the corresponding points for all samples in a variable space, the variance in the dataset can be determined. By definition, the central axis of a direction is called the first Principal Component (PC1), and most samples are located along it [55]. PC1 shows the direction of the maximum variation in the samples. The presence of samples that do not lie along PC1 indicates that the dataset needs another direction to manage the remained data points; this is known as the second Principal Component (PC2). The PC2 direction is orthogonal to the PC1 direction, and represents the second largest variance in the dataset (Figure 2-12). In the same

way, higher orders of PC can be defined and used to interpret the structure of the entire dataset.

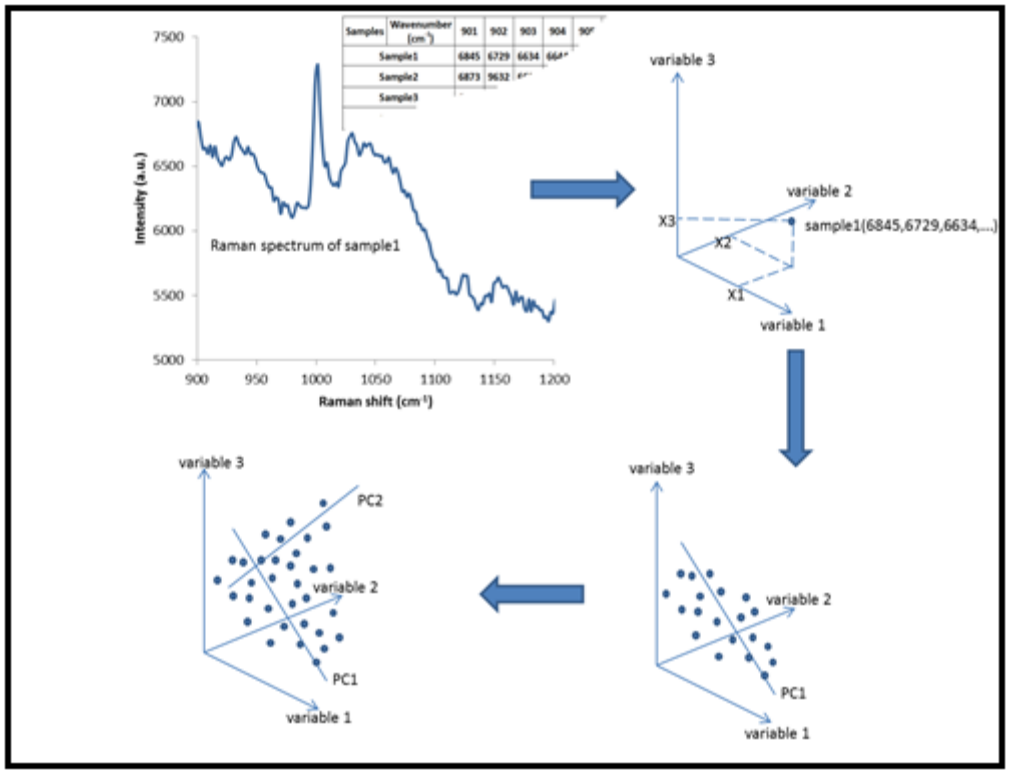

Figure 2-12 Raman spectrum of a sample and its presentation in multidimensional space

High PC numbers depend on the number of samples and variables. For n samples and p variables, the maximum PC number is either n-1 or p, depending on which is lower [55]. We preferred to use the lowest number of PCs, because higher PCs are related to smaller data variance, and can be considered noise.

To achieve an effective model, the dataset must be preprocessed before applying PCA. As a part of preprocessing the data must be mean-centered to avoid common offsets in the data points, and symmetrical compared to raw data [64]. All the data points are then measured relative to new origin point of the variable space, which is the mean of the data points (Figure 2-13). Another preprocessing task, known as scaling or weighting, expresses the data point variances in the same unit, and makes all variables truly comparable. For example, when one variable is measured in kg and another is measured in mg one variable dominates the other due

to its range, which means the variable variances are not comparable. Multiplying each observation by the inverse of the standard deviation ($\frac{1}{\mathrm{SDev}}$) means that centering and scaling are combined (autoscaling), which ensures that all variables have the same role in the model [55].

The main goals of PCA are data pattern interpretation and revealing hidden data information, and these require a bridge between the variable space and PC space. The loading and score plots, two important results of PCA, are explained in the following sections.

2.3.2.1.1 Loadings

Any PC can be considered a vector in variable space, and be represented by a combination of the unit vectors of different directions of this space [55]. The loading of a variable on a PC is defined as the cosine of the angle between them (Figure 2-13), and it defines the correlation between the variable and the PC. For p variables there are p loading coefficients for each PC that reflect the contribution of different variables to that PC.

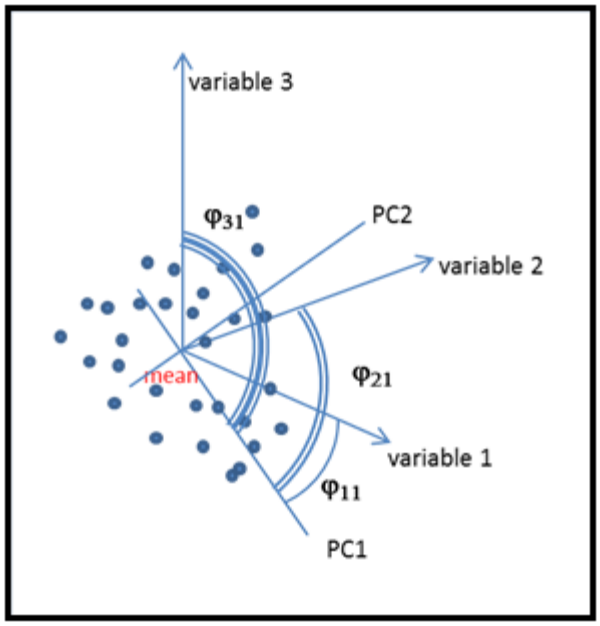

Figure 2-13 The loading of a variable

According to this correlation definition, the loading of each variable can range between -1 and 1, and might be used to describe the structure of the data. If the loading of a variable is high (close to +1 or -1), the contribution of that variable to that PC is high. Moreover, if two variables

have high loadings with the same PC, the correlation between the variables is high. The loading of variables for a specific PC can be plotted and used to reveal the most important variables, as shown in Figure 2-13. All the extremes in Figure 2-14 are considered important variables, and the most important is located around 1000 cm^{-1}.

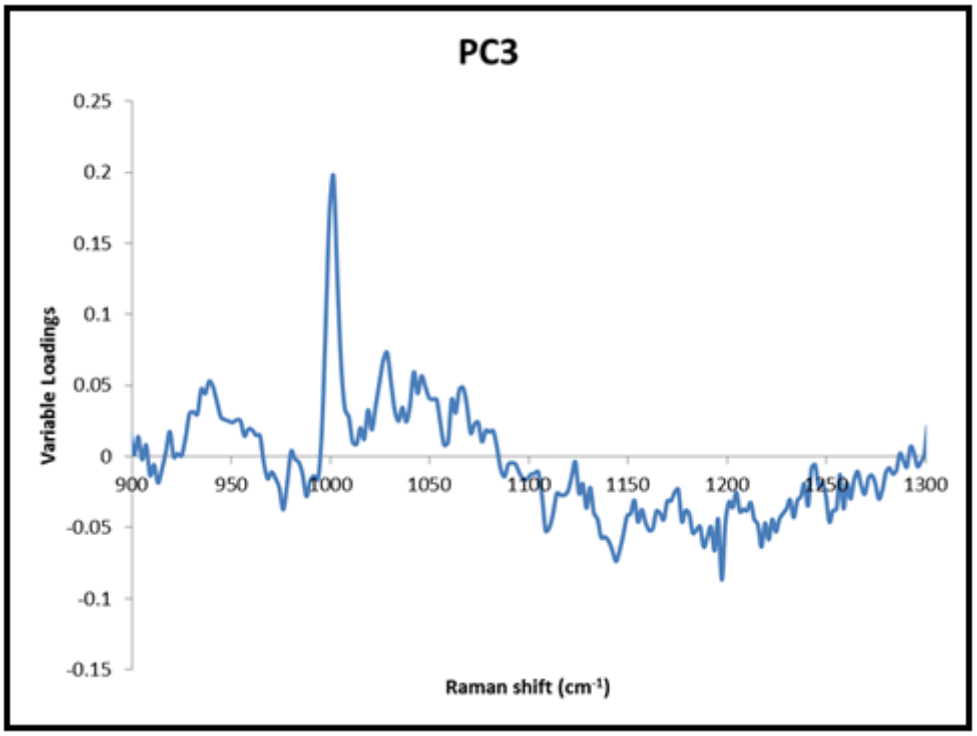

Figure 2-14 The loading plot of variables in the range 900-1300 cm^{-1}

2.3.2.1.2 Scores

Score is a parameter that shows the position or coordinate of a sample in PC space, and it can define the similarities and differences between samples. In the graphical explanation, the score of each sample is the signed distance from the origin in the direction of the PC axis in PC space (Figure 2-15).

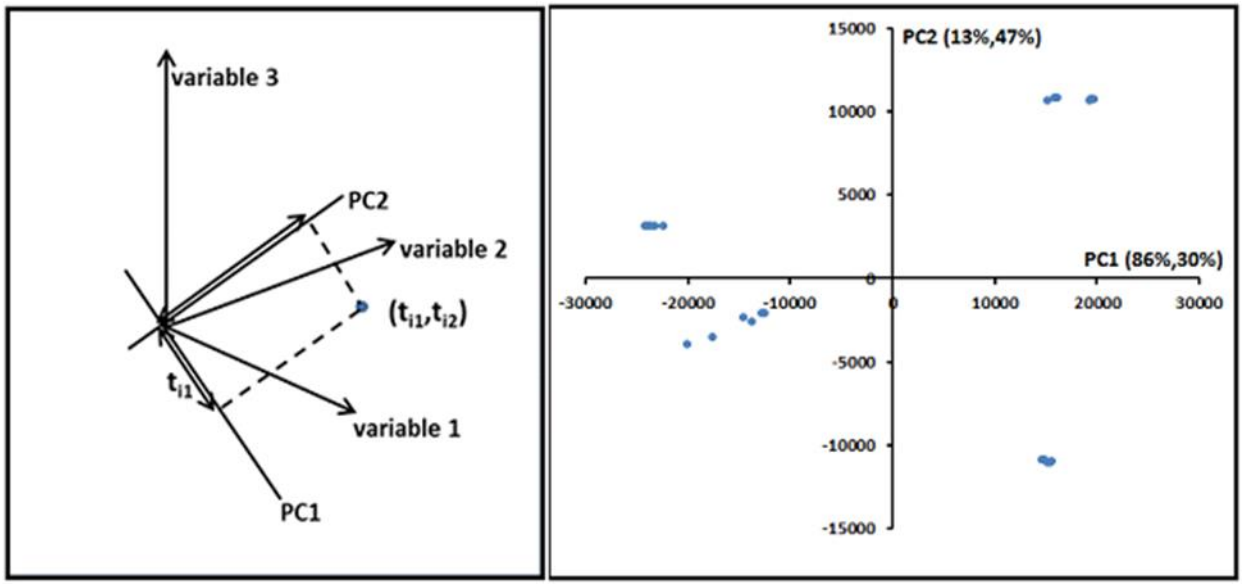

Figure 2-15 (a) The scores of a sample in two dimensional PC space,
(b) the score plot of a group of samples for two PCs

In Figure 2-15(a), the observation i in variable space is projected onto PC1 and PC2, and the projections along the direction of PC1 and PC2 are t_{i1} and t_{i2}, respectively. The score concept provides an interesting graphic of the observations (samples), and indicates how the samples are related to each other. This is called a score plot, and is usually the projection of observations on the PC1/PC2 plane because PC1 and PC2 have the largest variances. However, this projection can also be applied to any other pairs of PC. For example, the scores of a group of samples are presented in Figure 2-15(b), and they show four types of samples among these observations. The total variation of two PCs is about 99%, which explains a significant quantity of the variances of response in the analyte of about 77%.

2.3.2.2 PLS

Regression methods such as PLS typically create a fitting model for some observations. The PLS regression model describes how some independent variables (X) as predictor variables, explain other dependent variables (Y) as response variables. The dependent variables are usually the measurements that are expensive or difficult to take, while the independent variables are less expensive and easy to do. For example, in Raman spectroscopic application, which is the theme

of this study, the analyte in different samples is considered a dependent variable Y, and Raman spectra of different samples are considered X. When the number of independent variables is large compared to the dependent variables, it is good idea to project variable space to PC space with the PCA method, to reduce the dimensionality of the raw data. A PLS regression model can be written in a simple linear form of the relationship between X and Y, such as equation (14):

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + e \qquad (14)$$

or in matrix representation as in equation (15):

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \qquad (15)$$

where the **Y** matrix represents the observed response values, the **X** matrix represents predictor values, and the **E** matrix, known as the called error or residual, is the difference between the observed and predicted Y-values. $\mathbf{X}$ is an n by p matrix where n is number of samples and k is number of independent variables, and $\mathbf{Y}$ is an n by m matrix where m is number of dependent variables. $\mathbf{B}$ is a p by m matrix known as the B coefficient, or a regression coefficient matrix. The goal is to find the $\mathbf{B}$ matrix with the least error $\mathbf{E}$, by performing the PCA of the $\mathbf{X}$ and $\mathbf{Y}$ matrices simultaneously to identify the principal components that explain the covariance between them [65-66].

There are numerous algorithms for PLS modeling. The classical approach for a PLS regression algorithm is non-linear iterative partial least squares (NIPALS), presented by H. Wold [67]. It works directly on the centered X matrix, handles missing values in large datasets and considers only a few PCs. With this algorithm, PCs and models are estimated using iterative least squares by calculating one PC at a time, then the next, and so on. The algorithm causes higher errors if it includes a higher number of PCs [68]. Many researchers have worked on improving the stability of the results so it can handle large datasets or speed up the calculations. One of these is the statistically inspired modification of PLS (SIMPLS) method, presented by S. de Jong [69]. This algorithm is faster than NIPALS, and calculates the PCs as linear combinations of independent variables by maximizing the covariance between the X and Y matrices. It is also numerically stable for reasonable number of PCs [69]. Kernel is another algorithm that was described by Lindgren et al. [69], Rannar et al. [70], and improved by B.S. Dayal [71]. The improved kernel method is stable and fast for high numbers of PCs, and can handle large datasets [72]. Another algorithm, known as Krylov PLS, was introduced by M. Andersson. It uses the original X and Y

matrices based on looping the NIPALS algorithm, which is not stable for higher numbers of PCs [72]. Golub and Kahan presented the Bidiag algorithm, which is very fast but has higher errors than others [72-73].

These algorithms and others have been summarized by M. Andersson [73]. The Unscrambler® software used in this thesis can construct the PLS model using the NIPALS or Kernel algorithm. All calculations in this research were based on the Kernel algorithm, due to its capabilities mentioned above.

2.3.2.2.1 Calibration, validation, and prediction curves

The main goal of any regression method is to find a model that truly describes the relationships between datasets, and can also be used for predictions. PLS regression includes two steps: the first is to make a calibration curve and the second is to evaluate the calibration curve (validation). To perform PLS analysis on a large number of data they need to be divided: a calibration set to make the model, a validation set to find the best number of PCs and a prediction set to test the model independently [55, 74-75]. It is a common approach to validate the model using a validation set; this is known as test set validation (TSV). However, if the dataset is difficult to prepare or the analysis method is too expensive, the number of samples will not be enough to divide into the three sets. In that case, it is necessary to validate the model using an alternative method known as full cross validation (FCV), in which the entire dataset is divided into calibration, validation and test sets. The calibration and validation sets are used to make the model, and the test set is used to calculate predictions and evaluate it. This process is repeated with another sample until each sample appears only once in the test set [76]. The summation and average of the squared difference between the Y_ measured and Y_ predicted values for the test set provides the validation of Y.

Whether the TSV or FCV method is used for validation, the performance of a model is assessed by calculating a few parameters: the coefficient of determination of calibration (R_{cal}^2), the coefficient of determination of validation (R_{pre}^2), the root mean square errors of calibration (RMSEC), and the root mean square errors of prediction (RMSEP).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad (16)$$

$$SS_{res} = \sum_{i=1}^{q} \left(y_{i,ref} - y_{i,pred}\right)^2 \qquad (17)$$

$$SS_{tot} = \sum_{i=1}^{q} \left(y_{i,ref} - y_{ave}\right)^2 \qquad (18)$$

$$y_{ave} = \frac{1}{q} \sum_{i=1}^{q} y_{i,ref} \qquad (19)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{q} \left(y_{i,pred} - y_{i,ref}\right)^2}{q}} \qquad (20)$$

where q is the number of samples, $y_{i,pred}$ and $y_{i,ref}$ are the prediction and reference values for test sample i, respectively, and y_{ave} is the average of the reference values. R-squared is a statistical measure of how close the data are to the fitted regression line. As defined by the residual sum of squares (SS_{res}) and total sum of squares (SS_{tot}), R-squared provides the distance between the predicted, reference and average values of the dependent variables. R_{cal}^2 indicates the quality of the fitting, and shows how close the calibration set is to the fitted regression line. If the R_{cal}^2 is very close to 1 the calibration set is well fitted, while a value close to 0 means the fitting is poor. The R_{pre}^2 indicates how effective a fit can be expected for future predictions in a range of 0 to 1; the closer to 1, the better the prediction. The RMSEC is a measure of the dispersion of the calibration set from the regression line, while the RMSEP shows the dispersion of the validation set from the regression line, and describes the prediction error. Thus, the lower the RMSEP, the better the prediction accuracy.

Figure 2-16 illustrates a typical calibration curve, an important result of PLS regression models. The R-squared of the calibration and validation sets shows the sets are well fitted.

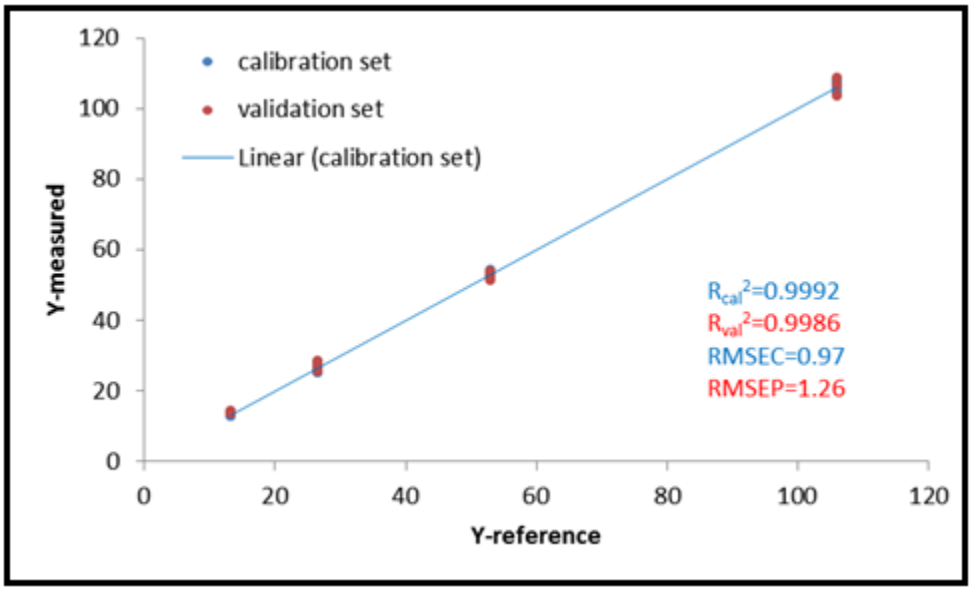

Figure 2-16 The typical calibration curve for PC4

A calibration curve can be plotted for any number of PCs. The calibration curve in Figure 2-16 is plotted for 4PCs, and the optimum number of PCs indicates the minimum RMSEP, as shown in Figure 2-17.

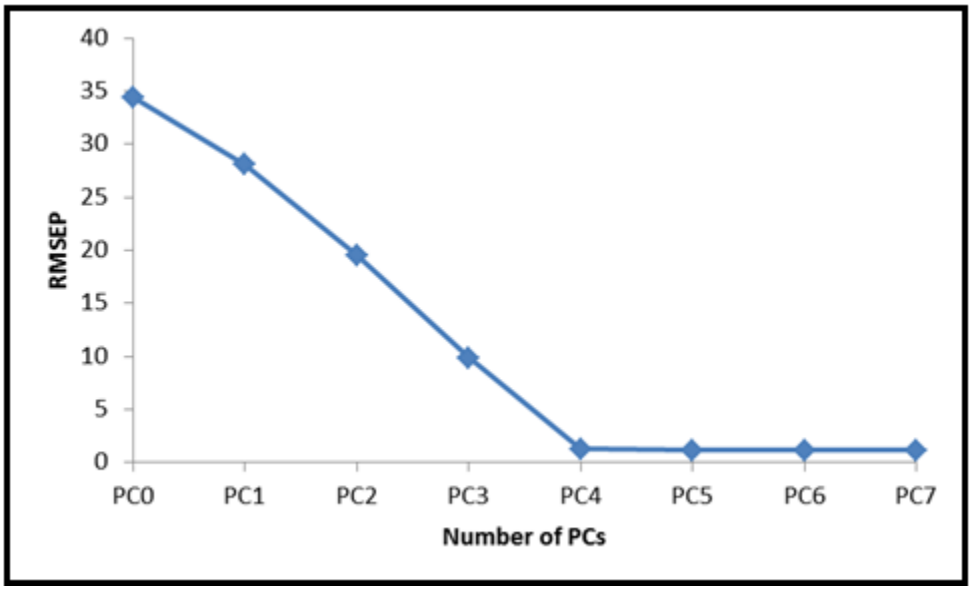

Figure 2-17 RMSEP vs number of P

2.3.2.2.2 Weighted B-coefficients

The weighted B coefficients (Bw_coefficients) are values that demonstrate how independent X variables affect dependent Y variables. The raw B coefficient values should be weighted with autoscaling, which measures the spread of a variable around its mean value. The Bw_coefficients become more comparable with weighting, which means that the higher the Bw_coefficient value, the greater the influence of the predictor variable on the criterion variable. In other words, Bw_coefficients represent the contribution of each X_i variable to the prediction of Y, which is dependent on how each X_i is correlated with a Y variable. These correlations can be positive or negative, and can be used to interpret the results of a PLS analysis. A positive Bw_coefficient of X_i means the higher the X_i, the higher the Y, while a negative Bw_coefficient of X_i means the lower the X_i, the higher Y. The zero value for the Bw_coefficient of X_i means there is no correlation between the variables. Figure 2-18 illustrates an example of Bw_coefficients corresponding to 4PCs, as a function of independent variables (in this example, the Raman shift). The important variables are circled.

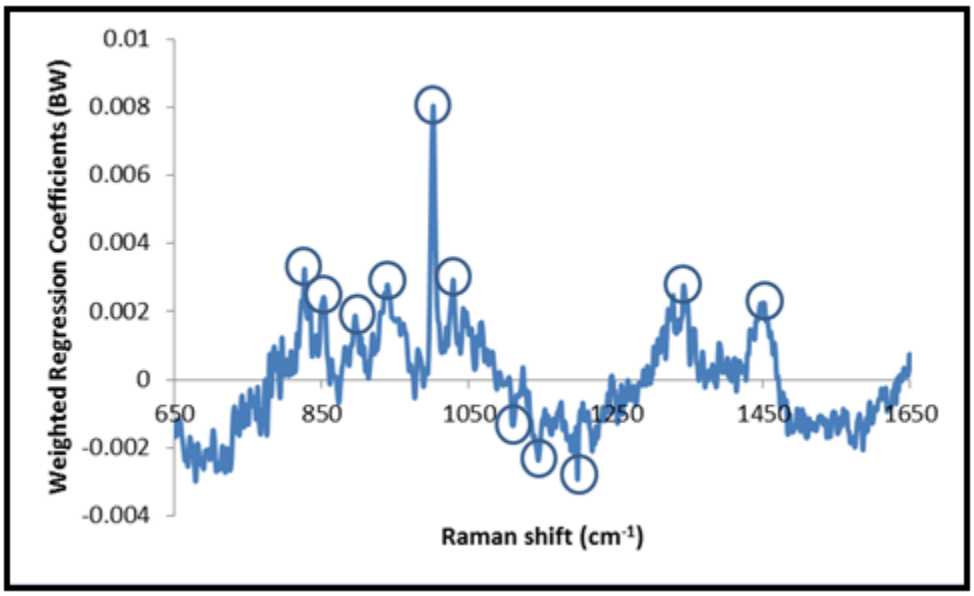

Figure 2-18 Weighted regression coefficients vs independent variables

2.3.2.3 Optimization

A PLS regression model that is to be used as a dataset interpreter or prediction feature must be optimized. Preprocessing is first step of the optimization process, as explained earlier in this chapter. The next step required for the optimization of a PLS model is the outlier removal procedure, which has been applied in this thesis. An outlier is a data point that has no relevant information and differs significantly from the other data points in the dataset [77]. The selection of the informative aspects of a variable's range is another important optimization method that can improve the performance of a constructed PLS model. Outlier removal and variable selection are discussed in the following sections.

2.3.2.3.1 Outliers removal

Outlier detection and removal is an essential aspect of MVA, as it improves the performance of the model. Although outlier removal can be considered a preprocessing step of data processing, here it is explained as an optimization technique implemented after the first model is made. Outlier detection, as a preprocessing method or optimizing technique, should always be applied to improve performance. The Unscrambler® software has a feature that recognizes potential outliers, and it was used frequently in this thesis.

In data processing, it is important to know the source of the outliers, because not all will be illegitimate contaminants [78]. Some researchers believe that if the cause of the outliers is not evident, their removal should not be allowed. However, removing them is strongly recommended in order to achieve a better model [78].

There are many sources that can cause outliers in MVA. Human actions, such as errors in recording the spectra or in the sampling (i.e. when the sample does not support the purpose of the experiment), instrumentation errors, methodological errors (i.e. when one factor or more changes during the experiment) could cause outliers, as well as legitimate cases, the likelihood of which is approximately one percent [79-80].

A simple way to detect outliers is a visual method using a distribution plot. The points in a distribution plot that are away from the main group of data points are identified as outliers, and the score plot, which was explained previously, is used to identify them. For example, in Figure

2-19 three groups of samples are clearly distinguishable, and only one sample away from the groups is identified as an outlier.

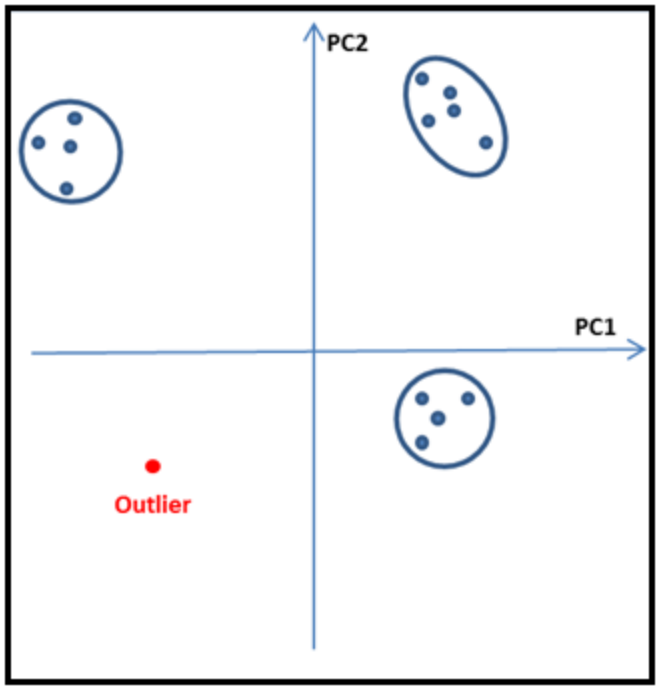

Figure 2-19 Score plot as tool to reveal the outliers

2.3.2.3.2 Variable selection

Another way to optimize a PLS regression model is to manage the range of independent (X-predictor) variables. The model is typically generated over the full range of X-predictors, which usually has both informative intervals and others that are not informative. The non- informative X-predictor ranges are not correlated with the response, and can be ignored to improve the model's prediction performance. Many reports describe the effect that selecting a specific range of X-predictors has on the performance of the PLS model, and reveal how non-informative ranges of variables can affect prediction power [81]. Variable selection is a feature that can reduce overfitting in a PLS model, and improve model performance by using cost effective approaches. However, in some cases variable selection causes greater mathematical complexity.

There are many statistical and data analysis studies that explain different variable selection methods. G. H. John et al. [82] reported that filter and wrapper methods are the main

approaches to variable selection, while Y. Saeys et al. introduced a method of embedded techniques (a combination of these two methods) [83-84]. The simplest technique is the filter method, which is based on introducing a threshold, (e.g. loading weight [85], regression coefficient [86]), and selecting the variables that satisfy the threshold. This method uses the loading weight or regression coefficient of the first model, then removes the variables that reduce them, according to the defined threshold. In other words, the second model with less variables represents higher loading weights or regression coefficients. The Jack-knife method is a filter technique that was coupled with PLS by H. Martens [87], and is typically used in spectroscopy and chemometrics applications [88-90]. With this method, the p-value of a regression coefficient is used as a criterion to select variables. The p-value is the probability of getting a result with more extreme deviation than what was actually recorded. FOCUS [91], Relief [92], decision trees [93] and correlation-based feature selection (CFS) [94] are algorithms that use the filter method to select a useful subset of variables. F. Liu [95] used the filter method, and demonstrated its ability to discriminate between varieties of fruit vinegars with NIR spectroscopy. Although this is a fast method, and has less risk of overfitting compared to other methods, the selected variables are not particularly effective at reducing the limit of detection [96]. Another drawback of the method is that the result depends on choosing an accurate first estimate of the loading weight or regression coefficient, which can be difficult. The second variable selection method is the wrapper technique, which is based on adding iteration algorithms to the filter method. The technique searches the variables and determines those variables that satisfy the threshold, and the search algorithm guarantees adequate error reduction. Genetic algorithm (GA) is a type of wrapper technique that, in combination with PLS, was introduced by K. Hasegawa et al. [97] and improved by R. Leardi et al. [98]. The GA-PLS method was inspired by genetic laws, and has been widely used in data processing to effectively search any large dataset. It uses generic algorithm to find a set of variables that generate an optimized calibration model, based on RMSEP criterion. The main advantages of GA are its capability to optimize continuous and discrete variable problems, to optimize multiple variables, and the fact that it can be run on parallel computers. The method is object-oriented, and can adjust its flexibility to work in diverse applications. The drawbacks of this

method are that it is time consuming, has a low rate of convergence, and a high risk of overfitting [83]. Interval PLS (iPLS), Interactive predictor weighted PLS (IPW-PLS), uninformative variable elimination in PLS (UVE-PLS), sub-window permutation analysis coupled with PLS (SwPA-PLS), covariance procedure (COVPROC) in PLS, regularized elimination procedure in PLS, and backward variable selection for PLS (BVSPLS) are wrapper methods summarized by T. Mehmood et al. [83]. Much research has been done regarding using this technique as a data analysis feature. For example, W. Cai et al. showed that the UVE-PLS wrapper method predicted nicotine content in tobacco samples more accurately [99]. B. Krakowska et al. used this method to prove that genuine diesel fuel samples and their counterfeit variants can be differentiated with less chance of overfitting [100]. Another method of variable selection is the embedded technique, which combines the iteration of variable selection with the iteration of a fitting model. As well, the interaction between variable selection and sample classification takes less time [96]. A. Telaar et al. used this method to detect the discriminating patterns in gene expression, and improved the performance of the statistical model [101]. Like the wrapper approach, this technique can include model construction noise [96].

The backward variable selection method for PLS (BVSPLS) is a type of wrapper method introduced by J. F. Pierna et al. [102]. The main criterion of BVSPLS is to evaluate the model, in terms of root mean square errors of prediction (RMSEP), in iterative steps. The algorithm first includes all the variables when constructing the first model, then the next model is made using all the variable except the first one, then the RMSEP of this model is compared with the RMSEP of the first model. If the RMSEP of the second model is decreased, the first variable could be removed, and if the RMSEP of the second model is increased, the first variable should be involved in the model. This procedure continues until the last variable of the range, at which point the final model has the minimum RMSEP. In this thesis, I further improve the backward variable selection for PLS (BVSPLS) method, which is discussed in Chapter 5.

2.4 Conclusion

Raman is an effective analytical technique that enables us to reveal valuable information about materials, identify different components, and predict quantitative parameters among datasets.

The low intensity Raman signal can be enhanced using the SERS technique, which is based on nanoparticles. The outputs of these techniques include spectra that need to be processed with suitable analytical techniques. MVA, PLS and, particularly, PCA are data processing techniques that classify the samples, create the calibration model, and apply it to predict related information.

Chapter 3. Raman and Surface-Enhanced Raman Spectroscopy of liquid samples using a cuvette

As discussed in Chapter 2, the simplest form of sampling in Raman/SERS experiments is to use a cuvette or capillary as a sample holder. The components of an optical setup should be selected wisely to avoid wasting energy, and though these simple sample holders cannot satisfy this condition, their simplicity encourages many to use them in experiments. This chapter first introduces the details of a Raman setup to detect analytes in a cuvette. We then discuss using the set-up for the detection of a) heparin concentration in serum, b) GLU-GABA concentrations in serum using SERS.

3.1 Introduction

The detection of the Raman signal of an analyte using a cuvette sample holder is much simpler than with other sample holders, such as HC-PCF. However, the detected Raman signal from a cuvette is usually weak, and the peaks reflecting the sample are completely overcome by other spectral in the background. Consequently, determining a concentration from the Raman signal is very difficult.

Despite this, the simplicity of using a cuvette often makes it the preferred technique over other types of in vitro sample holders, such as HC-PCF, when the speed of the experiment is important. The Raman efficiency of using a cuvette as a sample holder is not as high as with an HC-PCF, because when it is illuminated by a laser, the beam is scattered in all directions, making them difficult to collect. Thus, using the cuvette technique as the sample holder in Raman setup can cause more loss of the input beam at the sample than with HC-PCF.

The weak Raman/SERS signal of liquid samples in a cuvette can provide valuable information if the dataset is processed correctly using MVA, such as PLS or PCA. This chapter demonstrates

how to compensate for the weakness of using cuvette as the sample holder by using MVA as a data processing technique. The Raman spectra of heparin serum solutions and the SERS spectra of GLU-GABA solutions were both found using a cuvette. The results of these experiments were used to evaluate the power of MVA, and to precisely determine the concentrations in the sample analytes.

3.2 Raman/SERS setup using cuvette

The Raman setup introduced in Chapter 2 had a number of optical components, including a plano-convex lens, bandpass filter, dichroic filter, two objective lenses and a fiber bundle. These were used in forward scattering and backward scattering arrangements, to determine which was the more efficient Raman setup. The schematic diagrams of the arrangements are shown in Figure 3-1.

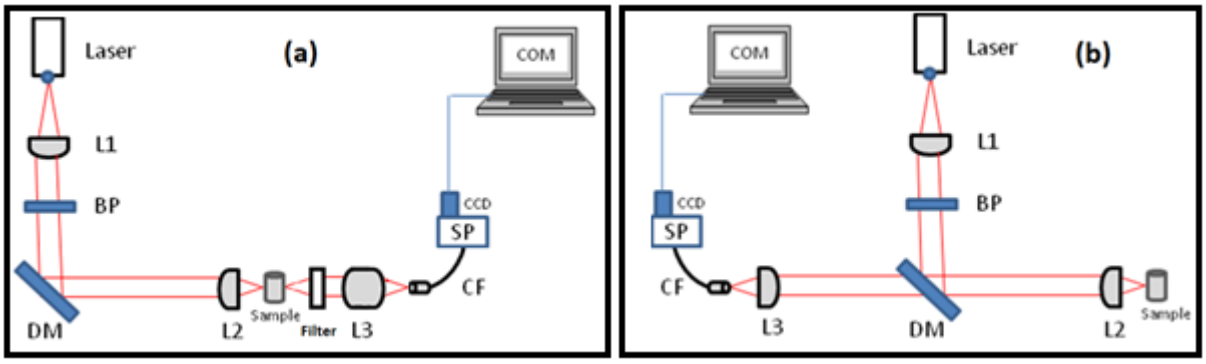

Figure 3-1 The schematic diagram of Raman setup corresponding to
a) forward and b) backward Raman scattering

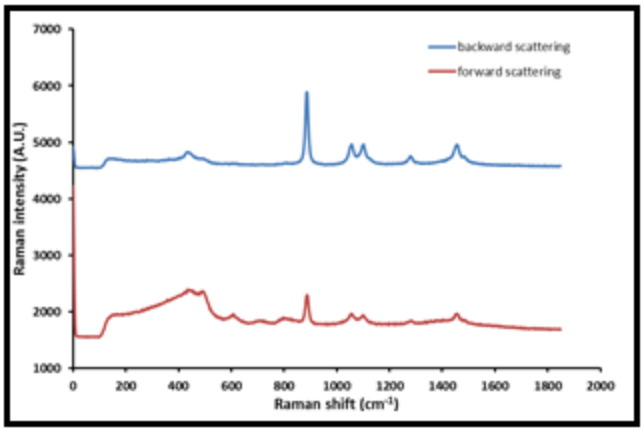

Figure 3-2 Forward and backward Raman spectra of ethanol

The forward and backward Raman spectra of ethanol were used to compare the two setups. Figure 3-2 shows the intensity of Raman peak at 886 cm^{-1} for the backward arrangement, which was approximately 2.5 times higher than the forward arrangement Raman peak. This verifies that the backward arrangement is more efficient than the forward. The comparison of laser intensity of the collecting fiber bundle in these configurations shows that a higher intensity of the laser beam at the collecting fiber bundle diminishes the Raman scattering signal in the forward arrangement at the zero point of the Raman shift, as shown in Figure 3-2.

The more efficient Raman setup also depends on choosing the best optical components in order to obtain the optimal sample spectra. We have used a 785 nm CW multimode diode laser to minimize the interference from the fluorescence background. Although longer wavelength Raman source provide less exciting power at sample (the Raman signal is proportional to λ^{-4}), we can compensate this lower power using proper focusing lens to have more power density at sample. The Numerical Aperture (NA) of a pigtailed diode laser and collecting fiber bundle is 0.22, and using the same NA is an effective way to estimate the focal lengths. The minimum spot size of laser beam at sample is diffraction limited and is proportional to the laser wavelength. A 10X objective lens (L2) with 20 mm focal length can provide 210 μm spot size which enable us to use a 10 mm path length cuvette or capillary as a sample container. To make sure the suitable objective lenses are used, we examined different objective lenses in Raman setup. The Raman intensity of ethanol at 886 cm^{-1}, which is its prominent peak, was chosen to

compare the efficiencies of the Raman setup using different lenses. First, the focal length of the collimating lens (L_1) was set to 10 mm. This provided a 5 mm diameter beam that was efficient enough to focus on the sample in the cuvette. Then other lenses (L_2) and (L_3) were chosen and, as shown in Table 3-1, the highest Raman intensity of ethanol was obtained if the focal lengths of L_1, L_2 and L_3 were 10, 20 and 25 mm respectively.

Table 3-1 Raman intensities of different focal lengths of lenses

Focal length (mm)			Raman intensity
f_1	f_2	f_3	(A.U.)
10	10	20	8328
	10	25	13211
	10	30	8156
	12	20	11552
	12	25	13243
	12	30	13085
	20	25	16940
	20	30	15851

3.3 Raman spectroscopy for clinical-level detection of heparin in serum using partial least squares analysis

3.3.1 Introduction

Heparin is a polysaccharide (complex sugar), and it is considered to be a clinically important blood anticoagulant. It is commonly administered to patient's blood during open heart surgery and kidney dialysis. Though suitable heparin treatment significantly decreases morbidity and mortality, it can also cause hemorrhagic complications from over-anticoagulation or heparin-induced blood disorders. Thus, it is critical for physicians to monitor the amount of heparin in blood accurately and quickly.

The guidelines on heparin monitoring is a complicated document that recommends the heparin doses to be administered to patients for various surgeries, and also discusses heparin monitoring methods, including their merits and demerits [103-105]. The clinical or physiological level of heparin is measured in terms of United States Pharmacopeia (USP) unit. Traditionally,

heparin monitoring is based on functional testing of anticoagulation, such as activated clotting time (ACT) or activated partial thromboplastin time (aPTT) [106-111]. ACT measures the anticoagulation effects of heparin by determining how long it takes for the blood with heparin to clot when induced by activators. The target range for ACT values varies depending on the surgery. For example, it is in the range of 400 to 600s for cardiopulmonary bypass surgery [112]. However, ACT test results are prolonged (up to 15 min) in cases of thrombocytopenia, thrombopathy and hemodilution, and therefore correlate poorly with actual heparin levels [113-114]. Compared to ACT, aPTT is a higher sensitive laboratory technique for monitoring unfractionated heparin, particularly in situations where the patient has coagulation disorders. It measures the time taken (aPTT) for the optical density of blood plasma to reach a specific threshold in the presence of activators. A normal aPTT is in the range of 24 to 37s which is lower than ACT, but the incubation time of activators (~10 min), followed by the addition of reagents (prior to aPTT measurement) makes the process time-consuming and tedious.

ACT or aPTT values correlate poorly with the actual heparin level in blood, and this could cause severe health complications. Therefore, alternative methods for monitoring heparin therapy that directly determine the quantity of heparin in a patient's blood are gaining considerable interest. Existing protocols based on heparin concentration monitoring include protamine sulphate titration, anti-Xa and others [115-118]. Protamine sulphate titration can cause excess post-operative bleeding or platelet activation if there is an overdose of protamine sulphate. Anti-Xa is used exclusively to monitor low-molecular weight heparin by measuring the heparin content indirectly, using the artificial factor X, an enzyme of coagulation, which is inversely related to the heparin activity [118]. The limitation of anti-Xa is that it is an offline method (i.e. laboratory technique), and involves numerous steps which make it very time-consuming.

Each of the above techniques (testing the anticoagulation effect or detecting the heparin concentration), has advantages and disadvantages, and the approximate detection time and accuracy are summarized in Table 3-2. The ideal technique for heparin therapy must be instantaneous, accurate and simple, and minimally affected by the patient's physical conditions or medical history.

Table 3-2 Different techniques of laboratory monitoring heparin

Method	Detection time (min)	Estimated detection accuracy (USP/mL)
ACT	~7-15	0.1
aPTT	~10	0.1
Anti-Xa	~60	0.01
Protamine sulphate titration	~5	0.1

Based on these considerations, we implemented Raman spectroscopy in conjunction with partial least-squares (PLS) analysis to measure the heparin concentration in serum at a clinical level. This analytical technique is useful in quantitative analysis, particularly when the Raman signal is weak or there is an overlap of Raman bands of interest and the sample media (e.g. serum, blood). In this section, we show how PLS analysis can help identify the spectral regions and deduce the sample quantity, as it scans the entire Raman spectrum or the spectral segments that contain Raman bands of interest [55,119-124]. Raman spectroscopy is a novel alternative to measure heparin content, compared to previously described laboratory methods such as fluorescence, surface plasmon resonance, field effect transistor and membrane-based ion-selective electrodes [125-127]. These methods involve indirect detection with heparin probes such as protamine or synthetic cationic polymers. Moreover, they are complicated, and based on either surface affinity capture or automated heparin protamine titration, which limit system sensitivity to detect lower concentrations of heparin in blood. In addition, the accuracy of such methods depends on the cross reaction of heparin with the labeling agent, which can give false results. Previously, Khetani et al. detected heparin quantity by enhancing its Raman signal with a strong light–sample interaction in the HC-PCF [12]. However, they had difficulty maintaining an identical light coupling condition from one sample filled HC-PCF to another. We used Raman (i.e. sample in cuvette) rather than enhanced Raman (i.e. sample in hollow core fiber as in Khetani et al.), and prepared numerous spectral datasets for multivariate analysis. Our standard Raman setup that uses the cuvette as the sample holder is simple, and directly measures heparin concentration up to clinical levels. In most surgeries, the clinical or physiological level of heparin is considered to be less than 10 USP of heparin per milliliter of the patient's blood. Compared to traditional methods, our approach for heparin monitoring is

faster, with a time of approximately one minute acquire spectral data and feed it to the prebuilt MVA model. The accuracy of the PLS model was tested by predicting the heparin concentration in a sample set that was not involved in its construction.

This section is organized in the following manner. We first discuss sample preparation, followed by a description of the quantification procedure for detecting heparin with the PLS regression model. Then we compare the predicted heparin concentration with the measured heparin concentration in a sample set that was not used in the construction of the PLS model.

3.3.2 Experimental details

3.3.2.1 Sample preparation

Blood samples from five cows were obtained from a local bovine slaughterhouse, and the serum was prepared by centrifuging the blood at 4000 rpm for 20 min. The clinical-grade heparin samples were purchased from Pharmaceutical Partners of Canada (PPC Inc.). The sample solutions were prepared by adding different quantities of heparin, in the range of 2 to 25 µL, to a fixed amount of serum (3 mL). To prepare the first set of serum-heparin sample mixtures, we divided the serum of the first cow into 10 equal 3 mL portions. We then added 2.5 µL of heparin to the first sample of the first cow, and 5 µL of heparin to the second sample of first cow and so on. The same procedure was followed for the second cow, while for the other three cows the sample preparation was started by adding 2 µL of heparin to 3 mL of serum with the same interval (2.5 µL); thus, there were 50 samples (5 × 10) overall. The concentration of heparin in the serum was labelled in terms of USP per mL of serum, in accordance with terminology used in clinical environments; USP represents the potency of the drug in clinical applications. The heparin concentrations in the 50 different serum-heparin samples from the blood of the five cows is shown in Table 3-3; it should be noted that 1 µL of heparin has a potency of 10 USP (0.094 mg). The heparin concentration/potency (in terms of its USP value in 1 mL of serum) was calculated from the actual volume of heparin that was added to 3 mL of serum, which is also shown in Table 3-3. The experimental configuration was based on an optimized setup, as explained in Section 3.2.

3.3.2.2 Multivariate Data Analysis

The serum contains several biological components, including albumin, glycoproteins, immunoglobulins and lipoproteins which contribute to the strong spectral (fluorescent) background. The weak Raman signal of heparin was completely overcome by the spectral background of the serum, and no direct correlation between the heparin concentration and its Raman bands could be found. The problem was exacerbated when detecting heparin at the physiological level, as the heparin concentration was below 10 USP/mL. In these circumstances, Raman spectral datasets for serum-heparin mixtures were used to construct a calibration model for PLS analysis using Unscrambler® X version 10.0 software (CAMO, Corvallis, OR). The PLS models were built from the spectral and analytical data. To address the fluorescence effect prior to PLS regression, the spectra were removed from the background with Unscrambler®. The Raman spectral data were normalized using multiplicative scatter correction (MSC), to correct the variability of the baseline data caused by scattering or other physical phenomena.

The calibration model was validated by test set validation (TSV). The spectral data corresponding to four cows (200 spectra), known as the training set/modeling group, were selected to construct the PLS model. The remaining cow's spectral data (50 spectra), also referred as the test group, was used to validate the constructed model. The details of the TSV procedure are further discussed in Section 3.3. The construction of an efficient PLS model involves careful selection of a PC, and the PLS model was evaluated against various statistical parameters, such as RMSEC, RMSEP and R^2. The optimal number of PCs was used in the calibration model.

3.3.3 Results and discussion

3.3.3.1 Raman spectral data

This section primarily focuses on the quantitative measurement of heparin in sample mixtures of heparin and serum. Fifty heparin-serum samples with compositions of heparin in the range of ~6 to 83 USP/mL) were prepared, and named SH1, SH2 . . . SH50, as seen in Table 3-3, and the Raman spectrum of pure clinical-grade heparin is shown in Figure 3-3. The assignment of Raman bands of heparin was reported by Atha et al. [128]. The Raman bands that overlapped

due to symmetric SO_3 vibration were located at ~1035 cm^{-1} (N-SO_3 vibration), 1045 cm^{-1} (6-O-SO_3 vibration) and 1060 cm^{-1} (3-O-SO_3 vibration). The two medium intensity peaks of heparin, at approximately 827 and 893 cm^{-1}, were assigned to the C-H deformation of R, and the α and β anomers of the 2-acetamido-2-deoxy-D-glucose residues along with the presence of low-intensity peak ~1000 cm^{-1} (C-N stretching) [128]. A spectral range of 600 to 1500 cm^{-1} was used for the quantitative analysis of the heparin, as it had the prominent Raman heparin peaks. However, the Raman spectrum of the serum-heparin mixture showed very few Raman heparin peaks at extremely low concentrations of heparin. This is evident in Figure 3-3, which shows how the strong fluorescence background of the serum (A) attenuates the 3200 USP heparin Raman signal (B) and it is expected to completely obscures the weak heparin Raman signal at the physiological level. Figure 3-3 also indicates a decrease in the fluorescence background of serum, and a consecutive increase in the heparin concentration. Moreover, the serum has various intrinsic chemicals that cause Raman peaks that interfere with the heparin Raman peak [122]. Consequently, correlating the Raman bands of heparin and its concentration was nearly impossible by just formulating a simple calibration model. Thus, the PLS models were constructed based on the Raman spectra of serum and heparin mixtures.

Table 3-3 The heparin concentration in serum for 50 sets of sample mixtures

Sample set #1		Sample set #2		Sample set #3		Sample set #4		Sample set #5	
Serum-heparin no.	Concentration (USP/mL)	Serum-heparin no.	Concentration (USP/mL)	Serum-heparin no.	Concentration (USP/mL)	Serum-heparin no.	Concentration (USP/mL)	Serum-heparin no.	Concentration (USP/mL)
SH-1	8.3	SH-11	8.3	SH-21	6.6	SH-31	6.6	SH-41	6.6
SH-2	16.6	SH-12	16.6	SH-22	15	SH-32	15	SH-42	15
SH-3	25	SH-13	25	SH-23	23.3	SH-33	23.3	SH-43	23.3
SH-4	33.3	SH-14	33.3	SH-24	31.6	SH-34	31.6	SH-44	31.6
SH-5	41.6	SH-15	41.6	SH-25	40	SH-35	40	SH-45	40

SH-6	50	SH-16	50	SH-26	48.3	SH-36	48.3	SH-46	48.3
SH-7	58.3	SH-17	58.3	SH-27	56.6	SH-37	56.6	SH-47	56.6
SH-8	66.6	SH-18	66.6	SH-28	65	SH-38	65	SH-48	65
SH-9	75	SH-19	75	SH-29	73.3	SH-39	73.3	SH-49	73.3
SH-10	83.3	SH-20	83.3	SH-30	81.6	SH-40	81.6	SH-50	81.6

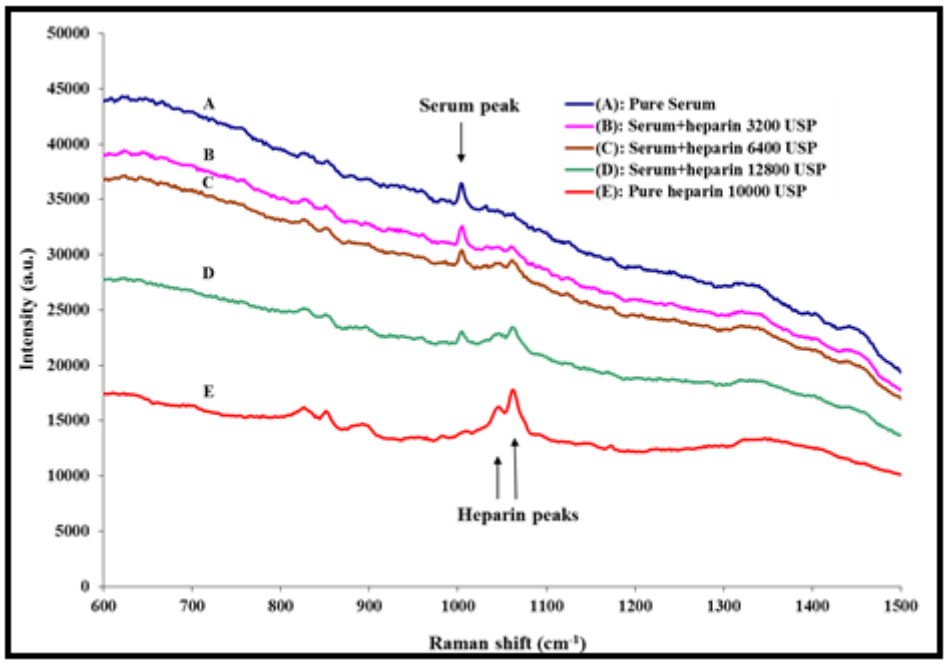

Figure 3-3 Raman spectra of pure serum, pure heparin, and mixtures of heparin and serum

3.3.3.2 Loading and score plots

As it was described in previous chapter, loading plots can be considered bridges between variable space and the principal component space, and they provide a projection view of inter-variable relationships. The loading of the first and second PC are shown in Figure 3-4(a). The prominent Raman bands are approximately 1040 to 1070 cm^{-1}, corresponding to symmetric SO_3 vibration represented in PC1. This is the degree of "systematic variation" in the overall

spectral range, and it forms the structure of the regression model. Another observation is the dip in PC1 near 1000 cm^{-1}, which is due to the decrease in the Raman peak of the serum at that point (Phenylalanine, C-C stretching) while the heparin amount was changed with respect to the serum [122]. The broad band in the PC1 wavenumber range of 1300 to 1450 cm^{-1} could be due to the spectral background of the cuvette. On the other hand, PC2 shows variation of the serum peak ~1000 cm^{-1}. The remaining portion of PC2 describes the "unexplained" component of the model which can be ascribed to random noise."

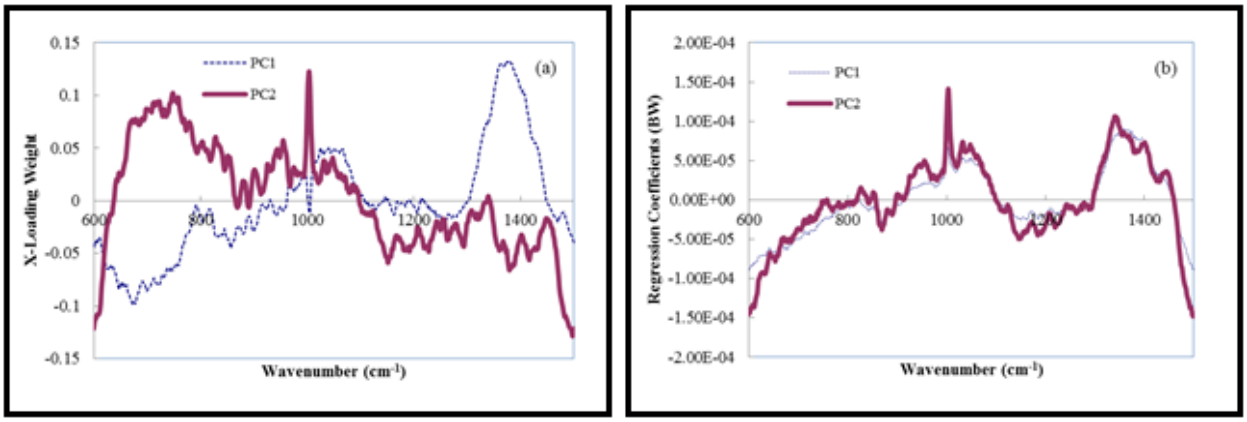

Figure 3-4 (a) Loadings of the first and second principal component of the MSC-corrected spectrum in the range of 600 to 1500 cm^{-1} (b) Regression coefficients of PLS model for PC1 and PC2

The regression coefficients plot in Figure 3-4(b) depicts the most important variables (wavenumbers) in the PLS model, and shows that wavenumbers at approximately 1000, 1035, and 1045 cm^{-1} were used for the quantitative analysis of the model. Similar to loading vectors, score vectors can be plotted against each other, as they are complementary in nature and provide significant information about the object and variables when examined together [55]. Score plots also indicate clustering of variables or the presence of outliers to be eliminated, as in the plots of PC1 and PC2 in Figure 3-5. The first two PCs indicate that 95% (X1 42%, X2 53%) of the X variance explains 59% (Y1 50%, Y2 9%) of the heparin response level. The figure also shows very distinguishable clusters in the samples, which means most of the samples in each cluster are similar. Loading and score plots have significant relevance in this work, as due to an overlap of the Raman band(s) of heparin (solute) and serum (solvent), the spectrum cannot

determine a specific correlation of heparin band intensity at low heparin concentrations. Loading plots, in particular, account for the variation of regression coefficients, and present the actual contribution of heparin to the overlapped Raman bands of heparin and serum.

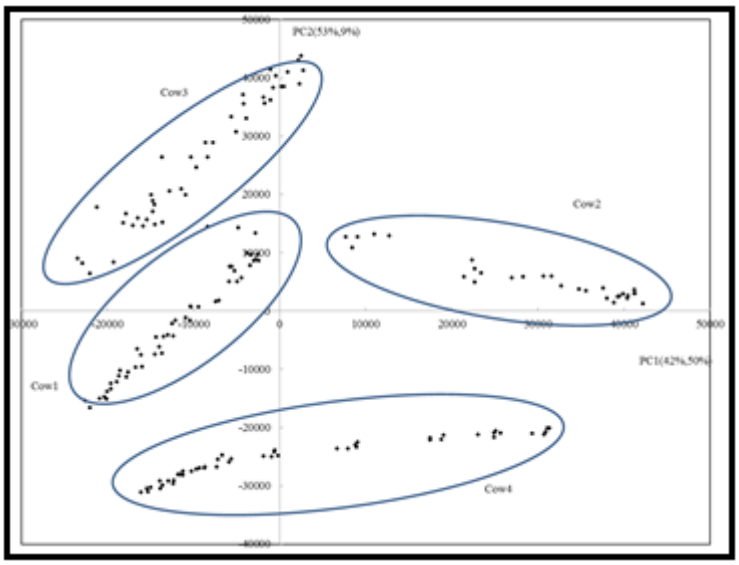

Figure 3-5 Score plot for first and second principal component of the MSC-corrected spectrum in the range of 600 to 1500 cm^{-1}

3.3.3.3 PLS model

The Raman spectra of 50 samples of serum and heparin were recorded, and to ensure consistency in the replicated measurements five Raman spectra were collected for each sample. Thus, the total number of Raman spectra was 250. The PLS models were developed by using the datasets of four cows to construct the model, and the dataset of one cow for independent prediction. For example, Model 1 was developed using the datasets of the second, third, fourth and fifth cow, Model 2 was developed using the datasets of the first, third, fourth, and fifth cow, and so on. A few sample outliers were identified by the Unscrambler® software, and removed from the analysis. Table 3-4 shows the results of the R^2, RMSEC and RMSEP of the five possible models, with and without MSC, for the spectral range of 600 to 1500 cm^{-1}. The table also indicates that MSC has reduced the RMSEP values in all the models. The PLS models were validated based on the TSV method, as it has been established that TSV gives less

prediction errors than full cross validation (FCV) in situations where the sample set is adequately large, which applies in our case [55].

Table 3-4 PLS models of heparin concentrations in serum with TSV

No.	Preprocessing	Test Set Validation			
		R^2 Cal	RMSEC (USP/mL)	RMSEP (USP/mL)	PCs
1	No MSC	0.86	8.74	4.11	3
	MSC	0.91	6.79	2.73	3
2	No MSC	0.90	7.38	9.83	4
	MSC	0.94	5.03	3.82	4
3	No MSC	0.93	6.12	8.75	5
	MSC	0.98	3.20	5.01	5
4	No MSC	0.93	6.28	10.7	5
	MSC	0.97	3.67	4.23	5
5	No MSC	0.93	5.87	6.56	6
	MSC	0.98	2.41	4.19	6

The PLS model obtained from the preprocessed data involved three to six PCs. The optimal number of PCs was determined by assessing the Y-variable residuals versus the PC numbers (not shown here), and determining the values of the PCs with residual variances tending toward zero. The number of PCs in each of the five PLS models was optimized to reduce the RMSEP values, as indicated in Table 3-4. Due to some variations in the blood samples from one cow to another, it was expected that the number of selected PCs would also vary, and the key was to establish higher degrees of prediction accuracy. All of the five PLS models with MSC preprocessing showed high R^2 values (> 0.91) and low RMSEP values (< 5 USP/mL). Table 3-4 shows that the average RMSEP values for all five PLS models with MSC preprocessing was ~4 USP/mL, with less fluctuation between them (standard deviation ~0.82). Thus, it is clear that

the RMSEP values were quite consistent from one model to another, which indicates the prediction accuracy for measuring heparin concentrations in the serum of all five PLS models.

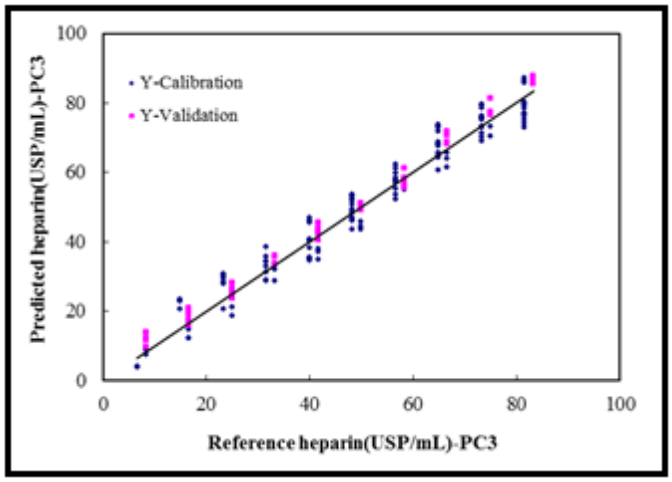

Figure 3-6 PLS regression model for predicting heparin content in serum in 600 to 1500 cm^{-1} spectral range using multiple scattering correction and test set validation

In Figure 3-6, the prediction results and calibration curve of one of the PLS models shows the measured and predicted values of heparin in serum. This model was based on preprocessed data (with MSC) and validated model with the TSV method. According to the calibration curve, the RMSEP error in TSV (in the range of 6 to 84 USP/mL) is about 2.73 USP/mL, which corresponds to ~3.2%. The conclusion is that when the amount of heparin is as low as ~8 USP, it can still be detected with the high accuracy required in a clinical environment.

3.3.3.4 Unknown sample prediction

The guidelines of heparin administration are complex, and depend on the patient's medical condition and other physical attributes (i.e. age, weight, etc.). A reliable and accurate method to monitor heparin must provide consistent results when a different dose is given to a patient. With this in mind, the next phase of research focused on predicting the heparin concentrations in unknown samples that were not involved in the construction of model. This is also referred to as "external validation", where the sample dataset is divided into a training set and a validation set. A comprehensive model that involved blood samples of four different cows (a training set) was constructed, and was then externally validated against the sample data

(validation set) which was set aside during construction. It predicted different amounts of heparin in the range of 8.3 to 83.3 USP⁄mL in a single serum sample. These results, summarized in Table 3-5, show that the PLS model can reliably predict different concentrations of heparin, with deviations in the range of ~2.2 to 3.2 USP⁄mL. The deviation is a function of the model's RMSEP, and confirms that the constructed model has good prediction capability for different concentrations of heparin in a single unknown sample.

We have demonstrated an alternative method of heparin detection, based on Raman spectroscopy and PLS analysis. This method is a direct method that does not need any activator and takes less than a minute (without including data processing time) and needs less than 2 mL samples. To improve its use for rapid monitoring of heparin in a surgical environment, the process of both spectral data acquisition and subsequent loading into the prebuilt calibrated model, could be fully automated. In addition, the time spent centrifuging the blood in this study could be further shortened by subjecting the blood to a "composite media" filter. In this case, the serum sample can be obtained within a few seconds, instead of the 15 to 20 minutes it takes a centrifuge to separate serum from blood.

Table 3-5 The prediction of different heparin concentrations in serum for one unknown sample

Measured heparin (USP/mL)	Predicted heparin (USP/mL)	Deviation (USP/mL)
8.3	9.7	2.2
16.6	15.8	2.3
25	25.3	2.3
33.3	33.6	2.4
41.6	40.6	2.6
50	50.2	2.7
58.3	58.3	2.8
66.6	68.6	3.1
75	76.4	3.2
83.3	86.4	3.1

3.4 Surface Enhanced Raman Scattering (SERS) spectroscopy for detection of glutamate and γ-aminobutyric acid in serum by partial least squares analysis

3.4.1 Introduction

Glutamate (GLU) and γ-aminobutyric acid (GABA) are the most prominent amino acid neurotransmitters in the central nervous system (CNS). GLU is enzymatically converted to GABA by the GLU decarboxylases [129]. The two amino acids, specifically GLU [130] and GABA [131], activate the receptor families. GLU is considered a major excitatory neurotransmitter, while GABA is predominantly inhibitory. Thus, both GLU and GABA play major roles in the CNS, and their imbalance can trigger a variety of neurological disorders, including Alzheimer's disease [132], Parkinson's disease (PD) [133], and others [134]. Cerebrospinal fluid (CSF) concentrations of GLU and GABA are also related to responses to phenobarbital treatment in primary epilepsy [135] and the levels of GLU in CSF are significantly correlated with ischemic events after subarachnoid hemorrhage [136]. The concentrations of different neurotransmitter amino acids, including GLU and GABA, are simultaneously measured in patients suffering from partial sensory deprivation [137]. Therefore, simultaneous measurement of GLU and GABA is critical for accurate diagnosis of neurological disorders, and potentially for developing novel neuropharmacological agents [138].

The best known technique to measure GLU and GABA is high performance liquid chromatography (HPLC) [139] combined with fluorescence (HPLC–FD) [140] or electrochemical detection (HPLC–ECD) [141], even though HPLC accessories are costly and the procedure is complex [142]. In addition, Liquid chromatography/mass spectrometry (LC–MS) and liquid chromatography/tandem mass spectrometry (LC–MS/MS) methods have been developed for the analysis of GLU and GABA in biological samples, such as human plasma and cerebrospinal fluid [143]. The LC–MS/MS method is very sensitive and can easily analyze non-volatile samples, but the complexity of its use and high cost are drawbacks.

The gas chromatography (GC) method with a mass spectrometry detector (GC/MSD), an electron capture detector (GC/ECD), a flame ionization detector (GC/FID), a thermionic emission detector (GC/TED) or a Fourier transform infrared detector (GC/FT-IR) are other

techniques used to monitor GLU-GABA [144]. These are very time-consuming, which can introduce variations due to reaction times and temperature [145]. Capillary electrophoresis-laser-induced fluorescence (CELIF) has been proposed as a new method for monitoring GABA [146], but it requires derivatization of the amino acids which increases the cost [147].

SERS has been used to monitor the concentration of various neurotransmitters in bulk solutions, including GLU, GABA, dopamine and norepinephrine [148-153]. Although SERS has been used independently for quantitative measurement of GLU and GABA [154], there are no reports of using it for simultaneous measurement of GLU and GABA. This is largely due to the challenges in resolving the Raman spectral features and the overlapped Raman/SERS bands of GLU and GABA, an obvious consequence of the similarity in their molecular structures. Moreover, most SERS studies have been conducted in aqueous solutions of either GLU or GABA, rather than in cerebrospinal fluid or blood serum. Hence, there is a definite need for cost-effective methodologies that can provide accurate and fast determination of GLU and GABA. Considering this, we used SERS in conjunction with PLS analysis to detect a mixture of GLU and GABA dissolved in water and serum.

In this section, we begin by explaining the sample preparation, then discuss the quantification procedure for detecting GLU and GABA in de-ionized (DI) water and serum, using the PLS regression model. Finally, to prove the accuracy of our method we compare the predicted GLU-GABA concentrations with the measured concentrations in a sample set that was not used in the construction of the PLS model.

3.4.2 Experimental details

3.4.2.1 Nanoparticle synthesis and sample preparation

The role of nanoparticles is critical to enhance the weak Raman signal of the molecules to be interrogated, particularly when the molecule concentration is extremely low. Colloidal nanoparticles were an obvious choice for our study, as small molecules like GLU and GABA are easily exposed in the electric field around the surface of aggregated nanoparticles, which amplifies their Raman signals.

GLU, GABA, hydroxylamine hydrochloride, sodium hydroxide and silver nitrate were purchased from Sigma Aldrich, and blood samples were obtained from a local bovine slaughterhouse. Ultra filtered serum (UFS) was prepared by centrifuging the blood at 4000 rpm for 20 minutes. The silver nanoparticles were prepared according to the method described by Leopold et al. [155]: 4.5 mL of sodium hydroxide solution (0.1 M) was added to 5 mL of hydroxylamine hydrochloride solution (0.06 M). The role of hydroxylamine hydrochloride is to reduce the silver nitrate, while the sodium hydroxide controls the size and dispersion of the produced collides. The mixture was rapidly added to 90 mL of silver nitrate solution (0.001 M) and shaken for a few seconds, which produced a milky-gray solution. The UV-vis. absorption/extinction spectra and transmission electron microscopy (TEM) images of two different batches of silver nanoparticles are shown in Figure 3-7. The UV-vis. absorption spectra of the two batches are quite similar with minimal shift in the absorbance peak, as also shown in Figure 3-7. This further suggests that the size/shape of nanoparticles from one synthesis to another are similar and reproducible with limited clustering/aggregation. Moreover, all spectral measurements were performed under identical experimental conditions, to achieve high quality, reproducible spectra.

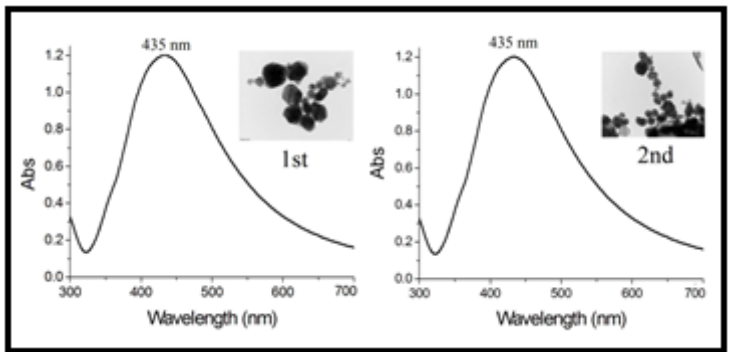

Figure 3-7 The UV-vis absorption spectrum of two batches of silver nanoparticles. Inset shows the TEM image of the nanoparticles

The sample mixtures of GLU and GABA were divided into three groups. The first and second groups had different sample mixtures of GLU-GABA in DI water, and the third group had GLU-GABA in blood serum. The first group was comprised of eight sample mixtures of GLU, GABA and nanoparticles in DI water in the millimolar range. The concentration of GLU in four of these samples was constant while the concentration of GABA varied, and the concentration of GABA

in the remaining four samples was constant while the concentration of GLU varied. We added 400 µL of nanoparticles to the eight different samples, as shown in Table 3-6. The second group of samples consisted of five mixtures of GLU-GABA in DI water in the micromolar range, as seen in Table 3-7. And the third group was prepared in the same way as the second group, but using cow serum. The third group consisted of the serum of five cows (A, B, C, D and E), and five different samples of GLU and GABA concentrations were prepared for each cow. The samples in each of these sets were labelled as A1-A5, B1-B5, C1-C5, D1- D5 and E1-E5. We then added 400 µL of prepared silver nanoparticles to each sample in the group. The experimental configuration was the same as in Section 3.2.

Table 3-6 GLU and GABA concentrations in eight different samples

Sample no.	GABA concentration (mM)	GLU concentration (mM)
1	0	4.7
2	0.47	4.7
3	0.94	4.7
4	1.9	4.7
5	4.7	1.9
6	4.7	0.94
7	4.7	0.47
8	4.7	0

Table 3-7 GLU and GABA concentrations in five different samples

Sample no.	GABA concentration (µM)	GLU concentration (µM)
1	10	90
2	28.6	71.4
3	50	50
4	66.7	33.3
5	87.5	12.5

3.4.2.2 Multivariate data analysis

Due to the similar molecular structures of GLU and GABA, the majority of their SERS bands overlap and, consequently, no direct relationship between their concentrations and SERS bands could be determined. Thus, SERS spectral datasets of GLU-GABA mixtures were used to create two calibration models for PLS analysis with Unscrambler® X version 10.0 software (CAMO, Corvallis, OR, USA). The PLS models were constructed with the same procedure as in Section 3.3, using the MSC technique and TSV method to obtain an efficient calibration model that can be evaluated against R^2, RMSEC and RMSEP.

3.4.3 Results and discussion

3.4.3.1 GLU-GABA mixture in DI water

3.4.3.1.1 Raman spectral data

First, the SERS spectra of GLU and GABA mixtures in DI water were recorded, mainly to identify the unique SERS peaks of GLU and GABA. Since water is a simple matrix compared to serum, well-resolved corresponding GLU and GABA peaks were obtained as there was virtually no interference from the Raman water peaks. For group 1, eight different mixtures of GLU, GABA and nanoparticles in the millimolar range were prepared, as shown in Table 3-6, and the amount of nanoparticles in each of eight samples was 400 µL. The SERS spectra of eight sets of sample mixtures with relatively different concentrations of GLU and GABA, in the range of 500 to 1600 cm^{-1}, are shown in Figure 2. According to the figure, the SERS peaks of 832 cm^{-1} (due to the contribution of deformation modes of C-O and N-H), 907 cm^{-1} (C-C-N band), 982 cm^{-1} (due to the completely ionized form of GLU), 1114 cm^{-1} (CH_2 band), 1295 cm^{-1} (CH_2 band), and 1356 cm^{-1} (CH_2 band) are GLU peaks, and 807 cm^{-1} (CH_2 groups or NH_2 band), 858 cm^{-1} (deformation mode of amino groups), 895 cm^{-1} (CH_2 band), 1057 cm^{-1} (C-N band), 1106 cm^{-1} (NH_2 band), 1217 cm^{-1} (CH_2 band or NH_2 band), 1235 cm^{-1} (CH_2 band or NH_2 band), 1332 cm^{-1} (CH_2 band) and 1441 cm^{-1} (CH_2 band) are GABA peaks. Other peaks, including 774 cm^{-1} (COO^- band), 934 cm^{-1} (completely ionized form of GLU, stretching mode of C-COO^- band), 1036 cm^{-1} (C-N, CH_2 bands) and 1386 cm^{-1} (corresponded to the symmetrical stretching of the COO^- group) are common SERS peaks of both GLU and GABA [151-152], [156-158].

Figure 3-8 shows the SERS spectra of aqueous mixtures of GLU and GABA with relatively different concentration ratios, and that the SERS peak intensity of GLU varies significantly compared to that of GABA, with respective variations of their amounts in the sample mixtures. This is because the GLU molecule has one more carboxylic group than GABA, which results in higher Raman modes of vibration. It should be noted that prominent SERS peaks near 832 cm^{-1} and 858 cm^{-1} are unique to GLU and GABA, respectively, and they can be distinguished from each other.

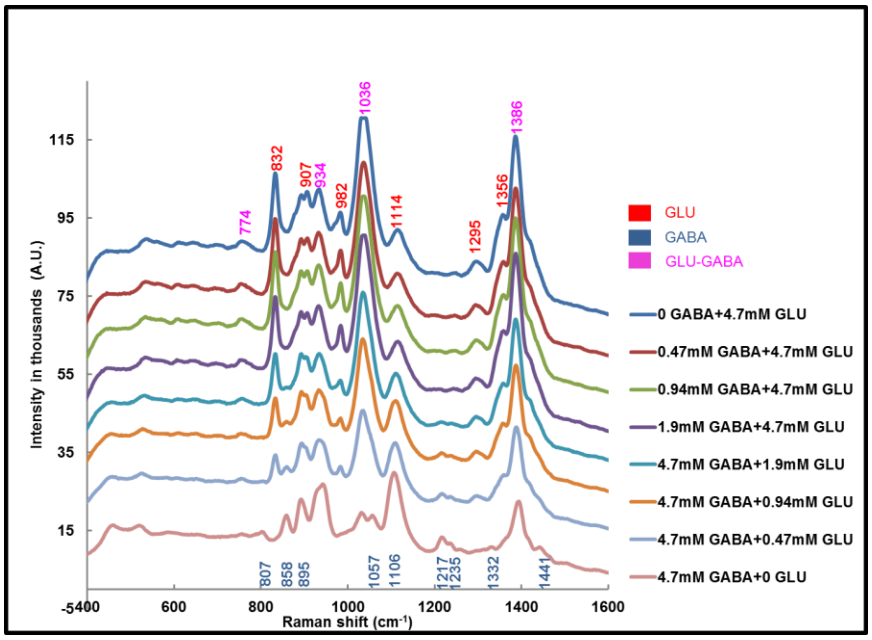

Figure 3-8 SERS spectra of GLU-GABA mixture in DI water

There was good correspondence between the SERS bands of GLU and GABA in DI water, as well as in the millimolar range and the concentrations. Thus, the evaluation of this linear association can be performed with a simple calibration model, as reported by V. Tiwari et al. [154]. The second group of samples consisted of five mixtures of GLU-GABA in DI water in the micromolar range, as shown in Table 3-7. A more precise calibration based on the PLS model was applied for this concentration.

3.4.3.1.2 Loading and score plots

The loading of the first and second PC are shown in Figure 3-9(a). The peaks at 614 cm^{-1} (COO^- band), 844 cm^{-1} (C-O and N-H), 934 cm^{-1}, 1036 cm^{-1}, 1130 cm^{-1} (NH_2 or NH_3^+ band) and 1386 cm^{-1} can be assigned to both GLU and GABA. The peaks at 684 cm^{-1} (O-Ag-O), 832 cm^{-1}, 874 cm^{-1} (CH_2 band), 993 cm^{-1} (completely ionized form of GLU), 1093 cm^{-1} (CH_2 band), 1162 cm^{-1} (CH_2 band), 1186 cm^{-1} (CH_2 band), 1255 cm^{-1} (CH_2 band), 1295 cm^{-1} and 1356 cm^{-1} are due to GLU only, and the peaks at 807 cm^{-1}, 893 cm^{-1}, 954 cm^{-1}, 1235 cm^{-1} and 1441 cm^{-1}, are due to GABA [151-152], [156-158]. Consequently, wavenumbers 684, 807, 832, 874, 893, 954, 993, 1093, 1162, 1186, 1235, 1255, 1295, 1356 and 1441 cm^{-1} can discriminate between GLU and GABA in the mixture. The regression coefficients plot is another feature that can highlight important wavenumbers in this analysis. However, it is not shown here because it is similar to the loading plot, and it provides virtually no new information.

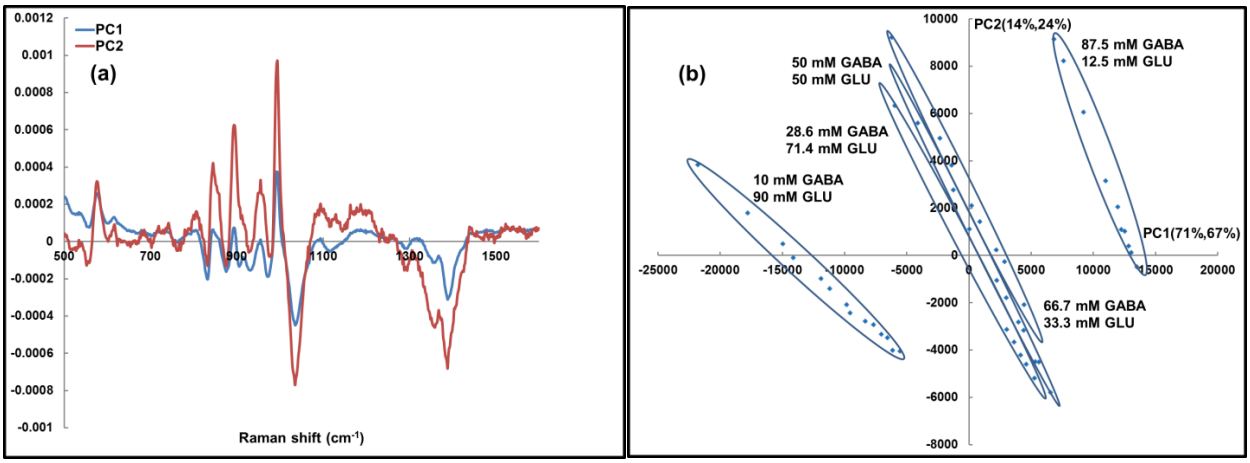

Figure 3-9 Loading plot (a) and score plot (b) of principal component of mixture in DI water in the range of 500 to 1600 cm^{-1}

As shown in Figure 3-9(b), the score plot of PC1 and PC2 is shown in Fig. 3-9(b) to find clusters of variables. This score plot show that 85% (X1 71%, X2 14%) of the X variance explains 91% (Y1 67%, Y2 24%) of the response mixture concentration. The figure reveals highly distinguishable clusters in the samples, which means that most of the samples in each cluster are similar, and different concentrations are distributed in different clusters. Loading and score plots have significant relevance in this work.

3.4.3.1.3 PLS model

The SERS spectra of five samples of GLU-GABA mixtures in DI water were recorded, and twenty Raman spectra were collected for each sample to ensure consistency of the replicated measurements. Thus, the total number of Raman spectra was 100. The PLS model was developed using two thirds of the dataset as a calibration set for constructing the model, and the remaining one third as a validation set for model evaluation. Approximately 15% of recorded spectra were identified as outliers by the Unscrambler® software, and removed from the analysis. According to this model, the R^2 (for calibration and validation) RMSEC and RMSEP for the spectral range of 500 to 1600 cm^{-1} are 0.99, 0.99, 1.2 and 1.4, respectively. All calculations are based on the TSV model using six PCs, which reduced the RMSEP value compared to the FCV model (not shown here). The result of the TSV model predictions are summarized in Table 3-8, and they indicate that the PLS model with the TSV validation method can reliably predict different concentrations of GLU-GABA in DI water, with deviation in the range of 0.9 to 1.3 μM. The deviation is an estimated uncertainty of each sample prediction based on a constructed calibration model, which was calculated by the Unscrambler® software we used.

Table 3-8 GLU and GABA concentrations prediction in 5 different samples (DI water mixture)

Sample no.	GABA concentration (μM) in DI water		GLU concentration (μM) in DI water	
	Prediction	Deviation	Prediction	Deviation
1	9.8	1.0	90.4	0.9
2	28.2	1.3	71.6	1.2
3	50.9	1.2	50.6	1.2
4	66.6	1.3	32.0	1.2
5	87.6	1.2	11.9	1.1

3.4.3.2 GLU-GABA mixture in serum

3.4.3.2.1 SERS spectral data

The ultimate goal of this study was to simultaneously measure the clinical-level concentration of GLU and GABA (less than 10 µM) in serum by SERS spectroscopy. The role of nanoparticles in enhancing weak Raman signals is important to determining the detection limit of GLU and GABA. It is worth mentioning that the signal enhancement factor calculated in this study involves only the ratio of enhanced Raman peak intensity and normal Raman peak intensity. It does not consider the ratio of the number of molecules sampled in bulk and those that are adsorbed on the nanoparticles surface, something that is usually done to evaluate Raman enhancement factors (also termed G factor). The SERS peaks of GLU (0.1M) in serum and GABA (0.1M) in serum were compared with the Raman peaks of those solutions (not shown here), and the Raman signal enhancement factor of silver nanoparticles was approximately 10. Figure 3-10 shows the SERS spectrum of different concentrations of pure GLU and pure GABA in serum. The SERS intensity of the GLU-GABA mixture in serum is less than the SERS intensities in DI water, as shown in Figure 3-11. This is because the optical activity of the nanoparticles was affected when the sampling matrix was changed from water to serum, resulting in a reduction of the Raman signal enhancing ability of the nanoparticles. Further, serum is much more complex than water, and it contains a variety of biomolecules which create a strong fluorescence background that overwhelms the SERS signal of the molecules under study (GLU and GABA). Consequently, correlating the Raman bands of GLU and GABA with their concentrations using a simple calibration model was virtually impossible, which is why the PLS models were constructed based on the Raman spectra of GLU and GABA in serum mixtures.

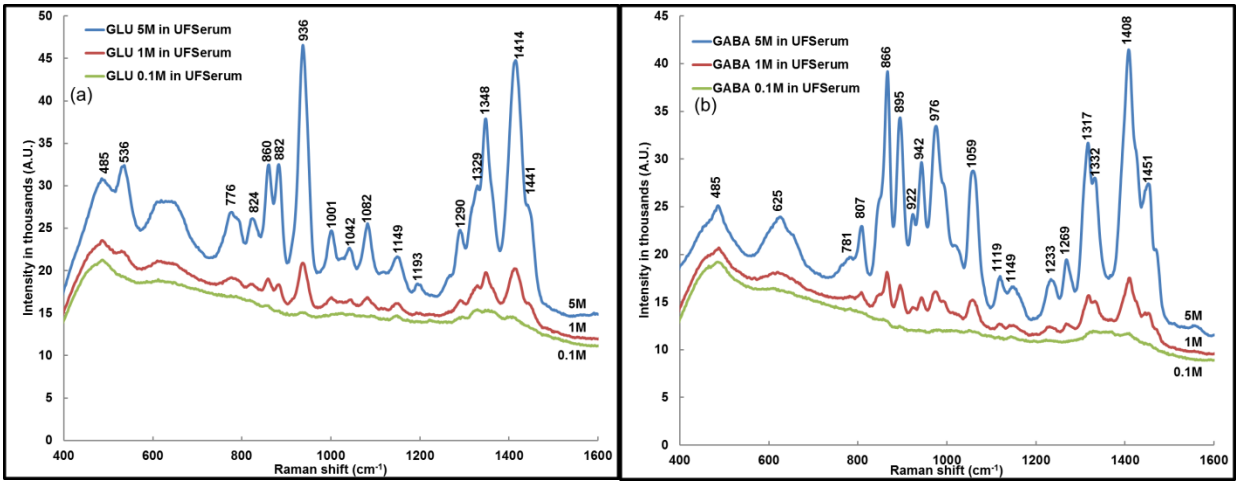

Figure 3-10 Raman spectra of GABA (a) and GLU (b) in serum

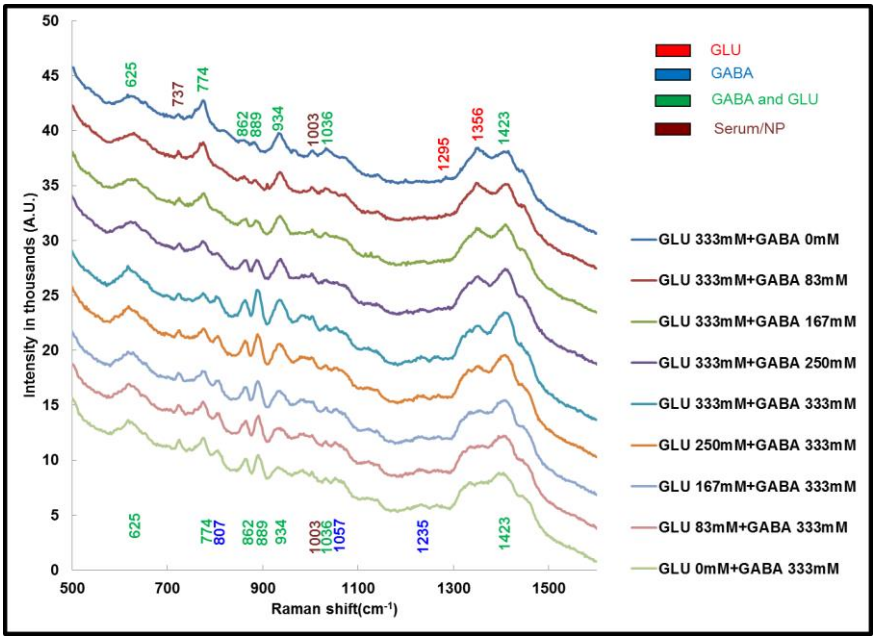

Figure 3-11 SERS spectra of GLU-GABA mixture in serum

3.4.3.2.2 Loading and score plots

The loadings of the GLU-GABA mixture in serum of the first and fourth PC are shown in Figure 3-12(a). In the loading plot the majority of peaks, including 646 cm^{-1} (COO^- band), 774 cm^{-1}, 791 cm^{-1} (NH_2 band), 844 cm^{-1}, 934 cm^{-1}, 1036 cm^{-1}, 1130 cm^{-1}, 1145 cm^{-1} (CH_2 band), 1323 cm^{-1} (CH_2 band) and 1386 cm^{-1}) show the contributions of GLU and GABA in the mixture. The peaks

at 954 cm^{-1} (COO^- band), 968 cm^{-1} (skeletal stretch mode), 1235 cm^{-1}, 1271 cm^{-1} and 1441 cm^{-1} show only the GABA contribution, while the peaks at 553 cm^{-1} (HOCC band), 832 cm^{-1}, 993 cm^{-1}, 1093 cm^{-1}, 1162 cm^{-1}, 1186 cm^{-1}, 1255 cm^{-1}, 1295 cm^{-1} and 1356 cm^{-1} show only the GLU contribution [151-152], [156-158]. The peaks at 710 cm^{-1} and 737 cm^{-1} are due to the presence of either nanoparticles, proteins or carbohydrates in the serum [159]. Moreover, the loadings of GLU-GABA mixture in serum of the fifth and sixth PC (not shown here) have a peak at 807 cm^{-1}, which is considered to be due to the GABA in the mixture.

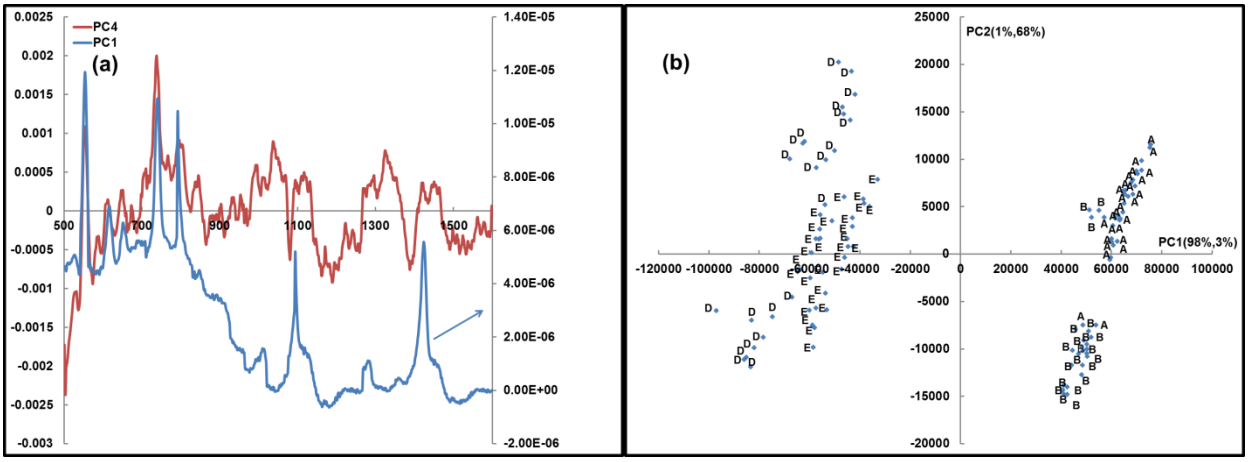

Figure 3-12 Loading plot (a) and score plot (b) of principal component of mixture in serum in the range of 500 to 1600 cm^{-1}

Thus, the wave numbers 553, 807, 832, 954, 968, 993, 1093, 1162, 1186, 1235, 1255, 1271, 1295, 1356 and 1441 cm^{-1} represent the GLU-GABA mixture in serum. It should be noted that some of discriminative wave numbers in water and serum are different, due to several biological components of serum compared to water. The complexity of serum causes almost 176 times lower Raman peak intensity (at 934 cm^{-1}) of GLU-GABA in serum (Figure 3-11) compared to GLU-GABA in water (Figure 3-8).

Similar to the DI water mixture, the score plot provides information about the relationship between different sets of the A, B, D and E samples (set C was set aside for independent prediction). The score plot of PC1 and PC2 is shown in Figure 3-12(b), and it reveals that 99% (X1 98%, X2 1%) of the X variance explains 71% (Y1 3%, Y2 68%) of the response mixture concentration. The percentage of response mixture concentration in serum (71%) is less than in

water (91%), due to the complexity of the serum. Although the clustering in the samples was not as clear as in the DI water case, As shown in Figure 3-12(b), the clusters of serum samples A and B are still distinguishable, while serum samples D are very close to those of samples E.

3.4.3.2.3 PLS model

There were five groups of samples (A, B, C, D and E), and each group was classified with a different serum. Every group consisted of five samples of GLU and GABA with relatively different concentration ratios. Thus, the total number of samples was 25 (5*5). The Raman spectra of the 25 samples of GLU and GABA mixture in serum were recorded, and 20 Raman spectra were collected for each sample to ensure consistency in the replicated measurements. Thus, the total number of Raman spectra was 500.

The PLS model was constructed using four sample sets of serum, while the fifth sample was set aside for independent predictions. For example, Model 1 was developed using the datasets of A, B, C and D, Model 2 was developed using the datasets of A, B, C and E, and so on. Approximately 9% of the recorded spectra were identified as outliers by the Unscrambler® software, and removed from the analysis. Table 3-9 shows the R^2, RMSEC and RMSEP results of each of the five possible PLS models, in the spectral range of 500 to 1600 cm^{-1}. The models were validated based on the TSV method, which has fewer prediction errors than the FCV method when the sample set is bigger than fifty.

Table 3-9 Five possible PLS models of GLU-GABA mixture in serum with TSV

Model no.	R^2_C	RMSEC (μM)	RMSEP (μM)	PCs
1	0.98	2.93	2.71	5
2	0.99	2.34	2.74	6
3	0.99	2.34	2.47	6
4	0.99	1.08	2.85	6

5	0.98	2.99	2.62	6

The PLS model obtained from preprocessed data involved five or six PCs. The optimal number of PCs was established by evaluating Y-variable residuals versus PC numbers (not shown here), and determining the values of PCs with residual variance tending to zero. As shown in Table 3-9, the number of PCs was optimized to reduce RMSEP values. All five possible models showed high R^2 (>0.98) and low RMSEP (from 2.5 to 2.8 µM), with an average RMSEP value of 2.7 µM and a standard deviation of variations of 0.11. The higher errors for RMSEC and RMSEP in serum compared to water were predictable, due to the lower percentage of response in serum. Regardless, the ability of this model to recognize the discriminative wavenumbers allowed us to accurately predict the different concentrations. The prediction results of the PLS model are shown in Figure 3-13, where the calibration curve indicates the measured and predicted values of GLU and GABA in the mixture. The model was based on preprocessed data and validated with the TSV method, and according to the calibration curve the RMSEP error in TSV (in the range of 10 to 90 µM) was approximately 2.5 µM, which corresponds to ~3%. The general conclusion is that the amounts of GLU and GABA are low and ~8 µM can be detected with error down to 2.7 µM, as required in a clinical environment.

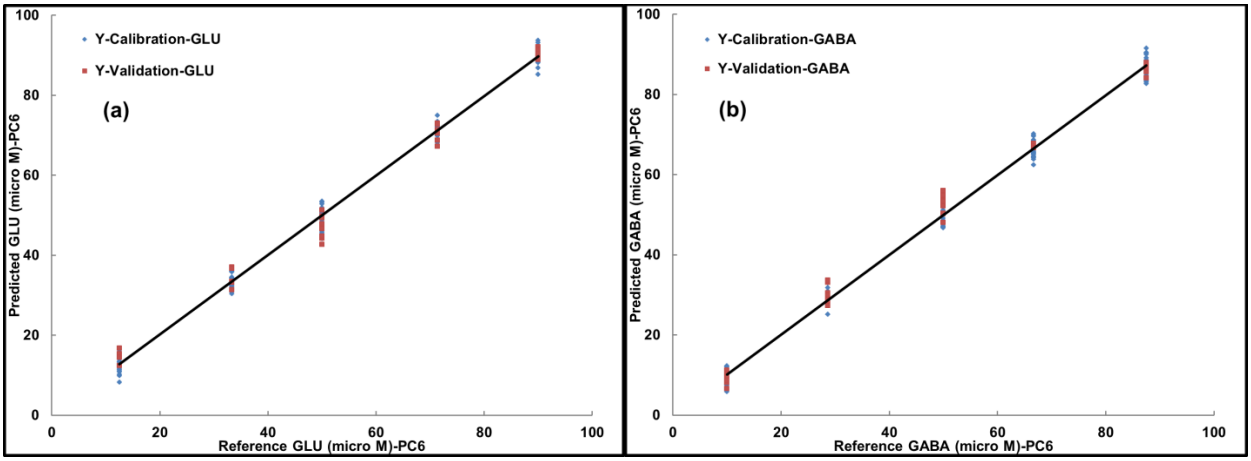

Figure 3-13 PLS regression model for predicting GLU-GABA content in serum in 500 to 1600 cm^{-1} spectral range using test set validation

3.4.3.2.4 Repeatability of measurements

The spectra of groups A, B, C, and D were recorded the first day, and those of group E were recorded the second day. The average SERS intensity of groups A, B, C and D were compared with the SERS intensity of group E by considering one of the SERS peaks of related spectra. For example, at 934 cm^{-1} the average SERS intensities of all recorded spectra of samples #1 (GABA_10 µM + GLU_90 µM) of groups A, B, C and D were calculated, then compared with the intensity of samples #1 of group E. The p-values, which give the probability of a true difference between groups of samples and repeatability, were calculated from analysis of variance (ANOVA) using MS Excel; the p-value and repeatability of samples #1 to samples #5 are shown in Table 3-10. A p-value less than 0.05 indicates a high repeatable measurement. The repeatability of these samples was between 0.69 and 0.89, which means that SERS intensities of identical concentrations of GLU and GABA in different serum (sample matrix) solutions (i.e. A to E) were highly repeatable (>0.7). Figure 3-14 shows the spectra of 20 raw spectra of sample #1 in DI water and serum. It demonstrates the reproducibility of recorded data, and compares the spectrum of samples in DI water and serum samples.

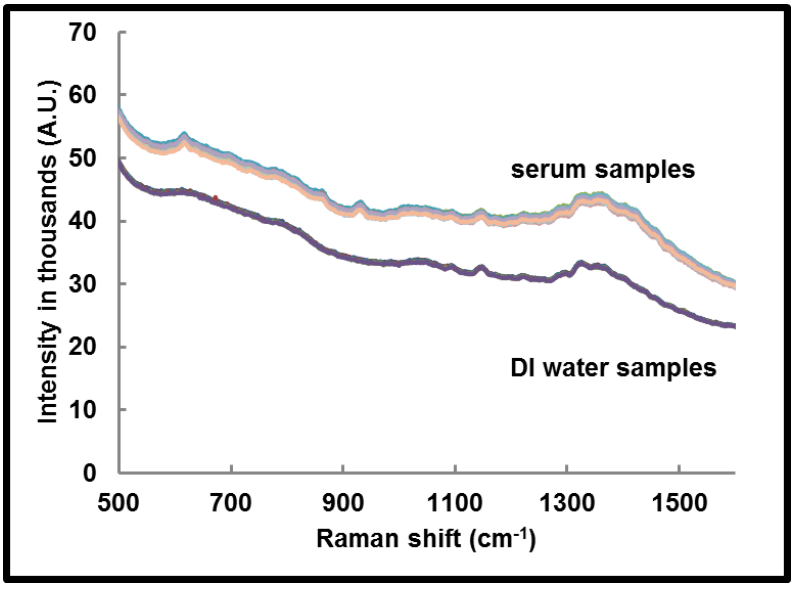

Figure 3-14 Twenty collected spectra of sample #1 in DI water and serum

Table 3-10 Repeatability of SERS intensity of different groups at wave number 934 cm^{-1}

Sample no.	Average of SERS intensities of sets A,B,C,D (day1) at 934 cm^{-1} (a.u.)	SERS intensity of set E(day2) at 934 cm^{-1} (a.u.)	P_Value	Repeatability	Standard deviation
1	1456	1569	0.023	0.74	0.077
2	2777	2897	0.021	0.75	
3	3261	3325	0.038	0.69	
4	3582	3408	0.010	0.89	
5	4128	4299	0.010	0.81	

3.4.3.2.5 Predicting GLU and GABA concentrations in unknown sample

The third group of samples in serum, PLS Model 3, was comprised of sample sets A, B, D and E, and was evaluated by external validation by predicting a set of unknown serum samples (sample set C). The model was validated against the sample set data (C), which was then set aside during model construction. It predicted different concentrations of GLU and GABA in serum, in the range of 10 to 90 µM. The results, summarized in Table 3-11, show that PLS Model 3 can reliably predict different concentrations of GLU-GABA in serum, with deviation in the range of ~2.2 to 2.7 µM.

Table 3-11 GLU and GABA concentrations prediction in five different samples (serum mixture)

Sample no.	GABA concentration (µM) in serum		GLU concentration (µM) in serum	
	Prediction	Deviation	Prediction	Deviation
1	8.4	2.4	93.5	2.6
2	27.2	2.2	73.6	2.5
3	48.0	2.4	46.5	2.7
4	62.8	2.5	35.6	2.6
5	85.9	2.4	15.4	2.7

Our standard procedure was to take 20 spectra of each sample and apply PLS to predict the sample concentration with specific accuracy. Then, to further establish the accuracy, we averaged the spectral data points of all the samples of a particular concentration (samples #1) from all the groups (A, B, C, D and E). The same procedure of averaging data points corresponding to samples of other concentrations (i.e. samples #2, #3, #4 and #5) from each group (i.e. A, B, C, D and E) was performed.

The PLS model of averaged spectral data points of samples corresponding to groups A, B, D and E was constructed, and then validated against the averaged sample set data C that was set aside during model construction. According to these models, the RMSEP was 2.2 μM for GLU and 2.0 μM for GABA, both of which were within the range of the previous results. The constructed models predicted different concentrations of GLU and GABA in serum, as summarized in Table 3-12. The deviations of predictions were between 1.3 μM and 1.8 μM, which indicate the accuracy of PLS models for averaged spectral data points of samples.

Table 3-12 GLU and GABA concentrations prediction in five averaged samples (serum mixture)

Sample no.	GABA concentration (μM) in serum		GLU concentration (μM) in serum	
	Prediction	Deviation	Prediction	Deviation
1	7.7	1.4	93.4	1.4
2	28.3	1.7	70.7	1.5
3	53.5	1.4	42.9	1.4
4	64.1	1.4	33.7	1.8
5	81.0	1.6	16.8	1.3

The detection limit of SERS for both GLU and GABA was approximately 8 μM, which is higher than the detection limit with methods such as LC-MS/MS and HPLC that are in the nM range.

However, the in-house instrumentation for our study used less expensive optics than techniques currently being employed for monitoring GLU and GABA. Moreover, our method does not require intricate sampling procedures and causes no sample degradation, as demonstrated by the similarity in the spectrum of the samples collected on different days. The method of monitoring GLU and GABA presented here, could be further improved if the process of spectral data acquisition and subsequent feeding into the pre-built calibrated model was fully automated. There is also interest in developing new nanoparticle synthesis protocols that enhance the GLU and GABA Raman signals in complex body fluids. Thus, the SERS/PLS-based method of detecting GLU and GABA described here could potentially be a viable clinical tool for the quantitative measurement of GLU and GABA in clinical environments.

3.5 Conclusion

We have established that Raman and SERS, in conjunction with PLS, can be used to monitor the biological components in serum at a clinical level in less than a minute using less than 2 mL of samples. Raman spectroscopy was used to monitor heparin concentrations in serum as low as 8 USP/mL which is required in a clinical environment. The SERS method was applied to distinguish GLU and GABA in serum in the 8 μM, as required in clinical environments. Despite the weak Raman/SERS signal of the analyte when using a cuvette sample holder, PLS analysis measured the concentration of heparin and GLU-GABA quantitatively and provided reliable estimates of their concentrations, regardless of variability in the serum samples. Though our method used a simple, low-cost experimental configuration, it delivered results accurate enough for it to be considered a viable alternative to some existing techniques to monitor heparin and GLU-GABA in clinical environments.

Chapter 4. Surface-Enhanced Raman Spectroscopy of liquid samples using HC-PCF

HC-PCF has emerged as a new generation, micro-structured fibers that can confine light within its core region. Due to their photonic band gap, these fibers can enhance the Raman signal of a sample by supporting strong light-matter interactions. HC-PCFs can be used as nanolitre sample containers, and are ideal for characterizing low-volume chemical and biological samples. Thus, they can be employed as a "reservoir and Raman signal enhancer", and used instead of cuvette for sample solution. This chapter presents the effect of SERS in conjugation with HC-PCF (SERS HC-PCF platform) on detection of leukemia cells as a clinical application.

4.1 Introduction

HC-PCF was discussed as an effective Raman/SERS sampling technique in Chapter 2. In addition to applying it in various spectroscopic techniques, there are practical issues when it is used as a sample container, such as the formation of air gaps in the HC-PCF channels that cause low efficiency light coupling. Replacing used HC-PCF with new HC-PCF for each experiment makes it time consuming and expensive, and the slow filling rate of HC-PCF is another drawback that could be improved [160].

A. Kethani et al. introduced a novel Raman setup, using HC-PCF laid out in an H-configuration to address these issues and increase its efficiency as liquid sample holder [160]. The optical Raman system of this setup is similar to that introduced in Section 3.2, with a few differences in the sample holder. In this setup, the cuvette is replaced by HC-PCF, and it is filled using the H-shaped pressure system [160]. This improved HC-PCF configuration is used here to verify the capability of SERS to monitor leukemia cells.

4.2 HC-PCF for monitoring leukemia cells using Surface Enhanced Raman Scattering (SERS)

4.2.1 Introduction

Acute myeloid leukemia (AML) is among the most recurrent types of pediatric cancers, and the leading cause of disease-related morbidity in children and adolescents [161-162]. AML causes abnormal production of blast cells in the bone marrow, and if not treated it suppresses the normal production of platelets and white blood cells within weeks, leading to life threatening bleeding and microbial infections. Hence, the early detection of AML [163] and evaluation of the minimal residual disease (MRD) after treatment, can improve a patient's life expectancy. The current standard techniques include flow cytometry [164], polymerase chain reaction [165], immunohistochemistry [166], microarray [167] and fluorescence-based assays [168], all of which are time-consuming and relatively difficult/costly to implement. Consequently, developing new, more affordable and faster technologies for AML detection remains a major challenge.

While HC-PCF offers higher interaction lengths between the light and the analyte and lower sample consumption, SERS provides large enhancement factors to increase the sensitivity of normal Raman signals, thereby enabling the detection of molecules in various applications. Thus, the integration of HC-PCF and SERS provides an ideal platform for the detection of biomolecules [169-170].

V.S. Tiwari et al. reported a detection scheme that exploits Raman spectroscopy to determine the optimal volume and size of silver nanoparticles, in order to maximize the Raman signal enhancement of rhodamine 6G (R6G) in HC-PCF [171]. They integrated HC-PCF with nanoparticles to create a portable sensor that can detect malignant cells, such as HL60 acute myeloid leukemia. The main advantage of the proposed sensor is the potential for rapid analysis and diagnosis.

This section is organized as follows. We first present a brief description regarding choosing HC-PCF, followed by discussion of nanoparticle synthesis and cell culture. Then we summarize our

findings on the enhancement of the Raman signal from leukemia samples, considering the apoptotic, live and necrosis cell cycle stages. The last section examines the Raman sensor detection limit with respect to flow cytometry.

4.2.2 Experimental details

4.2.2.1 Choosing HC-PCF

Choosing the fiber depends on the wavelength and refractive index of the liquid sample (details of this can be found in the V.S. Tiwari et al report) [171]. In their experiment, 785-nm was chosen as the excitation wavelength, so an HC-1550 hollow core photonic bandgap fiber from NKT Photonics was initially chosen. The fiber has a core size of 10.6µm (± 1µm), and supports the center 1550 nm wavelength with a bandwidth of ~200nm. Since leukemia cells are >10µm in diameter, HC19-1550 (a variant of the HC-1550 fiber) was formed by removing 19 cells from the cladding of the fiber, making the core diameter 20µm. HC19-1550 has a core size of 20µm (± 2µm) supporting a 1570nm center wavelength with bandwidth of 100nm. Figure 4-1 shows a cross-section of the HC19-1550 fiber, and the mode-field pattern of the HC-PCF fiber filled with leukemia cells and nanoparticles. The modal field diameter was found to be ~4.5mm as shown in Figure 4-1(b), which is quite close to the calculated theoretical value of 4.2mm. We found a similar mode field pattern for other sets of sample solutions with different leukemia cell cycle stages and different concentrations of leukemia cells. Thus, it was confirmed that the propagation properties did not change significantly with different leukemia samples.

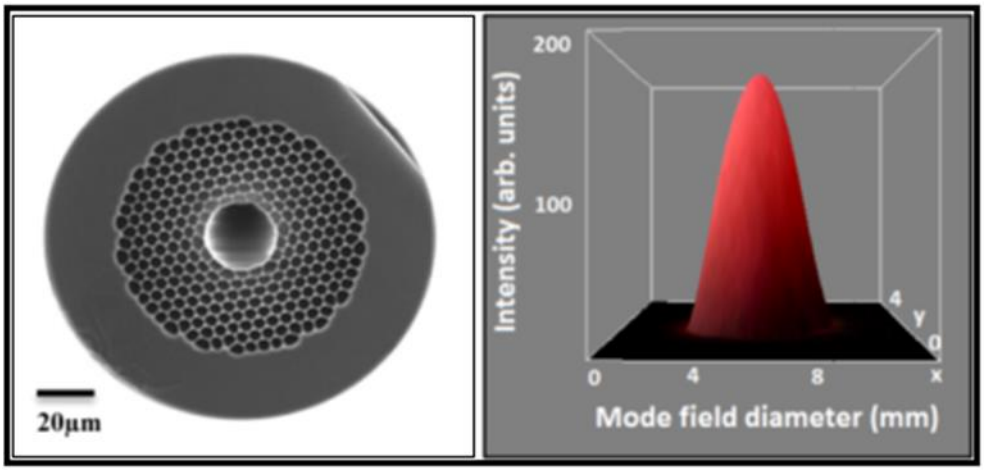

Figure 4-1 Hollow core photonic crystal fiber HC19-1550 (a) SEM image courtesy NKT Photonics Inc. (b) spatial distribution of modal field of leukemia sample solution filled HC-PCF imaged using a CCD camera (Canon) exhibiting a perfect Gaussian profile

4.2.2.2 Sample preparation

Acute promyelocytic leukemia (HL60) cells (ATCC® CCL-240) were cultured in Iscove's Modified Dulbecco's Medium (Sigma), supplemented with 20% fetal bovine serum, 1% antibiotics (streptomycin and penicillin) and 0.1% gentamicin. Cells were incubated at 37°C, 5.0% CO_2 and 100% humidity, and in all cases the cell density was kept between 0.1 and $1.0x10^6$ cells/ml. Apoptosis was induced by incubating the cells with 5.0 µM (*S*)-(+)-Camptothecin (CPT), a topoisomerase I inhibitor [172-174], for three hours in Hanks Buffer (Sigma) using a $1.0x10^5$ cells/ml density. The cells were then centrifuged at 1000 rpm for 5 minutes, and the pellet was re-suspended in 100 µL ($1.0x10^6$ cells/ml cell density) of Annexin binding buffer (Life Technologies), with 1.0 µL of 50 µg/mL propidium iodide (PI) and 5.0 µL Annexin V-FITC added to the cells. The mixture was incubated at room temperature for 15 minutes, then another 400 µL of Annexin buffer was added prior to sorting in a BD FACS-Aria flow cytometer. Cell sorting was carried out by gating non-stained cells (live), PI and Annexin V positive stained cells (necrotic; λ_{exc} = 488 nm, λ_{em} = 585 ± 21 nm), and Annexin V positive cells (apoptotic λ_{exc} = 488 nm, λ_{em} = 530 ± 15 nm). The number of cells in all cases were measured in a Vi-Cell (Beckman Coulter). Additional experiments were carried out for the non-stained cells following the scattering profile in the flow cytometer. In these experiments, the total number of cells varied

between 310 and 25000 cells/mL, using serial dilution in the cell culture medium without phenol red. Control experiments were carried out with the same dilution procedure as the cell culture medium, without cells.

Nanoparticles Synthesis: The role of nanoparticles is to enhance the weak Raman signal of molecules/cells resulting from their concentrations. Silver nanoparticles were prepared by chemical reduction in an aqueous solution of silver nitrate, according to the method described by Leopold et al. [155], and were used within two hours of preparation. The UV–vis absorption spectrum and TEM image of AgNP are shown in Figure 4-2. Silver spheres show only one principal plasmon band at ~430 nm, and the majority of silver nanoparticles were spherical with average size of ~60nm.

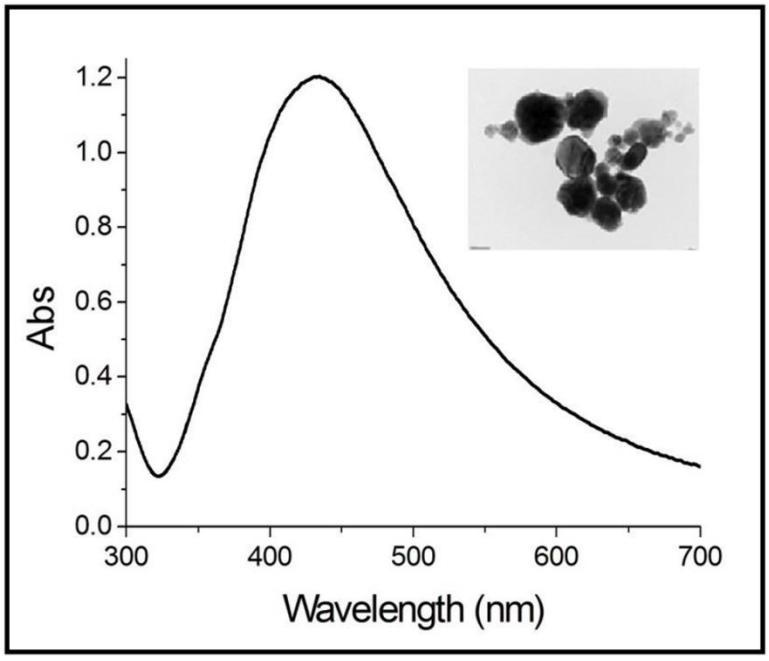

Figure 4-2 UV-Vis absorption spectrum of silver nanoparticles. Inset shows the TEM image of spherical silver nanoparticles of ~60nm size

4.2.2.3 Experimental configuration

The layout of the HC-PCF sensor is shown in Figure 4-3, and is the same as H-configuration introduced by A. Khetani et al. [160]. The only difference is the use of a 40X microscope objective lens (L1) with a NA of ~0.65. The light coupling efficiency of the leukemia and

nanoparticles solution in the HC-PCF was ~30%. The other segment of the sensor configuration is comprised of two parallel channels (tubing): one for sample (leukemia cells) input/output, and the other for purging fluid (water) input/output. The integration of HC-PCF is perpendicular, and with the two parallel fluidic channels it forms an H-shaped structure, as shown in the Figure 4-3. In order to flow Sample 1 (i.e. sample mixture) through the fiber channels, average pressure P1 was set higher than average pressure P2, which ensured that the fiber is filled. Similarly, to purge the sample from the fiber the pressure was reversed, and average pressure P2 was set higher than average pressure P1. The average pressure and the rate at which the sample is pumped into the fiber is discussed in detail [160].

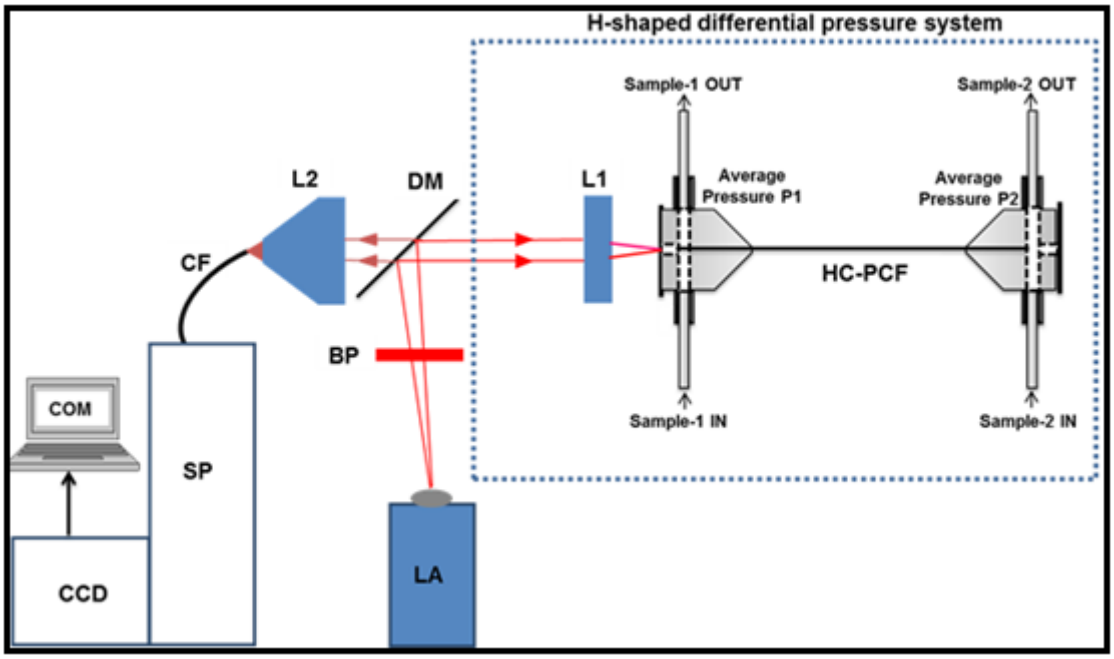

Figure 4-3 Schematic of the setup. LA: Laser; BP: Band pass filter; DM: Dichroic Mirror; L1: Light coupling lens; L2: light collector lens; CF: Collection fiber; SP: spectrograph; CCD: CCD camera; COM: Computer

4.2.3 Results and discussion

4.2.3.1 Enhancement of Raman signal with HC-PCF and nanoparticles

The first step of this experiment was to record the Raman spectra of HL60 cells with $1x10^6$ cells/ml in cuvette and a mixture of nanoparticles in HC-PCF. The Raman spectra of leukemia

cells in cuvette, as shown in Figure 4-4, have peaks at 1032 cm^{-1} (C-N stretching mode of phenylalanine) and 1318 cm^{-1} (protein). The mixture of leukemia cells and nanoparticles in HC-PCF produced a rich spectrum, with evident features at approximately 650 cm^{-1} (protein, C-S stretching, tryptophan, C-N stretching), 722 cm^{-1} (C-H rocking of CH2 methylene group in lipids), 789 cm^{-1} (O-P-O ring breathing modes of DNA/RNA bases), 1003 cm^{-1} (symmetric ring breathing mode of phenylalanine), 1032 cm^{-1} (C-N stretching mode of phenylalanine), 1093 cm^{-1} (O-P-O symmetric stretching mode of protein), 1119 cm^{-1} (C-N stretching mode of protein), 1283 cm^{-1} (amide III), 1318 cm^{-1} (protein), and 1436 cm^{-1} (CH bending of lipids) [162, 175]. Our aim was to determine the factor by which HC-PCF and nanoparticles enhance the Raman signal, which we found to be ~2,700. HC-PCFs are known to enhance Raman signals since they support strong modal field overlap with the sample, due to its photonic band gap property. The enhancement factor of the sensor was calculated by dividing the Raman signal of leukemia cells and nanoparticles from HC-PCF by the Raman signal of leukemia cells from the cuvette.

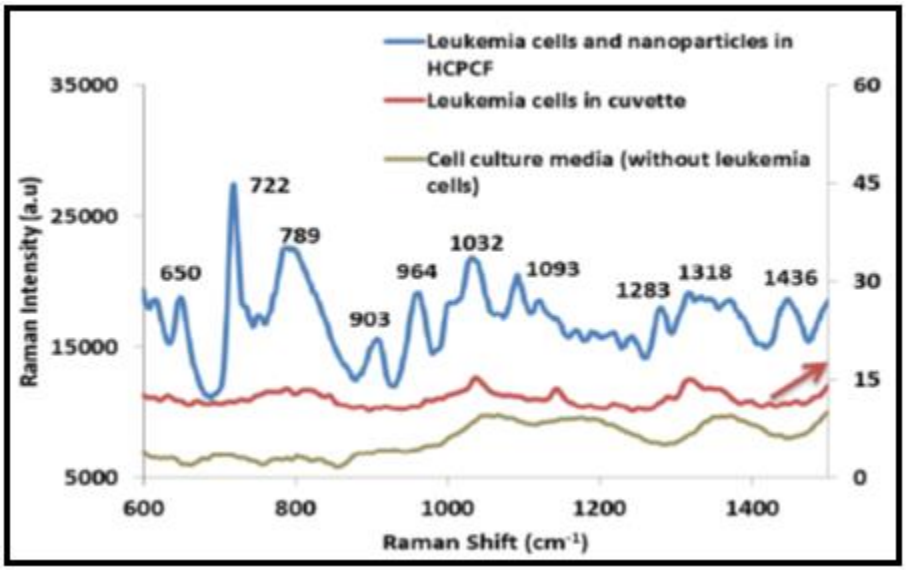

Figure 4-4 Enhancement of Raman signal of HL60 cells in HC-PCF using silver nanoparticles

4.2.3.2 Multivariate data analysis

In the next phase of our demonstration, we used statistical analysis to distinguish between the cell cycle states. As explained in Chapter 2, principal component analysis (PCA) and partial least

squares (PLS) are critical aspects of multivariate data analysis, with the role of verifying and detecting the classification and minimum levels of different leukemia cells. The Unscrambler version 10.3 (CAMO, Corvallis, OR, USA) was used to perform the multivariate data analysis. The score plot of PC1 and PC2 is shown in Figure 4-5, and it reveals different groups in the samples. PCA analysis on the different cycle stages of live, necrotic, and apoptotic HL60 cells, yields a distinctly distinguishable Raman signature.

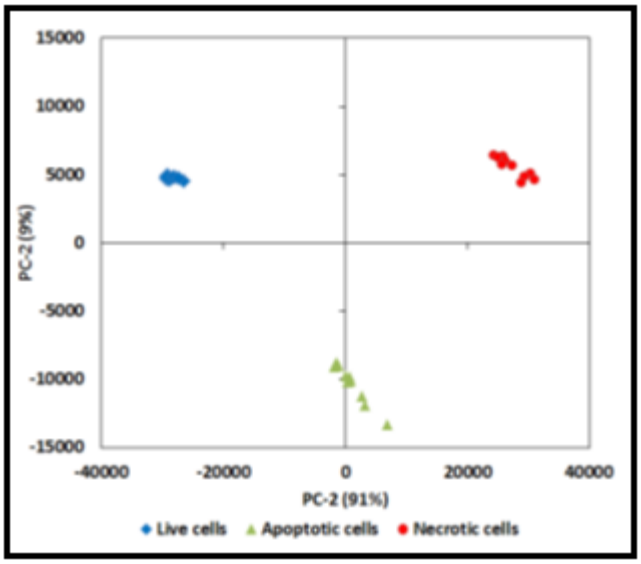

Figure 4-5 Plots of principal component analysis (PCA) analysis of Raman spectra of different leukemia cells stages showing distinguishable live, apoptotic and necrotic cell stages

Raman spectra of different cells cycle stages are shown in Figure 4-6, and they demonstrate the variations of Raman intensity at important wavenumbers. Most necrotic Raman bands show higher intensity than live and apoptotic cells, except at 722 cm^{-1} and 1001 cm^{-1}. In addition, the average intensity of the bands in apoptotic cells is higher than that of live cells. These spectra, as well as the score plot of PCA, enable us to distinguish the cells from one another.

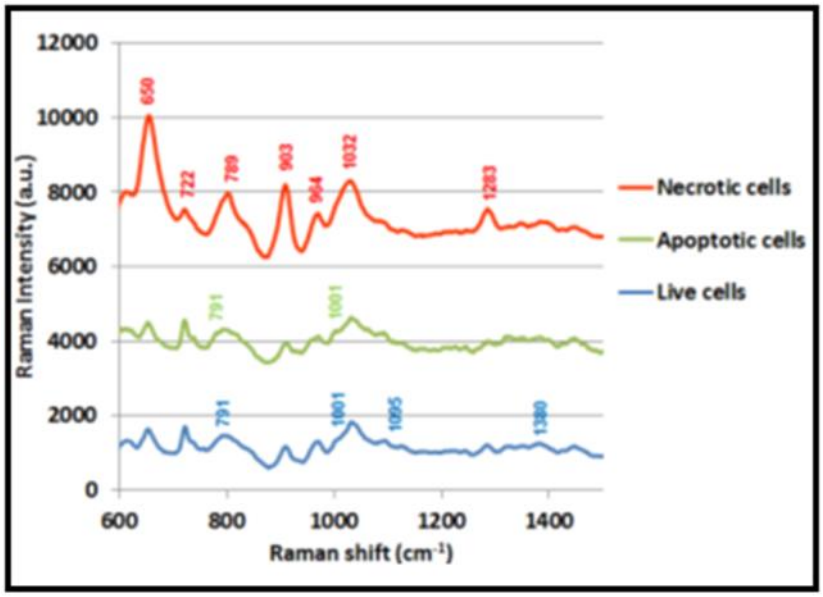

Figure 4-6 Raman spectra of leukemia cells cycle stages with distinguishing Raman peaks in live, apoptotic and necrotic cells

4.2.3.3 SERS in HC-PCF for different leukemia cells concentrations

In the next step of the experiment, we recorded the Raman spectra of different concentrations of leukemia cells and nanoparticles in HC-PCF. Figure 4-7 shows the SERS spectra at six different concentrations of leukemia cells. As discussed earlier, Raman peaks at 650 cm^{-1} (C-S stretching of protein, C-N stretching of tryptophan), 722 cm^{-1} (C-H rocking of CH2 methylene group in lipids), 789 cm^{-1} (O-P-O ring breathing modes of DNA/RNA bases), 1003 cm^{-1} (symmetric ring breathing mode of phenylalanine), 1032 cm^{-1} (C-N stretching mode of phenylalanine), 1093 cm^{-1} (O-P-O symmetric stretching mode of protein), 1119 cm^{-1} (C-N stretching mode of protein), 1318 cm^{-1} (protein) and 1436 cm^{-1} (C-H bending of lipids) were the prominent peaks of the six concentrations. After recording the Raman spectra of the HL60 cells, we applied multivariate analysis to correlate the Raman signals with the sample concentrations.

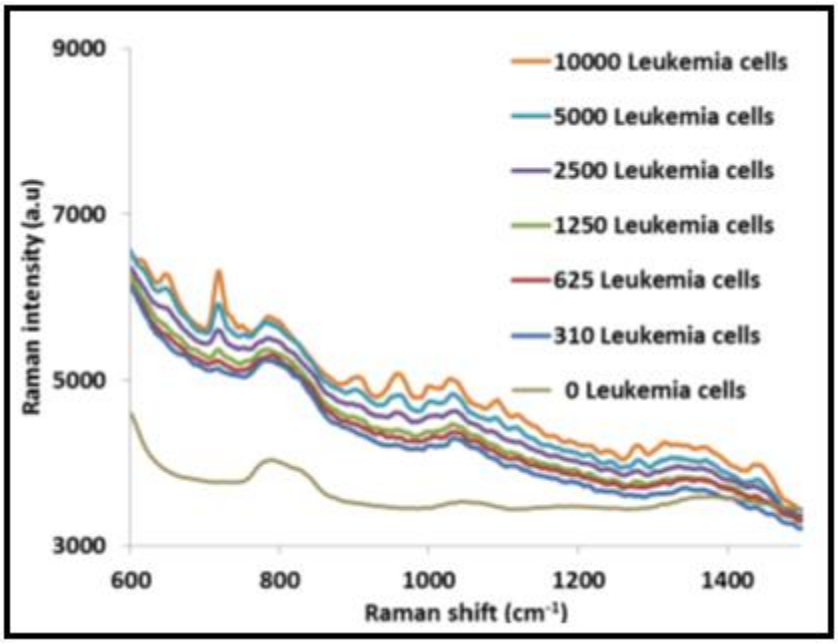

Figure 4-7 SERS spectra of different concentrations of live HL60 cells, expressed as cells/ml

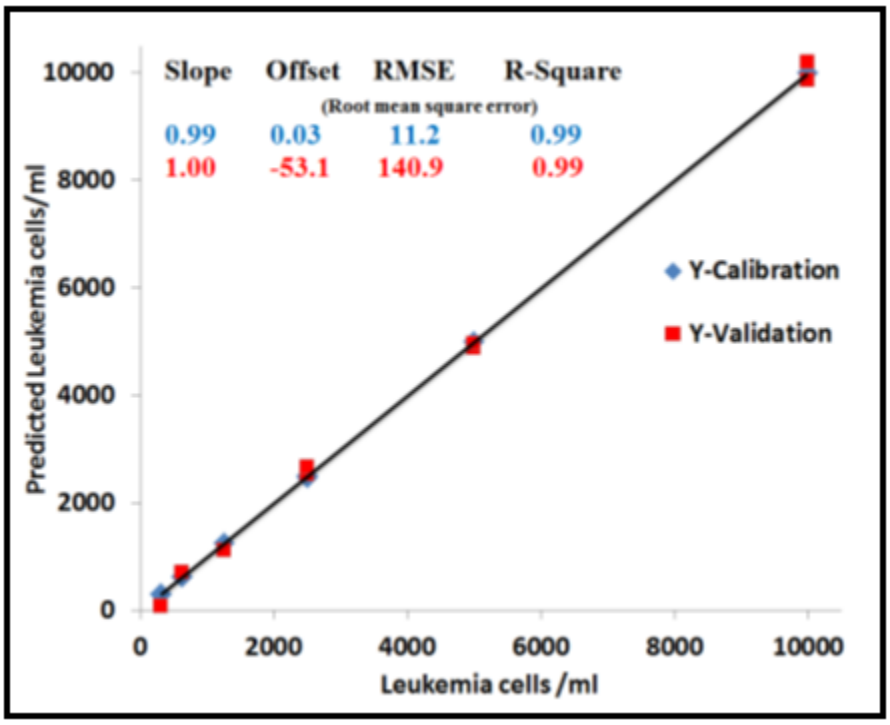

Figure 4-8 PLS prediction of different concentrations of leukemia cells/ml showing excellent correlation between calibrated and predicted samples with R2 = 0.99 and RMSEC and RMSEP of 11 and 140 cells/ml

Once the Raman spectra of the six samples were determined, ten Raman spectra were recorded for each sample to ensure repeatability of measurement. The Raman spectra of six samples were divided into three subsets: a calibration set, validation set and test set. The calibration and validation sets were used to create the model while the test set, as independent data, was used to evaluate the model. Three of the recorded spectra were identified as outliers by the Unscrambler software, and were not used in making the model. According to the PLS model of 45 records of samples in the calibration and validation sets, R^2 (for calibration and validation) RMSEC and RMSEP were 0.99, 0.99, 11 and 140, respectively. The calibration curve of this model is shown in Figure 4-8.

4.2.3.4 Comparing the HC-PCF sensor with flow cytometry

Flow cytometry experiments using non-stained cells were also conducted to compare the technique with our HC-PCF method. The data shown in Figure 4-9 clearly shows the lack of linearity between the number of events detected by the systems, and the total number of cells in the solution. Indeed, we found that the lower limit of detection was between 500 and 2500 cells/mL compared to solutions without cells. The red region in the inset in Figure 4-9 denotes the 500 cells/ml limit we consistently found in the background under our experimental conditions (see the right panels in Figure 4-9). In addition, the inset also shows how the point at 2500 cells/mL does not correlate with the total number of events detected by the equipment. Furthermore, a simple visual inspection of Figure 4-9 reveals an upward curvature, which also indicates poor correlation between the actual number of cells in the sample and those detected by the system.

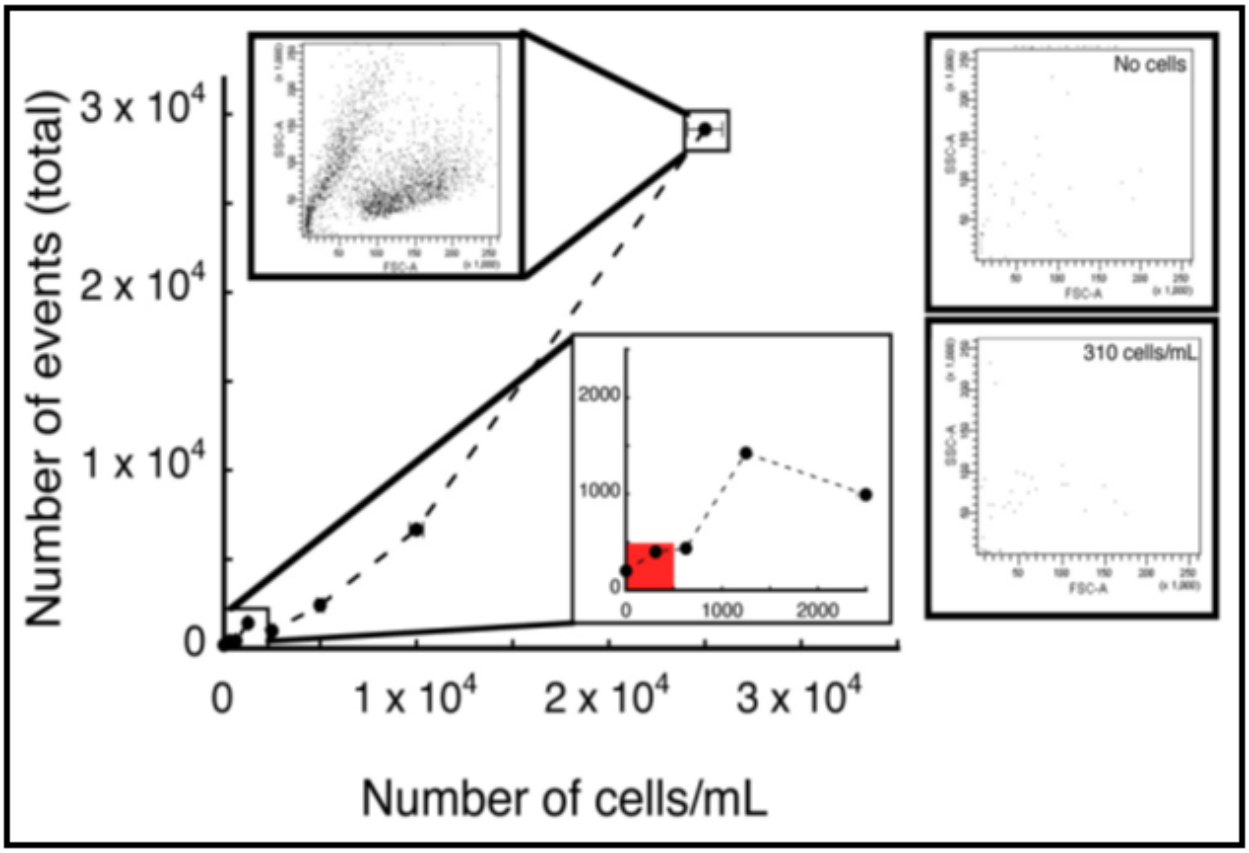

Figure 4-9 Number of events vs. total number of cells/mL for HL60 cells Number of events was quantified by using the scattering on 488-nm excitation of a non-stained cell suspension. The top inset shows the scattering profile for the sample containing 25,000 cells/mL. The bottom inset shows a magnification for the lower cell numbers with the red area representing the noise region or lower limit of detection for the system. Right panels include the scattering profile for the sample with 310 cells/mL and a control solution with no cells, which clearly shows the close similarity between the two plots

4.3 Conclusion

We have demonstrated a robust and sensitive platform for the monitoring and detection of leukemia cells. The scheme uses nanoparticles and hollow core photonic crystal fiber to enhance the weak Raman signal of leukemia cells, and it enabled us to achieve an enhancement of ~2700. In addition, we applied PCA statistical analysis to differentiate apoptotic, live and necrotic cells. We also used PCA analysis on different cell counts, which demonstrated the capability of the sensor to detect less cells. With our sensor, we successfully detected ~300 cells/ml which is 8 times better than flow cytometry method.

Chapter 5. An improved PLS regression method for Raman spectroscopy

As explained in Chapter 2, variable selection is one of the approved methods for optimizing a PLS model. Of the variable selection techniques explained earlier, this thesis focuses on the BVSPLS method, in which the most important variables are specified by using RMSEP criterion during the iteration steps. A better understanding of the IBVSPLS algorithm requires detailed familiarity.

5.1 Introduction

There are numerous approaches to improve the performance of any PLS model, one of which is based on removing the outlier samples from original dataset [78]. Outlier samples look entirely different than other samples and do not have the same (X, Y) linear relationships, so they are not well defined by the constructed PLS model. Most multivariate statistical software has features that identify outlier samples so they can be removed, as explained in Chapter 2. BVSPLS is another approach that was examed in Chapter 2 and its further improvement, which is the main target of this chapter, is described in the following sections.

5.2 Comparison between BVSPLS and IBVSPLS methods

5.2.1 BVSPLS algorithm

The first specification of the BVSPLS algorithm is related to its RMSEP criterion; all constructed models are compared in terms of their RMSEPs. The second specification is the number of iterations, which is related to the spectral range of measurement. In a Raman spectroscopic application, for example, if the first wavenumber of the spectral range is 401 cm^{-1}, the last wavenumber of the range is 1600 cm^{-1}, and the spectral resolution is 1 cm^{-1}, the number of iterations would be 1200. This means 1200 PLS models must be constructed, and each compared with the previous PLS model according to their RMSEPs. The constructed model with the lowest RMSEP in each iteration is retained, and used for comparison in the subsequent

iteration. The third specification of the algorithm is regarding the order of variable selection. The BVSPLS algorithm begins in a very straightforward way: the variable range starts with the first wavenumber of the range, and ends with the last wavenumber of the range. In the example above, the first variable evaluated is 401 cm^{-1}, and the last is the final wavenumber, 1600 cm^{-1}.

These three specifications allow us to investigate how the BVSPLS algorithm works. The first model is constructed using all the variables, then employed as a reference for comparison in the next iteration. The variable selection for verification of the PLS model begins with the first wavenumber being removed from the variable range, then the PLS model with the remaining variables is constructed and compared with the PLS model using all the variables. The first variable is removed or retained according to RMSEP of the model, after which the second variable is removed and the second constructed PLS model is compared with the previous PLS model. The PLS model with less RMSEP is considered the new reference model for the next iteration. This process is repeated until the last variable of the range, and the last PLS model has the lowest RMSEP based on the algorithm. In every iteration, the calibration dataset is used to construct the model, and the test dataset is used to evaluate it. Figure 5-1 illustrates the procedure.

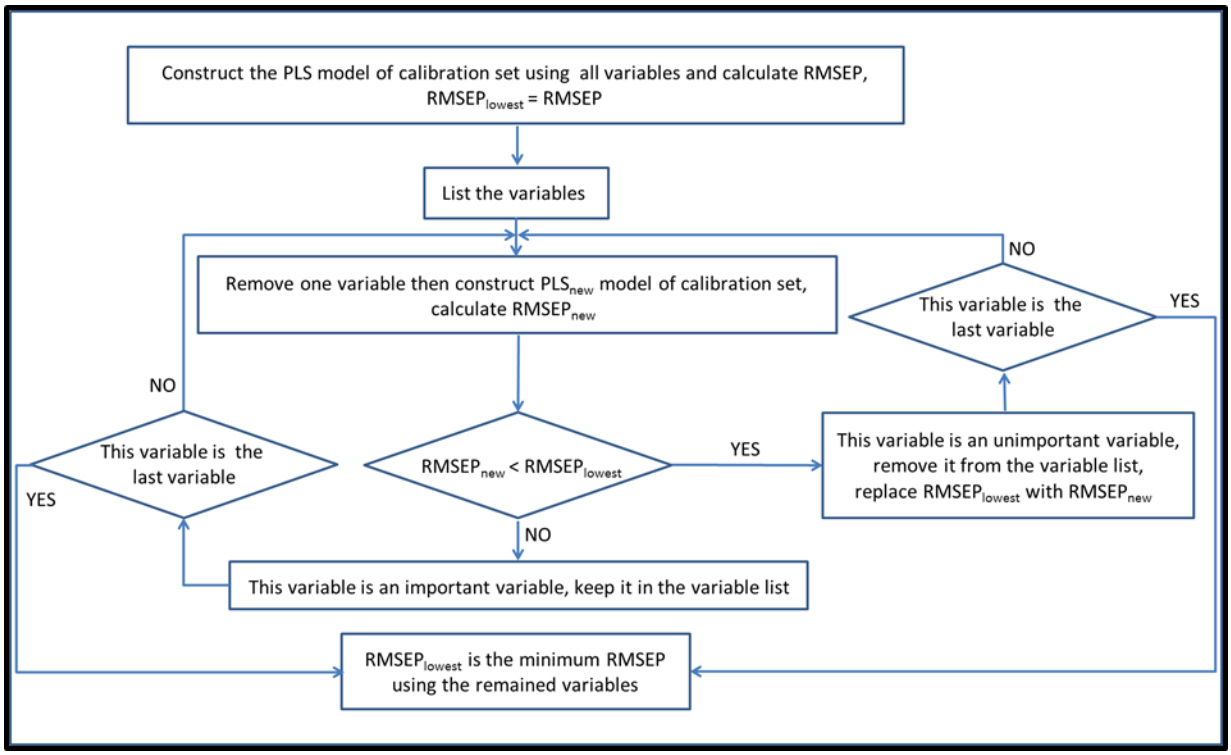

Figure 5-1 Flow chart of BVSPLS algorithm

5.2.2 IBVSPLS algorithm

This algorithm is based on the regression coefficients of a PLS model that are constructed based on a dataset, and they indicate the relationship between dependent and independent variables in the dataset. These coefficients can be found in any linear regression analysis in many applications, and an example of Raman spectroscopic dataset is used to better understand this algorithm. In Raman spectroscopy, the Raman spectrum of a sample typically exhibits prominent peaks at specific wavenumbers, which explains the molecular structure of the sample. The intensities of the peaks are related to the concentration of the sample. For example, Figure 2-9 illustrates the dependency between ethanol concentration (dependent variable) and the Raman intensity at different wavenumbers (independent variables). The prominent peaks at 433 cm^{-1}, 882 cm^{-1}, 1051 cm^{-1}, 1097 cm^{-1}, 1276 cm^{-1} and 1454 cm^{-1} have high intensities, which indicates that these wavenumbers and those around them are more important than the other wavenumbers.

When MVA is applied to a dataset of samples with different analyte concentrations, one of the main outputs is the regression coefficient list, which highlights the important and informative independent variables of the range [86,176-177]. To avoid the effects of small regression coefficients with large variances, datasets are weighted with the inverse of the standard deviation [86]. A higher absolute value in the weighted regression coefficient list means a variable has greater influence on the dependent variable. For example, in a constructed PLS model of a dataset of samples with different ethanol concentrations, the weighted regression coefficient list shows different values corresponding to each wavenumber, that express the correlation between the Raman intensities at each wavenumber and the absolute value of the weighted regression coefficients. Thus, the value of the weighted regression coefficient of a variable at 882 cm^{-1} is greater than the value of a weighted regression coefficient of a variable at 433 cm^{-1}, and so on. The weighted regression coefficient list of the PLS model is used to run the IBVSPLS algorithm.

Similar to BVSPLS, the IBVSPLS algorithm has three specifications; the only difference is the order of variable selection in the third specification. With the BVSPLS algorithm, all variables are verified individually from the first variable of the range to the last, while the IBVSPLS technique relies on this fact that some variables are more important than others. Based on this, the IBVSPLS algorithm requires the list of weighted regression coefficients of the first constructed PLS model. The list usually contains positive and negative numbers, and all variables with positive maximum weighted regression coefficients and negative minimum weighted regression coefficients have a significant impact on the dependent variable. The weighted regression coefficients list is not sorted, and to use it as a list of important variables it must be sorted from absolute minimum value to absolute maximum value. Consequently, the less important variables will be at the top of the list while the more important ones, higher positive or lower negative values of weighted regression coefficients will be at the end.

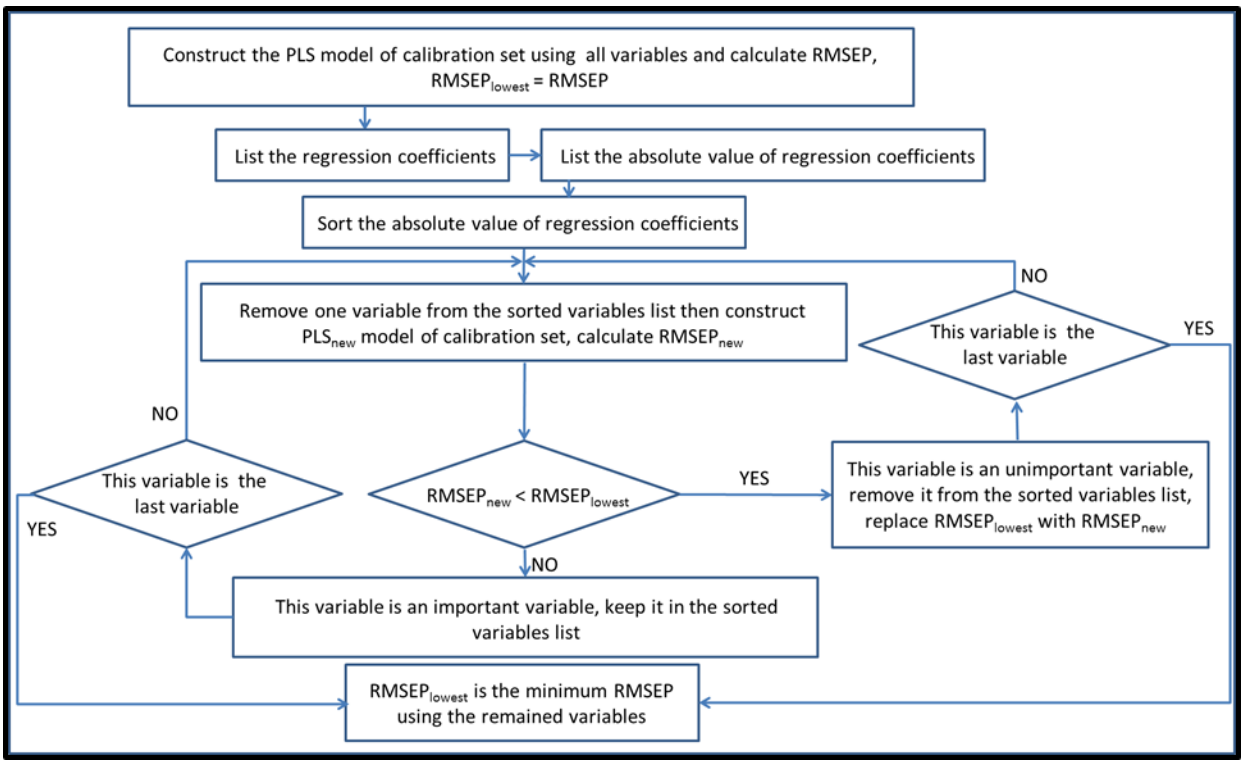

Figure 5-2 Flow chart of IBVSPLS algorithm

Figure 5-2 illustrates the procedures of the IBVSPLS method. The first PLS model is made by selecting all the variables, as shown, and continues by sorting the absolute value of the weighted regression coefficients from minimum to maximum. RMSEP is used to evaluate the models. Then the first variable of the sorted variables is omitted when the second PLS model is created. If the RMSEP of the second PLS model is less than that of initial model, the first variable is not important and can be disregarded. Then the second variable of the sorted variables is left out of the calculation, the third PLS model is created, and its RMSEP is compared with the RMSEP of the second model. If the new RMSEP is less than the previous one, the second variable is not important and can be ignored, while if it is more than the previous RMSEP the second variable is required for the new PLS. This procedure continues until the last variable of the sorted list has been addressed. Here, the constructed model using selected variables based on weighted regression coefficients shows a lower RMSEP than the BVSPLS model. Thus, the comparison between BVSPLS and IBVSPLS methods is based on using the weighted regression coefficients list. While both our proposal and the BSVPLS method use

the RMSEP as a criterion for selecting the variables, our method is based on the fact that some variables are more important than others, and this is depicted in a sorted, weighted regression coefficient list. With BVSPLS there is no ranking of the variables, and iteration starts from first variable of the range. The advantage of our method is that, when creating a PLS model it guarantees each iteration includes the more important variables and excludes the less important. In the next section, Raman spectroscopy datasets are used to evaluate the IBVSPLS algorithm, and compare it with the BVSPLS Jack-knifing and GA-PLS methods. This procedure could be extended to other types of datasets.

5.3 Experimental details

Two spectral datasets of Raman spectroscopic analysis were used in this study to evaluate the RMSEP of constructed regression models, as explained in Chapters 3 and 4. The first dataset was the Raman spectra of leukemia cells [178], while the second was the Raman spectra of heparin in serum [179].

An HC-PCF was used in the first case as a sample container, and a cuvette was used in the second case. The other experimental configurations were very similar, and were detailed in Chapter 3.

The PLS analysis was conducted using Unscrambler® X version 10.3 (CAMO, Corvallis, OR), which randomly split the datasets into different subsets. The GA-PLS method was performed using a Matlab code [180]. The Jack-knifing and GA-PLS models are techniques that use cross validation to build a PLS model, which is why they were applied ten times and the average RMSEP was calculated.

5.4 Results and discussion

The PLS, Jack-knifing, GA-PLS, BVSPLS and IBVSPLS methods were applied to different datasets. A group of samples was used to make the PLS models and the remaining data was applied to evaluate the results independently, in order to reduce overfitting in the constructed models. Table 5-1 shows the results of these methods when applied to different datasets.

Table 5-1 The results of PLS, Jack-knifing, GA-PLS, BVSPLS, and IBVSPLS models

Analyte	model	PC	RMSEP(test)	Number of used variables (percentage)	Improvement (%)
Leukemia	PLS	4	231.30	781(100%)	-
	Jack-knifing	4	217.42	448(57%)	6.0
	GA-PLS	4	206.79	111(14%)	10.6
	BVSPLS	4	227.13	411(53%)	1.8
	IBVSPLS-1	4	222.44	345(44%)	3.8
	IBVSPLS-2	4	207.02	167(21%)	10.5
Heparin	PLS	3	1.39	477(100%)	-
	Jack-knifing	4	3.83	181(38%)	-
	GA-PLS	7	2.70	164(34%)	-
	BVSPLS	3	1.27	280(59%)	8.6
	IBVSPLS-1	3	1.00	222(46%)	28.1
	IBVSPLS-2	3	0.88	155(33%)	36.7
	IBVSPLS-3	3	0.83	136(28%)	40.3
	IBVSPLS-4	3	0.82	116(24%)	41.0
	IBVSPLS-5	2	0.79	114(24%)	43.2

The methods were applied to each dataset, and the RMSEPs were compared. In the case of IBVSPLS, it was reapplied to the final model several times to achieve decreased RMSEP. The degree of RMSEP improvement is compared with the RMSEP of the PLS model which used all variables. The improvement (last column) is the relative percentage difference which is calculated from the difference between RMSEP of each model and PLS model, dividing by RMSEP of PLS model, and multiplying by 100.

5.4.1 Leukemia cells

A. Khetani et al. [178] introduced a Raman based portable sensor to detect malignant cells, such as HL60 acute myeloid leukemia (AML), a repetitive pediatric cancer that targets children and adolescents. The first dataset we used in this study was based on the Raman spectra of leukemia cell samples. As shown in Figure 5-3, the main Raman peaks of (10000 cells/mL) samples are at 650 cm^{-1}, 722 cm^{-1}, 789 cm^{-1}, 907 cm^{-1}, 960 cm^{-1}, 1032 cm^{-1}, 1087 cm^{-1}, 1246 cm^{-1}, 1285 cm^{-1}, 1318 cm^{-1}, 1339 cm^{-1}, 1368 cm^{-1} and 1436 cm^{-1}. These Raman peaks were assigned

to different bands of leukemia cells, as explained by A. Khetani et al. [178]. According to Table 5-1, all the methods improved the RMSEP of the model compared to the RMSEP of the PLS model. Using the Jack-knifing, GA-PLS, BVSPLS and IBVSPLS methods decreased the RMSEP of the model from 231.3 to 217.42, 206.79, 227.13, and 207.02, respectively. In the case of the GA-PLS method, the RMSEP was as usable as that of the IBVSPLS method. The last column of Table 5-1 shows the degree of improvement, based on the ratio of $RMSEP_{PLS}$ - $RMSEP_{model}$ to $RMSEP_{PLS}$. The GA-PLS and IBVSPLS methods had about 10% improvement for the first dataset, while Jack-knifing and BVSPLS showed less improvement. The number of used variables went from 781 in the PLS model, to 448, 111, 411 and 167 in the Jack-knifing, GA-PLS, BVSPLS and IBVSPLS models, respectively, which means the GA-PLS and IBVSPLS models use the most important variables in the wavenumber range. The important variables that were selected using the IBVSPLS method were verified by comparing the Raman spectra of the leukemia cells (10000 cell/mL) before and after selecting the variables, as shown in Figure 5-3. After using the IBVSPLS model, most of the remaining variables are located around the important Raman peaks of leukemia cells, and the non-informative variables are discarded. Moreover, the R^2 of the PLS, Jack-knifing, GA-PLS, BVSPLS, and IBVSPLS models were 0.9954, 0.9960, 0.9963, 0.9956 and 0.9963, respectively, which shows that data fitting to the model was improved with the IBVSPLS and GA-PLS methods. Therefore, IBVSPLS has better results than the Jack-knifing and BVSPLS methods, and it is as usable as GA-PLS.

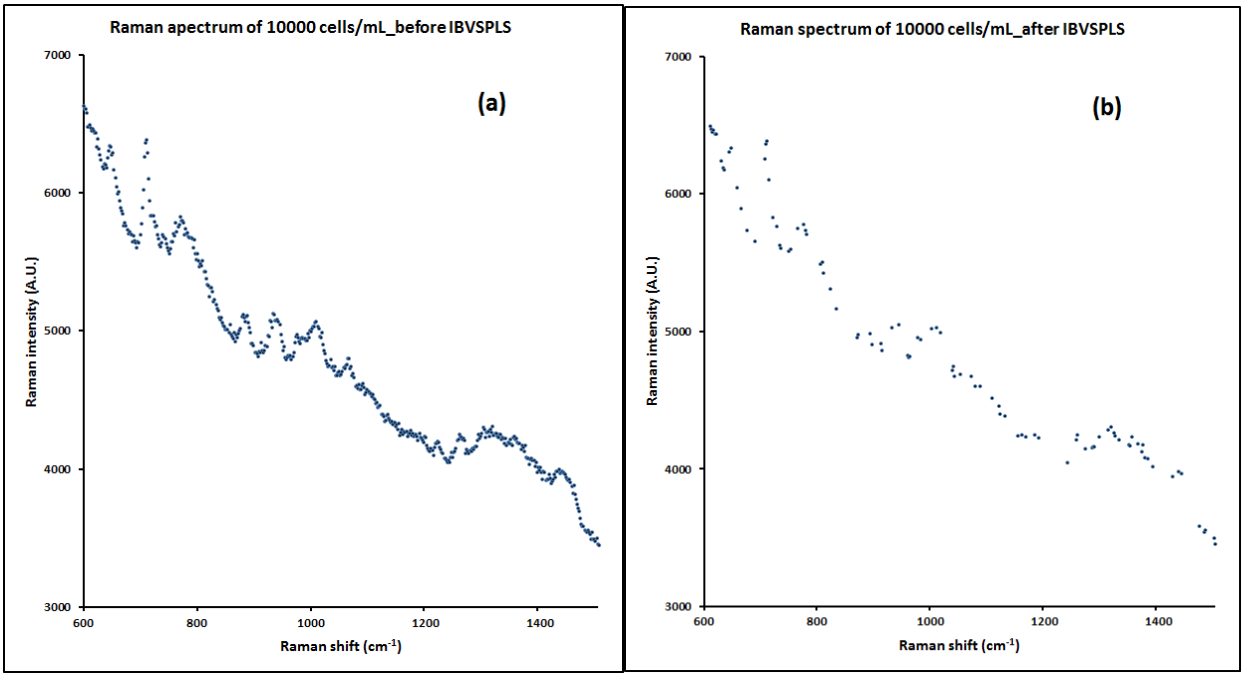

Figure 5-3 Illustration of Raman spectra of leukemia cells (a) before and (b)

5.4.2 Heparin in serum

This section describes how the heparin serum dataset from Chapter 3 was used to show the capability of the IBVSPLS method. The Raman spectra of the heparin serum sample before and after variable selection procedure are shown in Figure 5-4. The concentration of heparin in serum was labeled in terms of USP per mL, which reflects the potency of the drug in clinical applications [179]. The figure also shows the baseline corrected spectrum of the serum heparin mixture when the heparin concentration is 83.3 USP/mL. The assignment of Raman peaks at different wavenumbers were taken from the literature. According to Figure 5-4, the Raman peaks at 827 cm^{-1} (C_1-H deformation of α-anomer) [128], 893 cm^{-1} (C_1-H deformation of β-anomer) [128], 1035 cm^{-1} (N-SO_3 vibration) [128], 1045 cm^{-1} (6-O-SO_3 vibration) [128], 1060 cm^{-1} (2-O- SO_3 symmetric stretch) [128], are heparin peaks. The Raman peaks at 642 cm^{-1} (C-S tyrosine) [181-183], 854 cm^{-1} ((COC), proline (CCH) ring breathing, tyrosine) [184], 948 cm^{-1} (C-C α-helix, proline, valine) [184-185], 1003 cm^{-1} (C-N stretching, C-C symmetric stretching, phenylalanine) [128, 181], [183-184], 1121 cm^{-1} (C-C stretching, C-N stretching, proteins) [184, 186], 1167 cm^{-1} (ring cyclic stretch, in-plane C-H bending, tyrosine , CO-O-C asymmetric

stretching, lipids) [184, 187-188], 1244 cm^{-1} (β-sheet, SO_3 asymmetric stretch, amide III) [186, 188, 189], 1280 cm^{-1} (CH_2 wagging, amide III) [188], 1318 cm^{-1} (CH_3CH_2 twisting, amide III proteins) [184, 188], 1339 cm^{-1} (α-helix, C-H bend, C-C stretching mode, protein, phospholipids, phenylalanine) [184, 186, 188], and 1451 cm^{-1} (CH_3, CH_2 bending modes, protein) [184, 188] are the common peaks in serum and heparin. And the the Raman peaks at 665 cm^{-1} (C-S stretching mode, ring breathing mode, asymmetric vibrations of Te-O bonds, protein, nucleic acids, glass) [188, 190], 801 cm^{-1} (ring breathing mode, nucleic acids) [188], 876 cm^{-1} (δ-ring mode, C-C stretching, tryptophan, hydroxyproline) [183-184], 1206 cm^{-1} (C-C_6H_5 ring vibration, CH_2 twist, tyrosine ,tryptophan or phenylalanine) [184-185, 188], 1310 cm^{-1} (CH_3, CH_2 twisting or bending, CH_2 twisting, lipid) [185, 188] and 1401 cm^{-1} (CH_3 symmetric bending modes, COO^- symmetric vibration, methyl groups of proteins, amino acids) [158, 188] are the serum or glass peaks.

Figure 5-4 shows the Raman peaks of the sample after using 167 variables based on the IBVSPLS model. The RMSEP of the PLS model for heparin concentration, which used all the variables, was 1.39 USP/mL. The RMSEP decreased to 1.27 USP/mL in the BVSPLS model, and 1.00 USP/mL in the first run of the IBVSPLS model. Reapplying the IBVSPLS model reduced the RMSEP from 1.00 to 0.79 USP/mL. The IBVSPLS model showed an approximate improvement of 43.2% for the second set of data, while the BVSPLS model showed 8.6% improvement. The Jack-knifing and GA-PLS methods did not improve the RMSEP as much as the BVSPLS and IBVSPLS methods. The number of variables used in the PLS, Jack-knifing, GA-PLS, BVSPLS and IBVSPLS models were 477, 181, 164, 280 and 114, respectively, which indicates that excluding unimportant variables and using the IBVSPLS method can improve the limit of detection more than the Jack-knifing, GA-PLS or BVSPLS methods. The R^2 of validation was 0.9965, 0.9742, 0.9872, 0.9971 and 0.9988 for the PLS, Jack-knifing, GA-PLS, BVSPLS, and IBVSPLS models, respectively. This highlights the improvement of the constructed model after discarding unimportant variables, particularly with the IBVSPLS method.

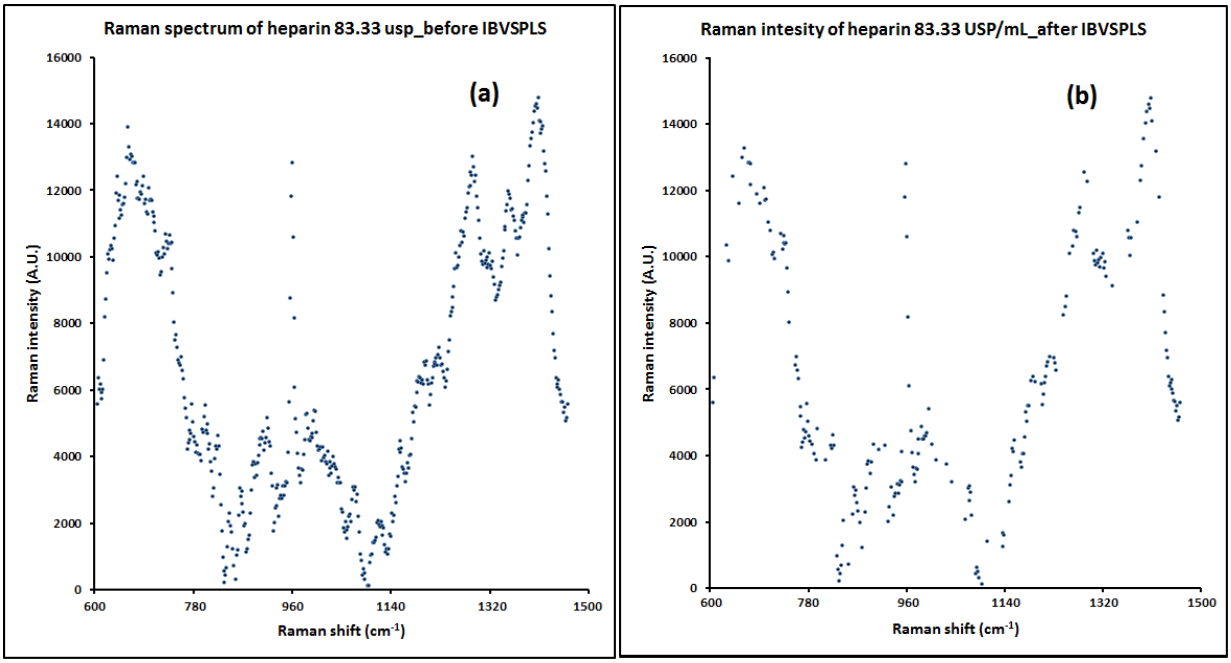

Figure 5-4 Illustration of Raman spectra of heparin-serum sample (a) before and (b) after IBVSPLS

5.5 Conclusion

In this chapter, the IBVSPLS model was introduced to improve the performance of PLS regression analysis by only selecting the relevant variables within a range of wavenumbers. Our method is based on the fact that weighted regression coefficient list contains both important and unimportant variables within the range. By sorting the absolute value of the weighted regression coefficients from minimum to maximum, non-informative variables that are related to lower weighted regression coefficients are identified, thus revealing the more important variables that can use RMSEP as the main selection criterion for the iterative steps. The IBVSPLS model that uses only informative variables has a lower limit of detection than the Jack-knifing, GA-PLS and BVSPLS methods.

6. Using SERS to detect PCOS disease

This chapter demonstrates how our enhanced Raman technique can be used to differentiate between PCOS and non-PCOS patients. We have determined that the use of SERS, in conjunction with PLS and PCA, allows us to detect PCOS in patient samples and measure their chemerin levels. Further, we applied the IBVSPLS regression method (introduced in Chapter 5) to reduce the LOD of chemerin in the samples.

6.1 Introduction

Polycystic ovary syndrome (PCOS) is a multi-factorial heterogeneous syndrome with complex pathologies that can occur when a woman's levels of the sex hormones estrogen and progesterone are out of balance. It affects up to 10% of those of childbearing age, and increases their risk of diabetes and cancer. There are some important questions about the pathogenesis of PCOS, the interrelationship between obesity and PCOS, and its etiology [191].

There is no specific test that can recognize PCOS definitively, and the rule-out method is the most common approach for PCOS diagnosis. Following the Rotterdam Criteria [192], the presence of at the least two characteristics from the following three are required to diagnose PCOS: 1) Oligo and/or anovulation, 2) clinical and/or biochemical signs of hyperandrogenism (e.g. non-classic congenital adrenal hyperplasia, hyperthyroidism, idiopathic hirsutism, familial hirsutism, Cushing's syndrome, androgen secretion) and 3) polycystic ovaries (e.g. cysts, ovarian hyperthecosis, stromal hyperthecosis). Other characteristics such as hyperprolactinemia can also be present [193]. Thus, the diagnosis of PCOS is exhausting for the patient, expensive and time consuming.

Chemerin is a chemoattractant protein known as Tazarotene-induced Gene 2, and acts as a ligand for the G-protein coupled receptor [194]. Chemerin levels correlate with insulin resistance and obesity [195-196], which are common comorbidities in PCOS [197-198]. Although some reports imply that there is a correlation between chemerin and PCOS, the role of chemerin and its contribution to PCOS pathogenesis is still under investigation [191]. Thus,

we decided to use SERS in conjunction with PCA to detect PCOS patients' samples, and to investigate the role chemerin plays in the disease using PLS.

There are various methods to measure chemerin levels. Enzyme-linked immunosorbent assay (ELISA) is a laboratory technique used to determine the concentration of chemerin in the ng/mL range, and has been used to measure serum chemerin levels in women with PCOS. The main drawback of this method is it cannot distinguish between the different chemerin isoforms [199-200]. Liquid chromatography/mass spectroscopy-mass spectroscopy (LC/MS-MS) is another technique to measure the chemerin level in serum [201]. One of most popular methods used to identify and quantify specific proteins is the Western Blot technique. While its popularity is due to its specificity (use of antibodies and molecular weight) and its ability to perform relative quantification, this method is time consuming (usually two days to obtain the results) and expensive due to specific reagent requirements and the cost of antibodies. Western Blot is also difficult to perform when the sample size is high, and numerous problems can arise (e.g. unusual or unexpected bands, absence of bands, weak signal, high background, uneven spots on the blot) which can lead to unexpected results. While these methods can all detect chemerin, they are highly complex and costly due to requirements for high-end equipment, skilled analysts and expensive tagged antibodies.

Though chemerin is a protein that has been analyzed and monitored with different techniques, this is the first time that SERS has been used. A collaboration between Dr. B. Tsang at the Ottawa Hospital and our group detected and measured the concentration in phosphate-buffered saline (PBS) and follicular fluid (FF). The simplicity of our setup, the highly informative SERS chemerin spectra, and the lower quantity of required clinical samples compared to a cuvette, encouraged us to pursue this research. In conjunction with PLS and PCA analysis techniques, this is a promising alternative to recognize PCOS and determine the chemerin contribution in its pathogenesis.

We begin by explaining the sample preparation, and then discuss differentiating between PCOS and non-PCOS using PCA analysis. The procedure for detecting chemerin in PBS and FF using the BVSPLS and IBVSPLS regression models is examined and compared to the

Western Blot method. Finally, we evaluate the role of chemerin in PCOS and non-PCOS patient samples using PCA and spectral analysis.

6.2 Experimental details

6.2.1 Capillary sample holder for SERS

The main goal of this study is to propose a simple, fast, inexpensive and accurate method to differentiate between PCOS and non-PCOS patient samples, and to measure the chemerin level in patient samples. Since FF is extracted from patients available samples are very limited, and one of the issues in this study is the minimum sample quantity required to record the Raman spectra. The Raman setup with a cuvette or HC-PCF sample holder requires a minimum of 1 to 3 mL, which is two orders of magnitude larger than the samples that can be collected from a patient. Therefore, we needed a sample container that does not require a large quantity of solution. Using a capillary as a sample holder is a good alternative to a cuvette or HC-PCF, and it requires less than 30 μL of sample. The capillary configuration is as discussed in Chapter 2, with the exception that it is used in a vertical configuration to reduce the sample quantity required. The one end open glass capillary with 25 mm length and 2 mm outer diameter is filled using microliter syringe. Although the sample consumption using a vertical capillary is lower than with cuvette or HC-PCF, the light-matter interaction is less effective. The proposed method is based on SERS, using a capillary as a container for small samples quantities.

6.2.2 Sample Preparation

The role and type of nanoparticles used here are the same as in Section 3.4, and the samples were prepared according to the method described by Leopold et al. [155]. We began this study using two pooled FF samples (one for PCOS and one for non-PCOS patients). The pool samples were collected from many patients, and are represented as an average. We then collected FF samples from twenty patients (10 PCOS and 10 non-PCOS), and used diluted chemerin in PBS to create the PLS model. PBS was chosen over FF due to the limited volume of FF that can be collected from patients. To further confirm the role of chemerin, we spiked pooled samples. Therefore the samples were divided into four groups;

1. Two unknown pooled FF samples from PCOS and non-PCOS patients.

2. Groups of samples that included FF from ten PCOS and ten non-PCOS patients.
3. Different levels of diluted chemerin in PBS. The 40 µL of 6.25 µM chemerin_PBS solution was used to prepare five different samples by diluting them with the solution. The chemerin concentration of these five samples was 6.25, 3.15, 1.56, 0.785 and 0.39 µM.
4. Pooled FF from PCOS and non-PCOS patients spiked with different amounts of recombinant chemerin (from 10 to 80 ng).

6.2.3 Multivariate Data Analysis

The SERS spectral datasets of PCOS and non-PCOS samples (the first, second and fourth group of samples) were used for PCA analysis to identify PCOS and non-PCOS samples, and the SERS spectral dataset of chemerin in PBS (the third sample group) was used to construct the calibration model for PLS analysis with Unscrambler® X version 10.3 (CAMO, Corvallis, OR, USA). The PLS model was constructed using baseline correction and TSV, to achieve a calibration model that can be evaluated against R^2, RMSEC and RMSEP. For data processing, the SERS spectra of chemerin_PBS (the third group of the dataset) was used to make a PLS model, and the chemerin concentrations of FF solutions (the second group of the dataset) were predicted using the model.

6.3 Results and discussion

The SERS spectra of the four sample groups were studied to determine specific peak(s) that can differentiate between PCOS and non-PCOS patient samples, and to evaluate the correlation between chemerin peaks and chemerin concentrations. Before recording the SERS spectra of the four sample groups, we first recorded the SERS spectra of nanoparticles and PBS (see Figure 6-1) to verify that they did not have specific Raman peak(s) in the samples due to PBS or NPs.

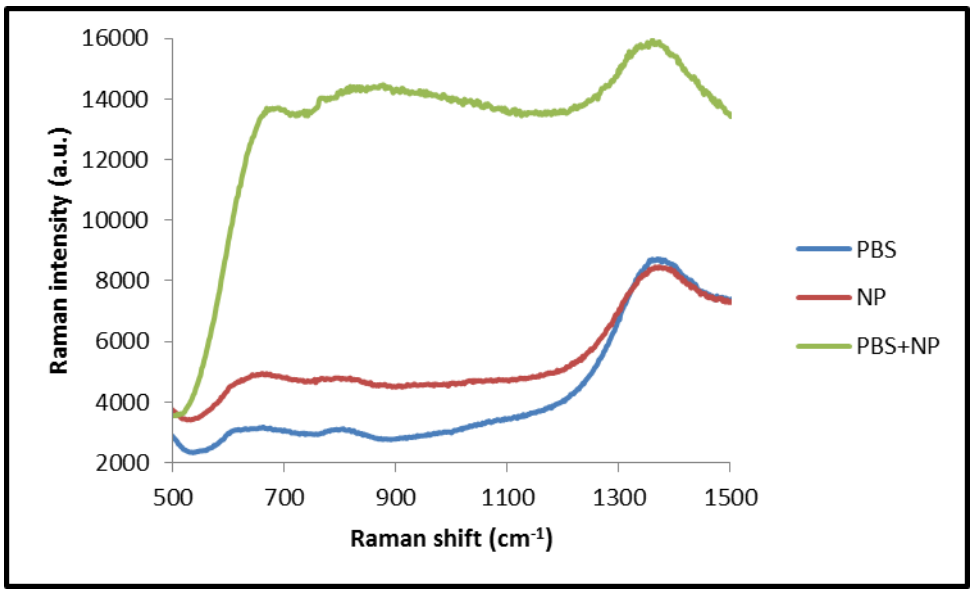

Figure 6-1 SERS spectra of PBS and nanoparticle solutions

6.3.1 Differentiating between PCOS and non-PCOS patients

This section investigates if our method can be used to differentiate between PCOS and non-COS patients. We first recorded the spectra of the first and second sample groups after adding 20 µL of nanoparticles to each sample. We initially worked with pool samples, which is common practice when the available samples are very low or highly expensive. The SERS spectra of the samples are shown in Figure 6-2. Both pooled PCOS and non-PCOS spectra showed some Raman peaks at the same wavenumbers, but with different intensities. PCA analysis was then used as a qualitative method on both the PCOS and non-PCOS data to reveal any hidden structures and clustering within the samples. Figure 6-3 shows the score plot of two components, and summarizes the variations in PCOS and non-PCOS of the first group of samples. The graph indicates that the recorded spectra of these samples are adequately separated.

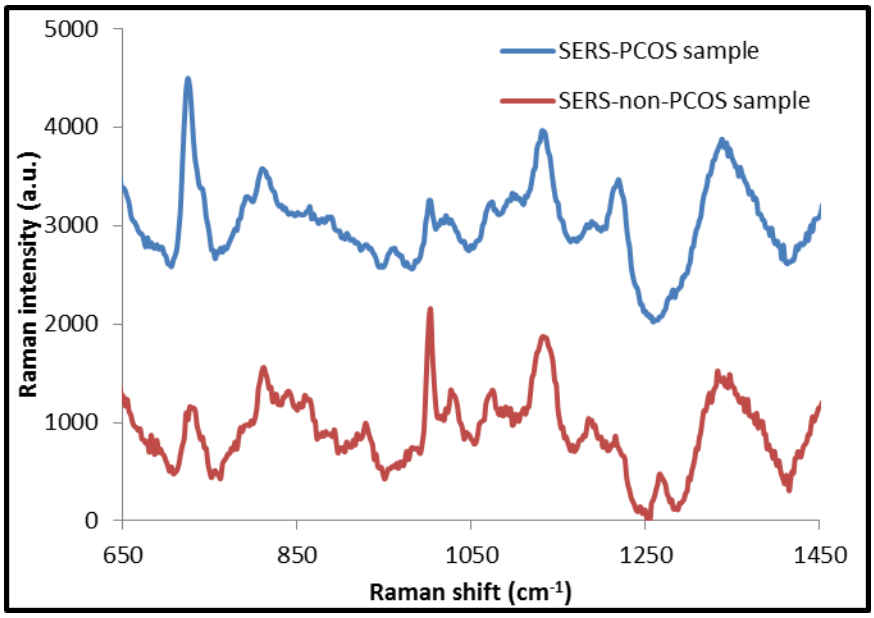

Figure 6-2 SERS spectra of chemerin_FF pooled sample

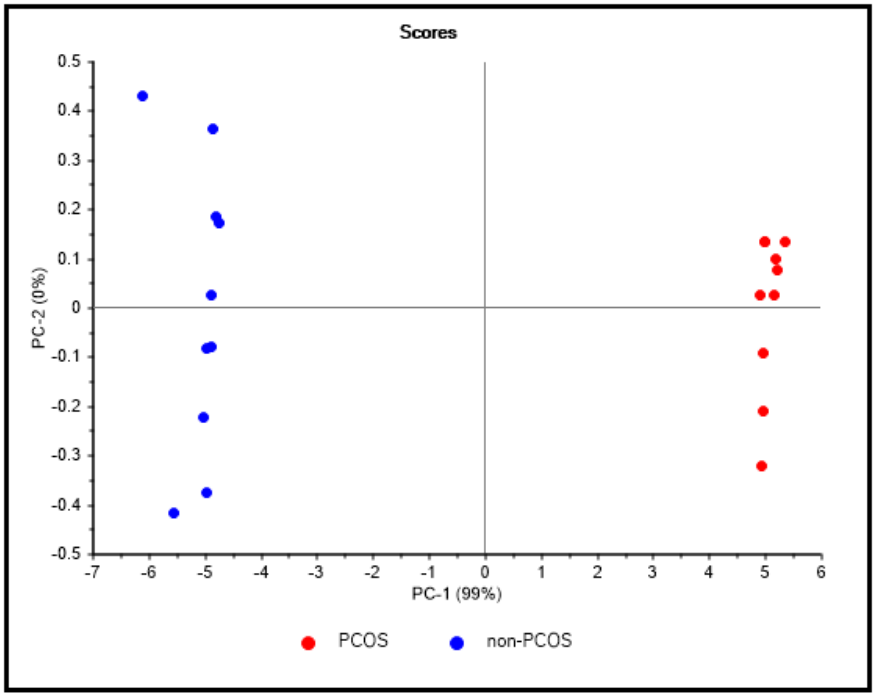

Figure 6-3 Score plot of principal component of PCOS and non-PCOS data of the first sample group

The verification of the SERS capability was followed by recording the SERS spectra of the second sample group (FF from 10 PCOS and 10 non-PCOS patients) from actual patients. The SERS spectra of these samples are shown in Figure 6-4. In the figure, there are SERS peaks at 691 cm^{-1} (C-S stretching lipids, methionine [185, 202]); 714 cm^{-1} (COO^- deformation glutamic acid [203], $N^+(CH_3)_3$ lipids [185]); 724 cm^{-1} (C-S-C asymmetric stretching; H_2O rock methionine [202]); 766 cm^{-1} (indole ring symmetric breathing tryptophan, CH_2 rocking and COO^- bending glutamic acid [202-205]); 781 cm^{-1} (W…O-Ct; W5…W1; N_t-C_α-Ct alanine [202], CO_2^- bending; CO_2 deformation valine [202]); 811 cm^{-1} (C-C-O symmetric stretching serine [202], C-C stretching collagen [188]); 838 cm^{-1} (CO_2^- ,$\gamma(CO_2^-)$ out of plane vibration leucine [202, 206], CH_2 rocking tryptophan [204], ring C-C symmetric stretching, Fermi resonance doublet, out-of-plane ring bending overtone tyrosine [203-205], deformative vibrations of amine groups [188]); 916 cm^{-1} (C_ε wagging; C_α-C_β lysine [202], C-C stretching phenylalanine [202], C-C-N stretching glutamic acid [202], C-C vibration pentanoic acid [207], C-COO^- and C-C stretching leucine [203, 206]); 942 cm^{-1} (γ(OH) aspartic acid [202], side chain N-C-N symmetric stretching; side chain C-C stretching arginine [202, 204], C-C stretching in α-helix conformation proline and valine [208-209]); 964 cm^{-1} (N_t-C_α; C_β-C_γ2; C_γ1-C_δ isoleucine [202], C_β wagging; N_t-C_α histidine [202], H-twisting on benzene ring tryptophan [202], dimer out-of-plane OH-O vibration, C-C stretching N-Methyl-D-aspartic acid [188, 203, 210]); 976 cm^{-1} (ring C-H out-of-plane bending phenylalanine [210], C-C stretching tryptophan [188]); 985 cm^{-1} (ring breathing proline [202], C_α–C_β, N_t- C_α , N-H stretching arginine [202, 211]); 1001 cm^{-1} (ring deformation phenylalanine [202], C-C symmetric ring breathing phenylalanine [184, 208], C-C and C-O stretching GLU [188], C-C stretching leucine [206], C-C deformation isophthalic acid [207]); 1020 cm^{-1} (indole ring breathing tryptophan [202-205], CH_2 twist, C-N stretch leucine [204, 206]); 1030 cm^{-1} (C-H in plane bending phenylalanine [202-205, 210, 212-213], C-C and C-N stretch N-methyl-D-aspartic acid [214], CH_2 twist leucine, C-N stretch leucine [204, 206], C-H deformation phenylalanine [209], N-C_α glycine [202], C_ε–Nζ, C_δ–C_ε lysine, deformation ring phenylalanine [202], ring breathing proline [202], C_γ–C_δ arginine [202]); 1072 cm^{-1} (N_t-C_α, C_α–C_β histidine [202], NH^+_3 rocking, C-N stretch leucine [202, 206], C-H formic acid [207]); 1091 cm^{-1} (C_γ–N_ε; N_ε-C_δ–H; N_ε-C_ε histidine [202], $C_\zeta N_\eta$1H2 asymmetric bend arginine [202], C-N stretch lipids [209], C-OH bend acetyl group, hyaluronic acid [215], C-N-H

asymmetric bend, COO^- and NH_2 vibrations arginine [202, 211]); 1129 cm^{-1} (C-N stretching proteins [184, 209], C-C, C-OH, C-N stretching, C-O-C glycosidic linkage protein [216], NH_3^+ wagging vibration aspartic acid [202]); 1186 cm^{-1} (C-H deformation, C-O stretching phenylalanine [202], NH_3^+ rocking leucine [202, 206], ring C-H in-plane bending phenylalanine [210]); 1215 cm^{-1} (γ (C-H), deformation ring phenylalanine [202]); 1246 cm^{-1} (amide III of collagen [202]); 1271 cm^{-1} ((=C-H) in plane bending, amide III protein, lipids [185], α-helix amide III collagen I protein [217], ring C-H in-plane bending phenylalanine [210], C-H deformation valine [202]); 1320 cm^{-1} (CH_3CH_2 twisting collagen [184], $N_tH_3^+$ asymmetric rocking, C_β-rock, C_t-C_α-H histidine [202], CH_2 wagging methionine [202], dimer C-O stretching, C_α–H bending, N-H rocking, C=O stretching, C_β-twist, C_tOO^- symmetric stretch, C_γ-rock, C_β-C_α-H_α; arginine [188, 202, 204, 210-211], C-C vibration isophthalic acid [207]) and 1334 cm^{-1} (N-C_α-H, NH_3 asymmetric rocking, N-C_α-H_α glycine [202]).

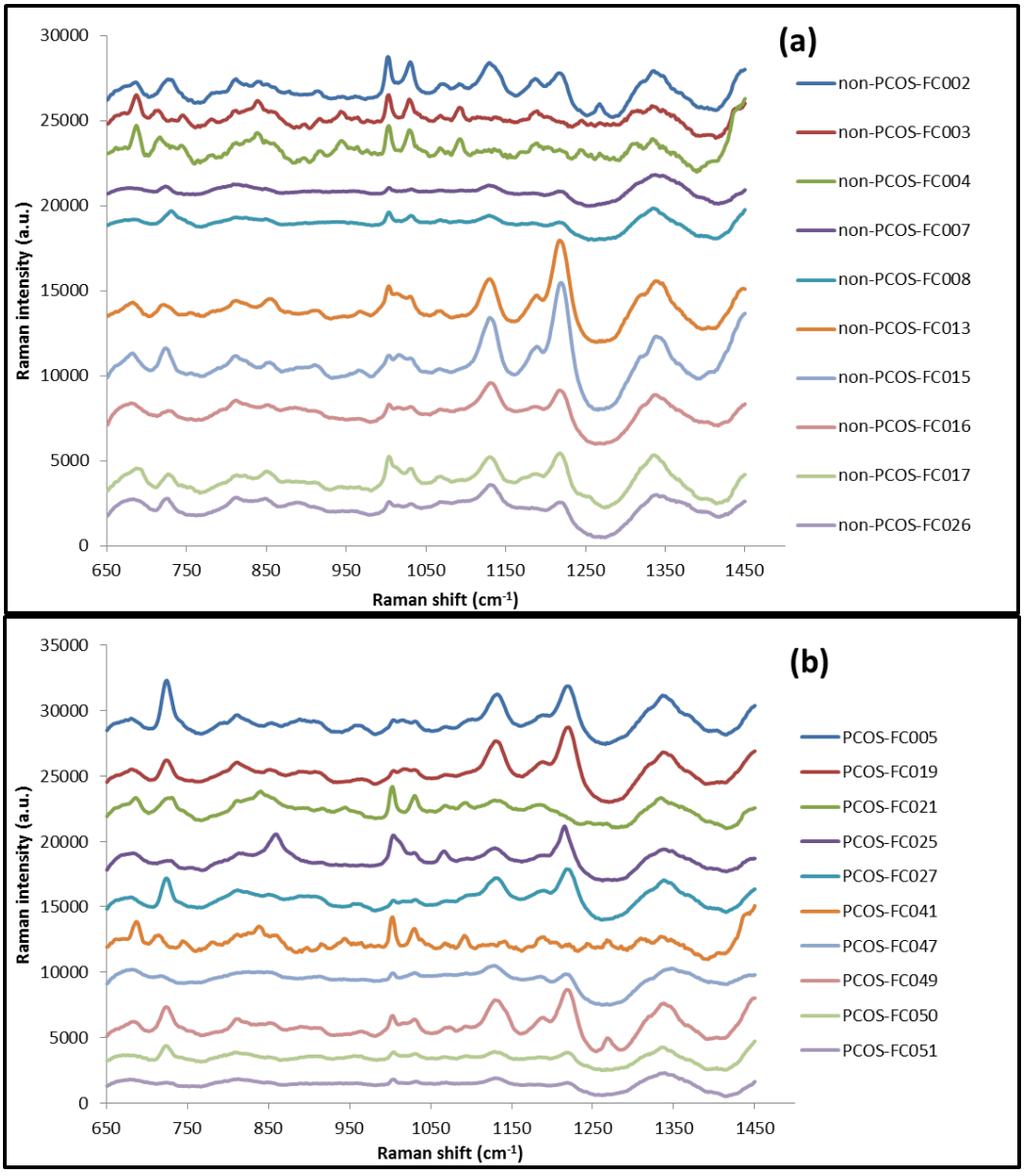

Figure 6-4 SERS spectra of chemerin_FF a) non-PCOS and b) PCOS patient samples

Based only on the spectrum, we could not differentiate between PCOS and non-PCOS patients. However, after applying PCA to the second sample group, the score plot identified two different clusters in the samples. These indicated that 80% of the PCOS samples were distinguishable from non-PCOS samples, as shown in Figure 6-5.

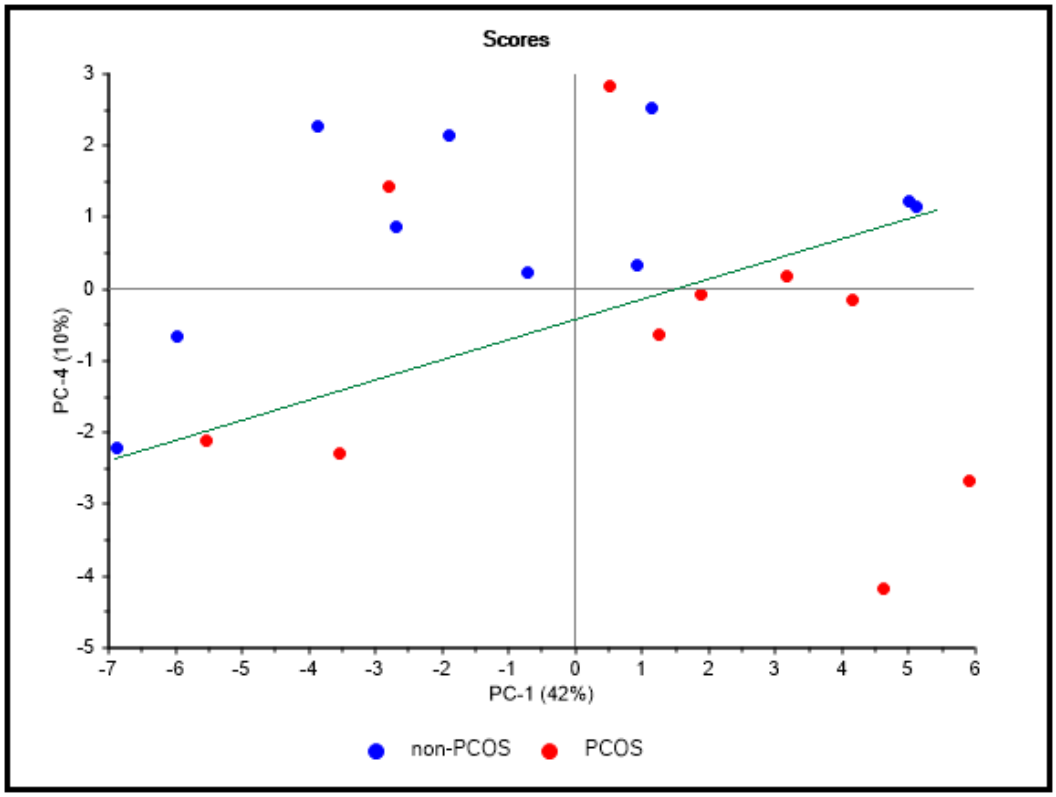

Figure 6-5 Score plot of principal component of PCOS and non-PCOS data of the second sample group

6.3.2 Investigating the role of chemerin in PCOS disease

The contribution of chemerin in the pathogenesis of PCOS patients is under investigation, particularly its level in PCOS and non-PCOS patients, and the etiology are still unclear. Using SERS allows us to quantify the role of chemerin in PCOS disease.

6.3.2.1 Using PLS to detect chemerin in PBS solution

To investigate the role of chemerin in PCOS disease, we began by determining the Raman spectrum of chemerin in PBS, a buffer solution commonly employed in biological research. Figure 6-6 illustrates the Raman spectra of different concentrations of chemerin in PBS when using a cuvette or HC-PCF as a container. As shown in the figure, the Raman spectrum of lower concentrations of chemerin is not informative, and thus cannot be used for low level chemerin detection. The spectrum of the highest chemerin concentration in PBS was 6.25 µM (not shown here), and did not reveal any specific peaks. This indicates that the Raman spectra of lower concentrations of chemerin in PBS cannot effectively develop a PLS model to predict chemerin

concentration in a patient sample. As a result, we chose to use the SERS technique in conjunction with PLS analysis.

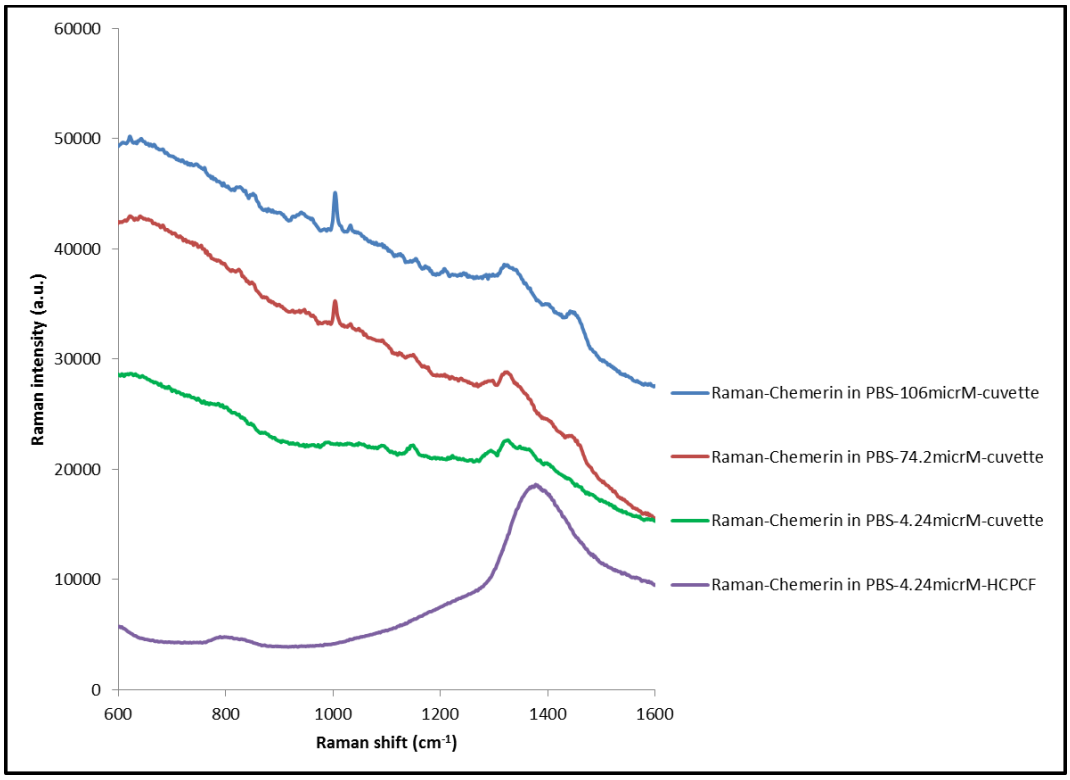

Figure 6-6 Raman spectra of chemerin in PBS

Using the third sample group, 20 µL of nanoparticles was added to prepare 3.15, 1.56, 0.785, 0.39, and 0.195 µM samples. The SERS spectra of these samples are shown in Figure 6-7. A spectral range of 650 to 1500 cm^{-1} was used in this analysis, and the figure indicates there are common peaks in chemerin in PBS and FF samples at 691, 714, 766, 985, 1001, 1020, 1030, 1072, 1091, 1215, 1271 and 1320 cm^{-1}. The peaks at 876 cm^{-1} (CH_2 rocking, C-N and N-H stretching arginine [204, 211], C-C stretching hydroxyproline [184, 217], C-C stretching, CH_2 rock methionine [202], C-C stretching aspartic acid [202], H-scissoring on indole ring tryptophan [202]); 1173 cm^{-1} (C_γ1-asymmetric bend, C_β-twist, C_γ1- C_β- C_γ2, C_δ-asymmetric rock isoleucine [202], NH_3 rocking, C-H deformation tyrosine, leucine [184-185], C-H stretching methionine [203]); 1224 cm^{-1} (PO_2^- asymmetric stretching nucleic acids [184]) and 1404 cm^{-1} (C_β-rock, C_γ-wagging, C_tOO^- symmetric stretch, N_t-C_α-H_α, arginine [202]) are due to chemerin.

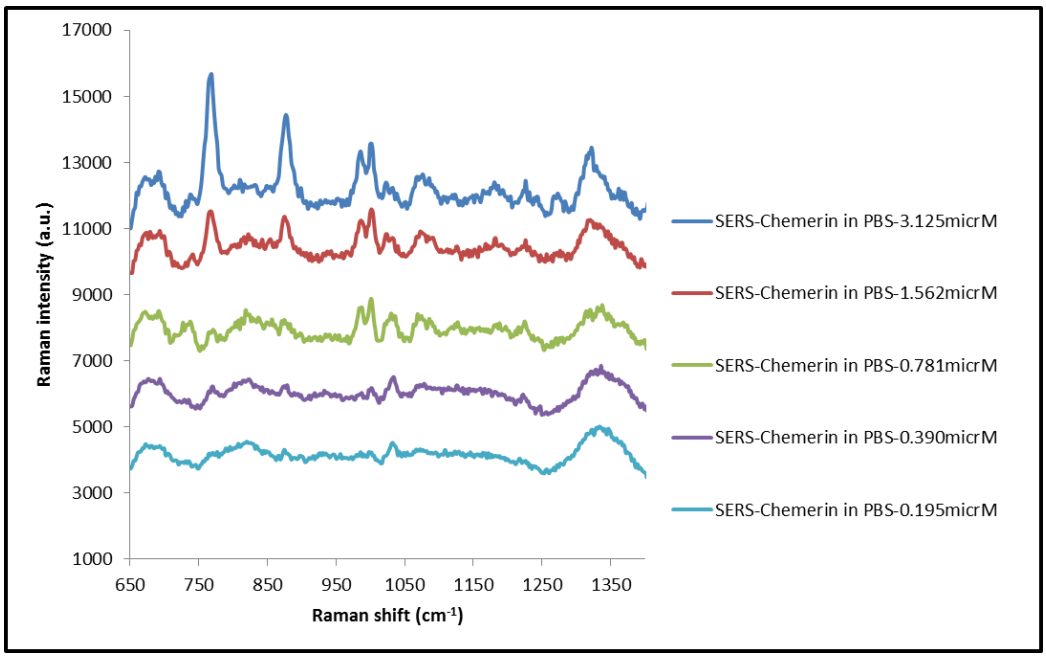

Figure 6-7 SERS spectra of chemerin in PBS samples

Figure 6-7 shows the correlation between chemerin peaks and concentrations. Fifty spectra of the third group of known samples (ten Raman spectra collected to ensure consistency of the replicated measurements) were used to build a PLS model, and twenty spectra of the second group were used to predict the chemerin concentration using the model. A TSV model was then developed using 67% of the known samples as a calibration set, and 33% as a validation set. These were chosen randomly by the Unscramble® software.

6.3.2.2 Loading and score plots

One result of PLS analysis is the loading plot, which indicates how the variables in different principal components correlate. The loading plot of the first and second PC is shown in Figure 6-8(a). According to the figure, the most important variables are 766, 876, 985 and 1001 cm^{-1}, which are the main chemerin peaks (see Figure 6-7). Another informative plot is the score plot, which indicates how different chemerin concentrations are explained by the variables. Figure 6-8(b) shows that 77% (X1 59%, X2 18%) of the X variance explains 99% (Y1 97%, Y2 2%) of the variation of chemerin concentrations in the samples. The figure also verifies that all the samples used to develop the PLS model are easily distinguishable from one another.

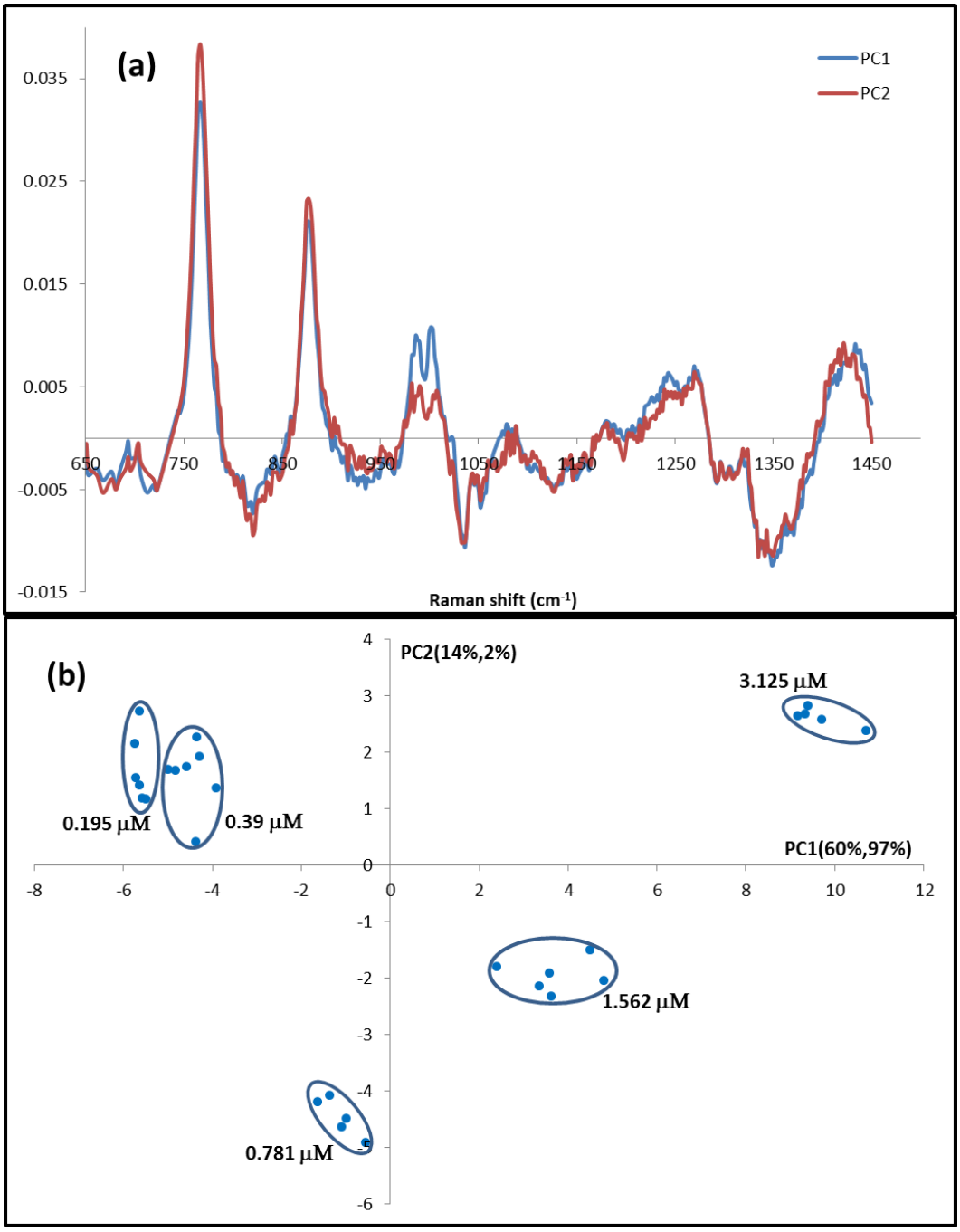

Figure 6-8 Loading plot (a) and score plot (b) of principal component of chemerin in PBS samples

6.3.2.3 PLS calibration model

According to the constructed PLS model, the R^2 (for calibration and validation), RMSEC and RMSEP for the spectral range of 650 to 1450 cm^{-1} are 0.99, 0.99, 0.016 and 0.126, respectively.

All these calculations are based on the TSV model using six PCs, which minimized the RMSEP value. The calibration curve of the developed PLS model is shown in Figure 6-9, while predictions of the chemerin concentration of PCOS and non-PCOS patient samples with this TSV model are shown in Table 6-1. The findings indicate that the PLS model with TSV validation can predict different concentrations of chemerin in FF, with standard deviations of 0.158 and 0.223 µM for non-PCOS and PCOS patient samples, respectively. This range of deviation with the SERS method was due to using the spectra of chemerin in PBS samples to predict the chemerin level of chemerin in FF samples, which is unavoidable. The correlation coefficient (R) between SERS and WB method was 0.81.

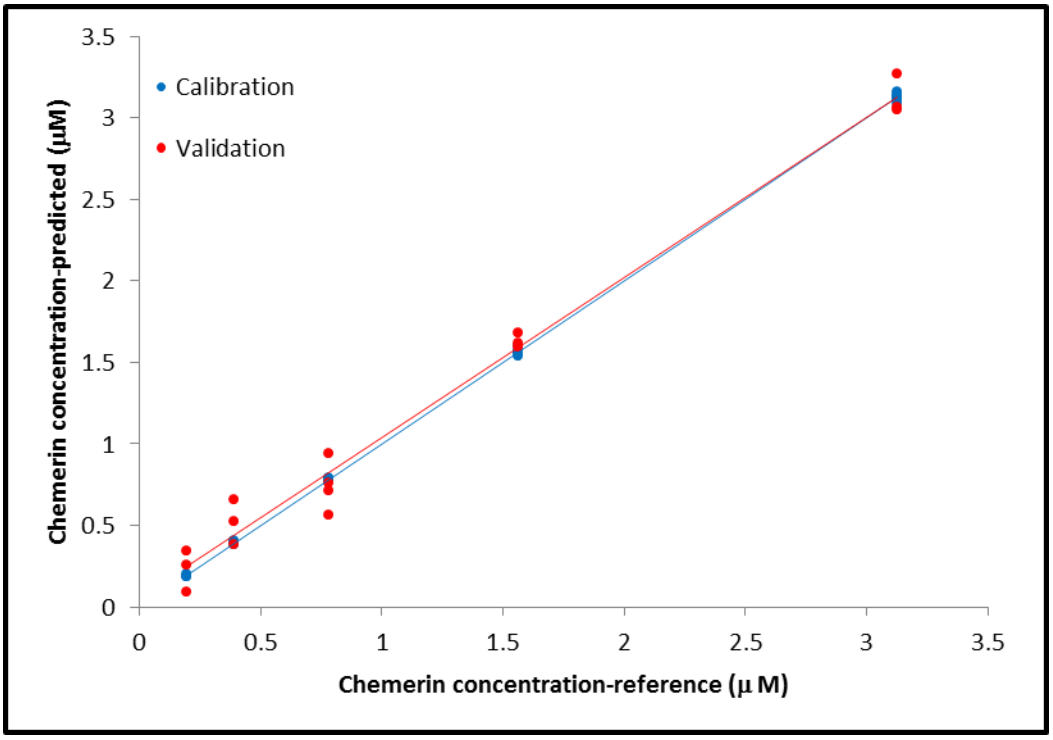

Figure 6-9 PLS regression model for of chemerin in PBS in 650 to 1450 cm^{-1} spectral range using test set validation

The average concentrations of chemerin of PCOS predicted by the SERS and Western Blot methods are 0.819 and 0.754 µM, respectively, and for non-PCOS samples they are 0.656 and 0.460. Although the predicted level of chemerin concentration using SERS is higher than with the Western Blot method, the differences between the average chemerin levels of PCOS and

non-PCOS samples with the SERS and Western Blot methods are very similar (0.163 and 0.294, respectively). This indicates that the chemerin level in PCOS and non-PCOS samples measured by the SERS and Western Blot methods are virtually equal, or in the same order of magnitude.

To improve the results of the regression process, we applied the new PLS regression analysis method introduced in Chapter 5. The R^2 (for calibration and validation) were as high as previous method at 0.99 and 0.99, respectively, while the RMSEC and RMSEP decreased to 0.003 and 0.049, respectively. This means that the LOD of chemerin with the SERS method is about 0.15 µM. According to this method, the average concentrations of chemerin in PCOS and non-PCOS predicted by SERS are 0.586, and 0.409 µM, respectively. Based on this method, the difference between the average chemerin levels of PCOS and non-PCOS samples with the SERS method is 0.177, which is better than previous results. This improvement is indicated by the standard deviations of measurements for non-PCOS and PCOS patient samples of 0.147 and 0.205 µM, respectively. This result shows that SERS accuracy is as good as Western Blot method. This method takes less than hour (including data processing) while Western Blot method takes almost 2 days.

Table 6-1 Chemerin concentration predictions in 20 non-PCOS and PCOS samples

Sample no.	Chemerin concentration (µM) in FF (SERS)-BVSPLS		Chemerin concentration (µM) in FF (SERS)-IBVSPLS		Chemerin concentration (µM) in FF -WB	
	Prediction	STDV	Prediction	STDV	Prediction	STDV
Non-PCOS-2	0.961	0.158	0.689	0.147	0.722	0.154
Non-PCOS-3	0.547		0.310		0.338	
Non-PCOS-4	0.534		0.314		0.453	
Non-PCOS-7	0.448		0.251		0.428	
Non-PCOS-8	0.878		0.642		0.459	
Non-PCOS-13	0.734		0.476		0.607	
Non-PCOS-15	0.616		0.363		0.344	
Non-PCOS-16	0.640		0.355		0.299	
Non-PCOS-17	0.597		0.322		0.658	
Non-PCOS-26	0.610		0.370		0.293	
PCOS-5	0.717	0.223	0.575	0.205	0.479	0.276
PCOS-19	0.614		0.398		0.766	
PCOS-21	1.250		0.984		1.304	
PCOS-25	0.649		0.408		0.459	
PCOS-27	0.758		0.607		0.600	

PCOS-41	0.981	0.706	0.824
PCOS-47	0.617	0.444	0.709
PCOS-49	1.040	0.743	1.003
PCOS-50	0.953	0.694	0.946
PCOS-51	0.611	0.301	0.453

6.3.2.4 Spiking pooled samples with chemerin

Based on the results of chemerin in PBS, we believe that chemerin is involved in PCOS disease. To further confirm this, we spiked pool samples with chemerin, and studied the impact (the fourth sample group). Both the PCOS and non-PCOS pool samples were spiked with 10, 20, 40, 60 and 80 ng of chemerin dissolved in PBS. The SERS spectra of all the spiked samples were recorded after adding 20 µL of nanoparticles to each sample, as shown in Figures 6-10.

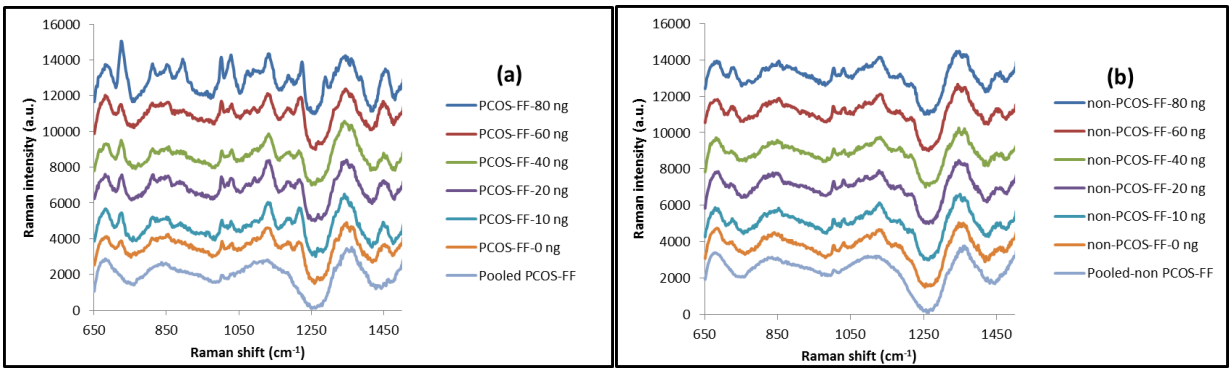

Figure 6-10 Baseline corrected SERS spectra of a) PCOS and b) non-PCOS patient samples spiked with chemerin

According to Figure 6-10(a), there is a correlation between the Raman intensity at different wavenumbers and the amount of chemerin in the PCOS sample. The main wavenumbers that show higher intensities and more correlation between the Raman peaks and the amount of chemerin are 724, 1001, 1028 and 1224 cm^{-1} (due to chemerin_PBS or FF), while the other peaks do not show this correlation clearly. The correlation also can be seen in the Raman spectra of non-PCOS spiked samples, as shown in Figure 6-10(b). However, this is not as evident in PCOS spectra, as the intensities of the Raman peaks of non-PCOS samples are less than PCOS samples.

In Figure 6-10, the comparisons between PCOS and non-PCOS samples spiked with the same amount of chemerin reveal that all the PCOS sample Raman peaks, particularly 724 cm^{-1}, have higher intensities than the Raman peaks of the non-PCOS samples, as shown in Figure 6-2.

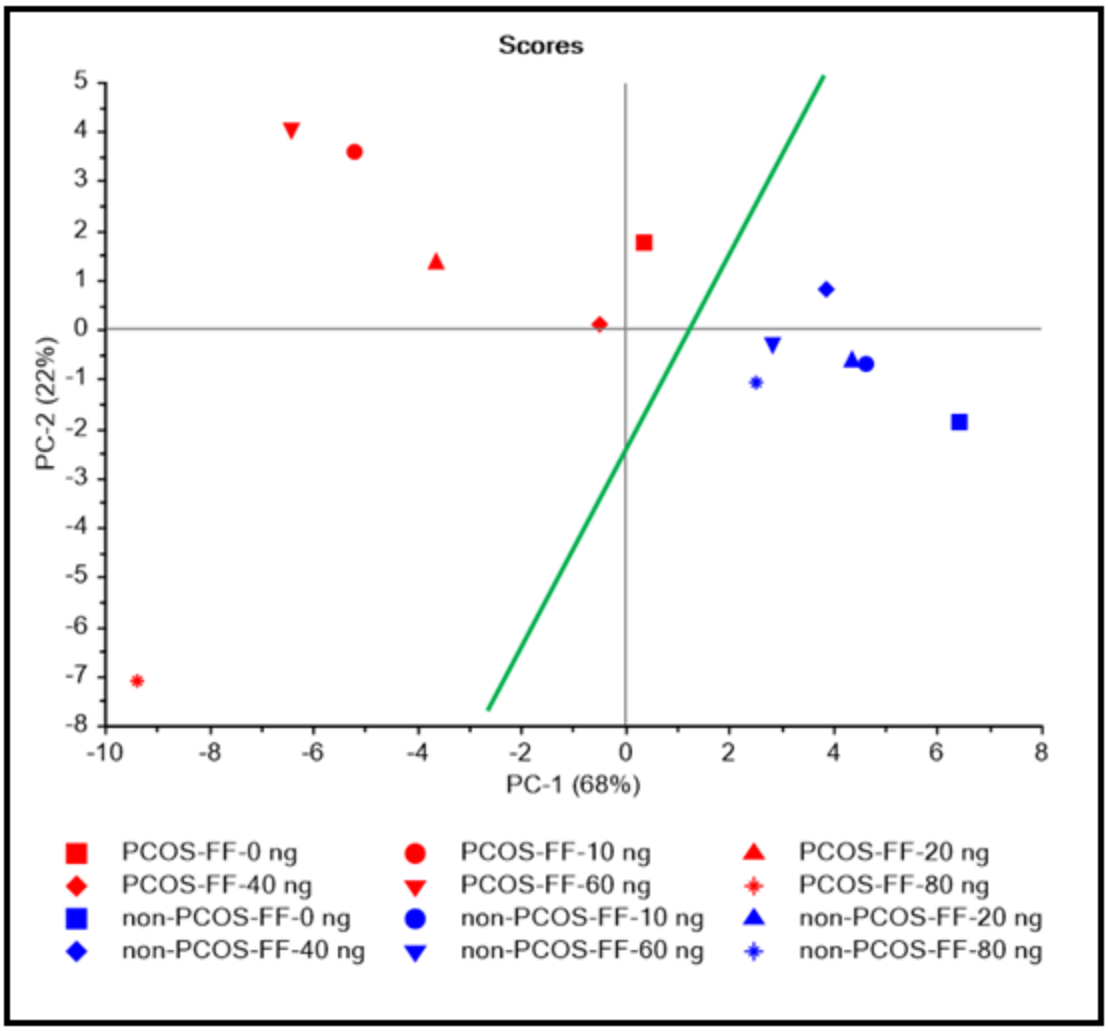

Figure 6-11 Score plot of principal components of PCOS and non-PCOS data of the fourth sample group

Figure 6-11 shows the result of applying PCA to the fourth sample group. All the spiked PCOS samples are clearly separated from the spiked non-PCOS samples, which verifies that the Raman method with PCA can be used to distinguish PCOS patient samples from non-PCOS patient samples.

6.4 Conclusion

We have demonstrated that our SERS setup using a capillary container with PCA is a fast and accurate alternative method to detect PCOS disease. Our setup was also used to investigate the role of chemerin in PCOS disease, and to measure the concentration of chemerin in PBS and FF samples. The IBVSPLS regression method provided reliable estimates of chemerin concentrations, and reduced the LOD of chemerin in FF patient samples to approximately 0.15 μM.

Chpater 7. Summary and future work

7.1 Summary

The main objective of this thesis is to improve the LOD of Raman biosensors by using SERS and/or HC-PCF in conjunction with statistical methods. To summarize:

- In Chapter 2, the background of Raman spectroscopy, cuvette/capillary, HC-PCF sampling and their advantages and limitations were reviewed. The chapter continued with an explanation of SERS theory and its recent progress. Finally, the importance and requirements of MVA, MVA categories, PCA and PLS were explained.

- In Chapter 3, Raman and SERS techniques, in conjunction with PLS and PCA, were implemented using cuvette as a simple sampling method. The power of PLS models to improve the LOD of heparin in serum at clinical level was evaluated. The role of PLS and PCA statistical methods to detect clinical levels of GLU and GABA in serum using SERS technique with cuvette sampling was experimentally evaluated.

- In Chapter 4, SERS based HC-PCF sampling using PLS data processing was used to detect leukemia cells at a clinical level. This also revealed that live, apoptotic and necrotic leukemia cells were distinguishable using PCA data processing. In this chapter, I have shown that the LOD of experiment with an integrated HC-PCF with nanoparticles is improved using PLS and PCA statistical methods.

- In Chapter 5, a new method to improve the LOD of Raman biosensors by optimizing the PLS regression model was introduced. The backward variable selection method for PLS (BVSPLS) as a wrapper technique was improved by choosing suitable variables among the sorted regression coefficients. The improved PLS model was evaluated by recalculating heparin prediction (Chapter 3) and leukemia prediction (Chapter 4).

- In Chapter 6, SERS with capillary sampling was used in conjunction with PCA to distinguish between PCOS and non-PCOS patient samples, and the role of chemerin in PCOS patient

samples was confirmed using the PCA technique. The proposed method in Chapter 5 was used to measure the chemerin concentration of FF samples, and it validated that the improved PLS technique can reduce the LOD of chemerin in FF patient samples.

7.2 Future work

This section discusses potential future research.

7.2.1 Verification of new optimization with other data set

The new optimization procedure was examined using heparin and leukemia cell datasets. Although only these Raman datasets were applied to verify the performance of the IBVSPLS method, the technique is an effective general approach, and could be tested with other datasets from different research fields.

7.2.2 Verification of new optimization with large data sets

The IBVSPLS is a relatively time consuming method that needs to be addressed. New IBVSPLS datasets could be selected from different dataset sizes, to reveal how the LOD improvement depends on the size of dataset. This is an important question regarding the capability of the method, and how it increases processing time with a large dataset compared to other relevant methods.

7.2.3 Implementation of new optimization with MATLAB code

The multivariate analysis has many steps, from sample arrangement and preprocessing to regression analysis, principle component analysis and optimization procedures. We used Unscramble software to preprocess or create PLS and PCA models, after which IBVSPLS was applied as an optimization procedure to improve the LOD of the analyte. We use this software manually in different steps of the method, which is time consuming. In future work, we could implement this optimization procedure with MATLAB code, to decrease the computation costs.

7.2.4 PCOS diagnosis using HC-PCF

In Chapter 6, the detection of PCOS and non-PCOS patient samples was manipulated using a simple capillary. The main reason we did not use the integrated HC-PCF with a differential

pressure system (as discussed in Section 4.2.2.3) was because it required a large sample volume. Filling a 10 cm HC-PCF requires approximately 1 μL of solution, while the leading tube and glass vial containers need 1 to 3 mL. In future work on this type of Raman biosensor application, we could miniaturize the leading tube and sample container to minimize the required sample volume. The main challenge with this configuration is how to apply and control the sample flow rate during Raman recording, which enables reproducibility of recording for any sample, and simplifies the multivariate analysis.

References

[1] C. V. Raman and K. S. Krishnan, “A new type of secondary radiation,” Science, 121, 501-502 (1928).

[2] R. L. McCreery, “Raman spectroscopy for chemical analysis, Vol 157 in Chemical analysis,” New York: John Wiley & Sons (2000).

[3] M. Fleischman, P. Hendra, and A. McQuillan, “Raman spectra of pyridine adsorbed at a silver electrode,” Chem. Phys. Lett., 26(2), 163-166 (1974).

[4] J. T. Motz, M. Fitzmaurice, A. Miller, S. J. Gandhi, A. S. Haka, L. H. Galindo, R. R. Dasari, J. R. Kramer, and M. S. Feld, “In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque,” J. Biomed. Opt., 11(2), 021003 (2006).

[5] A. Campion and P. Kambhampati, “Surface-enhanced Raman scattering,” Chem. Soc. Rev., 27(4), 241-250 (1998).

[6] R. Lewandowska, “Classical or transmission Raman, SERS or TERS- which Raman spectroscopy technique Is right for you,” Spectroscopy, 28(6), 32-42 (2013).

[7] L. Hecht, A. L. Phillips, and L. D. Barron, “Determination of enantiomeric excess using Raman optical activity,” J. Raman Spectrosc., 26(8-9), 727-732 (1995).

[8] J. Popp and W. Kiefer, “Raman scattering, fundamentals,” Encyclopedia of Analytical Chemistry, Wiley Online Library (2006).

[9] A. Khetani, A. Momenpour, V. S. Tiwari, and H. Anis, “Surface enhanced raman scattering (SERS) using nanoparticles,” in Silver nanoparticle applications in the fabrication and design of medical and biosensing devices, E. I. Alarcon, M. Griffith, and K. I. Udekwu (Editors), New York: Springer (2015).

[10] E. Marcatili and R. Schmeltzer, “Hollow metallic and dielectric waveguides for long distance optical transmission and lasers (long distance optical transmission in hollow dielectric and metal circular waveguides, examining normal mode propagation),” Bell Syst. Tech. J., 43, 1783-1809 (1964).

[11] W. R. Trutna and R. L. Byer, “Multiple-pass Raman gain cell,” Appl. Opt., 19(2), 301-312 (1980).

[12] A. Khetani, V. S. Tiwari, A. Harb, and H. Anis, “Monitoring of heparin concentration in serum by Raman spectroscopy within hollow core photonic crystal fiber,” Opt. Express, 19(16), 15244–15254 (2011).

[13] C. M. Smith, N. Venkataraman, M. T. Gallagher, D. Müller, J. A. West, N. F. Borrelli, D. C. Allan, and K. W. Koch, “Low-loss hollow-core silica-air photonic bandgap fiber,” Nature, 424, 657-659 (2003).

[14] J. Irizar, J. Dinglasan, J. B. Goh, A. Khetani, H. Anis, D. Anderson, C. Goh, and A. S. Helmy, “Raman spectroscopy of nanoparticles using hollow-core photonic crystal fibers,” IEEE J. Sel. Top. Quant., 14(4), 1214-1222 (2008).

[15] X. Yang, T. C. Bond, J. Z. Zhang, Y. Li, and C. Gu, “Photonic crystal fiber Raman sensors,” Proc. SPIE, 8559, 855902-1 (2012).

[16] G. Antonopoulos, F. Benabid, T. A. Birks, D. M. Bird, J. C. Knight, and P. St. J. Russell, “Experimental demonstration of the frequency shift of bandgaps in photonic crystal fibers due to refractive index scaling,” Opt. Express, 14(7), 3000-3006, (2006).

[17] N. thang, “Stimulated raman scattering in gas filled hollow-core photonic crystal fiber,” Diss., Max Plank Institute, Germany (2013).

[18] F. M. Cox, A. Argyros, and M. C. J. Large, “Liquid-filled hollow core microstructured polymer optical fiber,” Opt. Express, 14(9), 4135-4140 (2006).

[19] A. Khetani, “Photonic crystal fiber as a biosensor,” Diss., University of Ottawa, Canada (2008).

[20] A. Khetani, V. S. Tiwari, A. Harb, and H. Anis, “Monitoring of heparin concentration in serum by Raman spectroscopy within hollow core photonic crystal fiber,” Opt. Express, 19(16), 15244-15254 (2011).

[21] S. Abalde-Cela, P. Aldeanueva-Potel, C. Mateo-Mateo, L. Rodrı´guez- Lorenzo, R. A. Alvarez-Puebl, and L. M. Liz-Marzan, “Surface-enhanced raman scattering biomedical applications of plasmonic colloidal particles,” J. R. Soc. Interface, 7(suppl 4), S435–S450 (2010).

[22] X. M. Qian and S. M. Nie, “Single-molecule and single-nanoparticle SERS: From fundamental mechanisms to biomedical applications,” Chem. Soc. Rev., 37(5), 912–920 (2008).

[23] Y. Oh, S. Park, M. Kang, J. H. Choi, Y. Nam, and K. Jeong, “Beyond the SERS: Raman enhancement of small molecules using nanofluidic channels with localized surface plasmon resonance,” Small, 7(2), 184–188 (2011).

[24] C. J. Choi, Z. Xu, H. Y. Wu, G. L. Liu, and B. T. Cunningham, “Surface-enhanced nano domes,” Nanotechnology, 21(41), 415301–415307 (2010).

[25] M. Moskovits, ”Surface-enhanced spectroscopy,” Rev. Mod. Phys., 57(3), 783-826 (1985).

[26] S. M. Morton and L. Jensen, “Understanding the molecule-surface chemical coupling in SERS,” J. Am. Chem. Soc., 131(11), 4090-4098 (2009).

[27] L. Jensen, C. M. Aikens, and G. C. Schatz, “Electronic structure methods for studying surface-enhanced Raman scattering,” Chem. Soc. Rev., 37(5), 1061-1073 (2008).

[28] P. L. Stiles, J. A. Dieringer, N. C. Shah, and R. P. Van Duyne, “Surface-enhanced Raman spectroscopy,” Ann. Rev. Anal. Chem., 1(1), 601-626 (2008).

[29] M. E. Stewart, C. R. Anderton, L. B. Thompson, J. Maria, S. K. Gray, J. A. Rogers, and R. G. Nuzzo, “Nanostructured plasmonic sensors,” Chem. Rev., 108(2), 494-521 (2008).

[30] C. J. Orendroff, L. Gearheart, N. R. Jana, and C. J. Murphy, “Aspect ratio dependence on surface-enhanced Raman scattering using silver and gold nanorod substrates,” Phys. Chem. Chem. Phys., 8(1), 165-170 (2006).

[31] C. J. Orendroff, A. Gole, T. K. Sau, and C. J. Murphy, “Surface-enhanced Raman spectroscopy of self-assembled monolayers: sandwich architecture and nanoparticle shape dependence,” Anal. Chem., 77(10), 3261-3266 (2005).

[32] G. McNay, D. Eustace, W. E. Smith, K. Faulds, and D. Graham, “Surface-enhanced Raman scattering (SERS) and surface-enhanced resonance Raman scattering (SERRS): A review of applications,” Appl. Spectrosc., 65(8), 825-837 (2011).

[33] J. M. Chalmers and P. R. Griffiths “Handbook of vibrational spectroscopy,” New York: John Wiley & Sons (2002).

[34] K. L. Kelly, E. Coronado, L. L. Zhao , and G. C. Schatz, “The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment,” J. Phys. Chem. B, 107(3), 668-677 (2003).

[35] G. A. Baker and D. S. Moore, “Progress in plasmonic engineering of surface-enhanced Raman scattering substrates toward ultra-trace analysis,” Anal. Bioanal. Chem., 382(8), 1751-1770 (2005).

[36] X. Zhang, M. A. Young, O. Lyandres, and R. P. Van Duyne, “Rapid detection of an anthrax biomarker by surface-enhanced Raman spectroscopy,” J. Am. Chem. Soc., 127(12), 4484-4489 (2005).

[37] R. Bogue, “Nanosensors: A review of recent progress,” Sensor Rev., 28(1), 12-17 (2008).

[38] H. Liang, Z. Li, W. Wang, Y. Wu, and H. Xu, “Highly surface-roughened ‘flower-like’ silver nanoparticles for extremely sensitive substrates of surface-enhanced Raman scattering,” Adv. Mater., 21(45), 4614-4618 (2009).

[39] M. Baia, L. Baia, and S. Astilean, “Gold nanostructured films deposited on polystyrene colloidal crystal templates for surface-enhanced Raman spectroscopy,” Chem. Phys. Lett. 404(1-3), 3-8 (2005).

[40] R. Zhang, B. Xu, X. Liu, Y. Zhang, Y. Xu, Q. Chen, and H. Sun, “Highly efficient SERS test strips,” Chem. Commun., 48(47), 5913-5915 (2012).

[41] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, and J. P. Remon, “Applications of Raman spectroscopy in pharmaceutical analysis,” TrAC-Trend. Anal. Chem., 21(12), 869-877 (2002).

[42] S. Sasic, “Pharmaceutical applications of Raman spectroscopy,” New Jersey: John Wiley & Sons (2008).

[43] R. Pandey, S. K. Paidi, J. W. Kang, N. Spegazzini , R. R. Dasari, T. A. Valdez, and I. Barman, “Discerning the differential molecular pathology of proliferative middle ear lesions using Raman spectroscopy,” Scientific Reports, 5, Article ID 13305 (2015).

[44] P. C. A. M. Buijtels, H. F. M. Willemse-Erix, P. L. C. Petit, H. P. Endtz, G. J. Puppels, H. A. Verbrugh, A. van Belkum, D. van Soolingen, and K. Maqueli, “Rapid identification of mycobacteria by Raman spectroscopy,” J. Clin. Microbiol., 46(3), 961–965 (2008).

[45] J. Moros, S. Garrigues, and M. de la Guardia, “Evaluation of nutritional parameters in infant formulas and powdered milk by Raman spectroscopy,” Anal. Chim. Acta, 593(1), 30–38 (2007).

[46] K. Kong, C. Kendall, N. Stone, and I. Notingher, “Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection,” Adv. Drug Deliv. Rev., 89, 121–134 (2015).

[47] D. Yang and Y. Ying, “Applications of Raman spectroscopy in agricultural products and food analysis: A review,” Appl. Spectrosc. Rev., 46(7), 539-560 (2011).

[48] R. M. Seifar, J. M. Verheul, F. Ariese, Udo A. Th. Brinkman, and C. Gooijer, “Applicability of surface-enhanced resonance Raman scattering for the direct discrimination of ballpoint pen inks,” Analyst, 126(8), 1418-1422 (2001).

[49] A. Raza and B. Saha, “Application of Raman spectroscopy in forensic investigation of questioned documents involving stamp inks,” Sci. Justice, 53(3), 332-8 (2013).

[50] R. A. Halvorson and P. J. Vikesland, “Surface-enhanced raman spectroscopy (SERS) for environmental analyses,” Environ. Sci. Technol., 44(20), 7749–7755 (2010).

[51] P. Vandenabeele, H. G. M. Edwards, and L. Moens, “A decade of Raman spectroscopy in art and archaeology,” Chem. Rev., 107(3), 675-686 (2007).

[52] P. Vandenabeele, H. G. M. Edwards, and J. Jehlička, “The role of mobile instrumentation in novel applications of Raman spectroscopy: archaeometry, geosciences, and forensics,” Chem. Soc. Rev., 43(8), 2628-2649 (2014).

[53] I. R. Lewis and H. G. M. Edwards, “Handbook of Raman spectroscopy from the research laboratory to the process line,” New York: Marcel Dekker (2001).

[54] K. R. Beebe, R. J. Pell, and M. B. Seasholtz, “Chemometrics: A practical guide,” New York: John Wiley & Sons (1998).

[55] K. H. Esbensen, “Multivariate data analysis - in practice,” Oslo: CAMO (2004).

[56] J. A. Hartigan, “Clustering algorithms,” New York: John Wiley & Sons (1975).

[57] M. Miljkovic, T. Chernenko, M. J. Romeo, B. Bird, C. Matthäus, and M. Diem, “Label-free imaging of human cells: algorithms for image reconstruction of Raman hyperspectral datasets,” Analyst, 135(8), 2002-2013 (2010).

[58] R. Xu and D. Wunsch II, “Survey of clustering algorithms,” IEEE T. Neural Networ., 16(3), 645-678 (2005).

[59] T. P. Wrobel, L. Mateuszuk, S. Chlopicki, K. Malek, and M. Baranska, “Imaging of lipids in atherosclerotic lesion in aorta from ApoE/LDLR-/- mice by FT-IR spectroscopy and hierarchical cluster analysis,” Analyst, 136(24), 5247-5255 (2011).

[60] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, “Review of multidimensional data processing approaches for Raman and infrared spectroscopy,” EPJ Tech. Instrum., 2(8), 1-38 (2015).

[61] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, 1995, 20(3), 273-297

[62] J. A. F. Pierna, V. Baeten, and P. Dardenne, “Screening of compound feeds using NIR hyperspectral data,” Chemometr. Intell. Lab., 84(1-2), 114–118 (2006).

[63] J. Suykens, “Advances in learning theory: Methods, models and applications”, Proc. of the NATO Advanced Study Institute on Learning Theory and Practice, Belgium, (2003).

[64] R. Bro and A. K. Smilde, “Centering and scaling in component analysis,” J. Chemometr., 17(1), 16-33 (2003).

[65] A. Hoskuldson, “Pls regression methods,” J. Chemometr., 2(3), 211-228 (1998).

[66] M. S. Lewis-Beck, A. Bryman, and T. F. Liao, “The SAGE encyclopedia of social sciences research methods,” California: SAGE (2003).

[67] H. Wold, “Nonlinear estimation by iterative least squares procedures,” in Research Papers in Statistics, F. David (Editor), New York: John Wiley & Sons (1966).

[68] F. Lindgren, P. Geladi, and S. Wold, “The kernel algorithm for PLS,” J. Chemometr., 7(1), 45–59 (1993).

[69] S. d. Jong “SIMPLS: An alternative approach to partial least squares regression,” Chemometr. Intell. Lab., 18(3), 251-263 (1993).

[70] S. Rännar, F. Lindgren, P. Geladi, and S. Wold, “A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm,” J. Chemometr., 8(2), 111-125 (1994).

[71] B. S. Dayal and J. F. McGregor, “Improved PLS algorithms,” J. Chemometr., 11(1), 73-85 (1997).

[72] G. Golub and W. Kahan, “Calculating the singular values and pseudo-inverse of a matrix,” J. Soc. Ind. Appl. Math., 2(2), 205-224 (1965).

[73] M. Andersson, “A comparison of nine PLS1 algorithms,” J. Chemometr., 23(10), 518-529 (2009).

[74] R. E. Shaffer, G. W. Small, and M. A. Arnold, “Genetic algorithm-based protocol for coupling digital filtering and partial least-squares regression: Application to the near-infrared analysis of glucose in biological matrices,” Anal. Chem., 68(15), 2663-2675 (1996).

[75] J. F. Baldrich, “Experimental design applied to the selection of samples and sensors in multivariate calibration,” Diss., Universitat Rovira I Virgili, Spain (1997).

[76] J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. van Velzen, J. P. M. van Duijnhoven, and F. A. van Dorsten, “Assessment of PLSDA cross validation,” Metabolomics, 4(1), 81–89 (2008).

[77] C. C. Aggarwal, “Outlier analysis,” New York: Springer (2013).

[78] V. Barnett and T. Lewis, “Outliers in statistical data,” Chichester: John Wiley & Sons (1994).

[79] J. W. Osborne and A. Overbay, “The power of outliers (and why researchers should ALWAYS check for them),” Practical Assessment, Research & Evaluation, 9(6), (2004).

[80] V. J. Hodge and J. Austin, “A survey of outlier detection methodologies,” Artif. Intell. Rev., 22(2), 85-126 (2004).

[81] A. J. Miller, “Selection of subsets of regression variables,” J. R. Statist. Soc. A, 147(3), 389-425 (1984).

[82] G. H. John, R. Kohavi, and K. Pfleger, “Irrelevant features and the subset selection problem,” Proceedings of the Eleventh International Conference on Machine Learning, New Brunswick, New Jersey, 121-129 (1994).

[83] T. Mehmood, K. H. Liland, L. Snipen, and S. Saeb, “A review of variable selection methods in partial least squares regression,” Chemometr. Intell. Lab., 118, 62-69 (2012).

[84] Y. Saeys, I. Inza, and P. Larranaga, “A review of feature selection techniques in bioinformatics,” Bioinformatics, 23(19), 2507–2517 (2007).

[85] I. E. Frank, “Intermediate least squares regression method,” Chemometr. Intell. Lab., 1(3), 233-242 (1987).

[86] A. G. Frenich, D. Jouan-Rimbaud, D. L. Massart, S. Kuttatharmmakul, M. M. Galera and J. L. M. Vidal, “Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares,” Analyst, 120(12), 2787-2792 (1995).

[87] H. Martens and M. Martens, “Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR),” Food Qual. Prefer., 11(1-2), 5-16 (2000).

[88] B. Molleer, “Near infrared transmission spectra of barley of malting grade represent a physical-chemical fingerprint of the sample that is able to predict germinative vigour in a multivariate data evaluation model,” J. Inst. Brewing, 110(1), 18–33 (2004).

[89] J. P. Wold, B. J. Marquardt, B. K. Dable, D. Robb, and B. Hatlen, “Rapid quantification of carotenoids and fat in Atlantic salmon (salmo salar L.) by Raman spectroscopy and chemometrics,” Appl. Spectrosc., 58(4), 395-403 (2004).

[90] Y. Peng, M. Knadel, R. Gislum, K. Schelde, A. Thomsen, and M. H. Greve, “Quantification of SOC and clay content using visible near-infrared reflectance-mid-infrared reflectance spectroscopy with Jack-knifing partial least squares regression,” Soil Sci., 179(7), 325-332 (2014).

[91] H. Almuallim and T. G. Dietterich, “Learning with many irrelevant features,” Proceedings of the Ninth National Conference on Artificial Intelligence, Anaheim, California, 547-552 (1991).

[92] K. Kira and L. A. Rendell, “A practical approach to feature selection,” Proceedings of the Ninth International Workshop on Machine Learning, Aberdeen, UK, 249–256 (1992).

[93] R. Kohavi and G. H. John, “Wrapper for feature subset selection,” Artif. Intell., 97(1-2), 273-324 (1997).

[94] M. A. Hall, “Correlation-based feature selection for machine learning,” Diss., Waikato University, New Zealand (1999).

[95] F. Liu, Y. He, and L. Wang, “Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis,” Anal. Chim. Acta, 615(1), 10–17 (2008).

[96] S. G. Nancy and S. A. Balamurugan, “A comparative study of feature selection methods for cancer classification using gene expression dataset,” Journal of Computer Applications (JCA), 6(3), 78-84 (2013).

[97] K. Hasegawa, Y. Miyashita, and K. Funatsu, “GA strategy for variable selection in QSAR studies: Ga-based PLS analysis of calcium channel antagonists,” J. Chem. Inf. Comput. Sci., 37(2), 306-310 (1997).

[98] R. Leardi and A. L. Gonzalez, “Genetic algorithms applied to feature selection in PLS regression: How and when to use them,” Chemometr. Intell. Lab., 41(2), 195–207 (1998).

[99] W. Cai, Y. Li, and X. Shao, “A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra,” Chemometr. Intell. Lab., 90(2), 188–194 (2008).

[100] B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, and M. Sznajder, “Detection of discoloration in diesel fuel based on gas chromatographic fingerprints,” Anal. Bioanal. Chem., 407(4), 1159–1170 (2015).

[101] A. Telaar, G. Nurnberg, and D. Repsilber, “Finding biomarker signatures in pooled sample designs: A simulation framework for methodological comparisons,” Adv. Bioinform., Article ID 318573 (2010).

[102] J. A. F. Pierna, O. Abbas, V. Baeten, and P. Dardenne, “A backward variable selection method for PLS regression (BVSPLS),” Anal. Chim. Acta, 642(1-2), 89-93 (2009).

[103] T. Baglin, T. W. Barrowcliffe, A. Cohen, and M. Greaves, “Guidelines on the use and monitoring of heparin,” Br. J. Haematol., 133(1), 19–34 (2006).

[104] J. D. Olson, C. F. Arkin, J. T. Brandt, M. T. Cunningham, A. Giles, J. A. Koepke, and D. L. Witte, “Laboratory monitoring of unfractionated heparin therapy,” Arch. Pathol. Lab. Med., 122(9), 782–798 (1998).

[105] “Point-of-care monitoring of anticoagulation therapy; Approved guideline,” CLSI document POCT14-A, Wayne, PA: Clinical and Laboratory Standard Institute (2004).

[106] P. D. Raymond, M. J. Ray, S. N. Callen, and N. A. Marsh, “Heparin monitoring during cardiac surgery. Part 2: Calculating the overestimation of heparin by the activated clotting time,” Perfusion, 18(5), 277–281 (2003).

[107] E. K. Heres, K. Speight, D. Benckart, J. Marquez, and G. P. Gravlee, “The clinical onset of heparin is rapid,” Anesth. Analg., 92(6), 1391–1395 (2001).

[108] S. Kitchen, I. Jennings, T. A. Woods, and F. E. Preston, “Wide variability in the sensitivity of APTT reagents for monitoring of heparin dosage,” J. Clin. Pathol., 49(1), 10–14 (1996).

[109] S. A. Spinler, A. K. Wittkowsky, E. A. Nutescu, and M. A. Smythe, “Point of care anticoagulation monitoring. Part 2: Unfractionated heparin and low molecular weight heparin,” Ann. Pharmacother., 39(7-8), 1275–1285 (2005).

[110] Y. Nosé, “Hemodialysis patients’ deaths in the USA by contaminant suspected heparin originating from China,” Artif. Organs, 32(6), 425–426 (2008).

[111] D. Perry and T. Todd, “Activated partial thromboplastin time [APTT],” http://www.practical-haemostasis.com/Screening%20Tests/aptt.html

[112] J. A. Young, C. T. Kisker, and D. B. Doty, “Adequate anticoagulation during cardiopulmonary bypass determined by activated clotting time and the appearance of fibrin monomer,” Annals Thorac. Surg., 26(3), 231–240 (1978).

[113] P. D. Raymond, M. J. Ray, S. N. Callen, and N. A. Marsh, “Heparin monitoring during cardiac surgery. Part 1: Validation of whole-blood heparin concentration and activated clotting time,” Perfusion, 18(5), 269–276 (2003).

[114] T. Ammar, C. F. Fisher, K. Sarier, and B. S. Coller, “The effects of thrombocytopenia on the activated coagulation time,” Anesth. Analg., 83(6), 1185–1188 (1996).

[115] R. T. Hall, P. G. Rhodes, E. A. Turner, and W. J. Braun, “Protamine sulfate titration for heparin activity in neonates with indwelling umbilical catheters,” J. Pediatr., 88(3), 467–472 (1976).

[116] M. A. Smythe, J. C. Mattson, and J. M. Koerber, “The heparin anti-Xa therapeutic range: are we there yet?” Chest, 121(1), 303–304 (2002).

[117] O. Shigeta, H. Kojima, Y. Hiramatsu, T. Jikuya, MDa, Y. Terada, N. Atsumi, Y. Sakakibara, T. Nagasawa, and T. Mitsui, “Low-dose protamine based on heparin-protamine titration method reduces platelet dysfunction after cardiopulmonary bypass,” J. Thorac. Cardiov. Sur., 118(2), 354–360 (1999).

[118] I. Weinberg, “Anti Xa,” http://www.angiologist.com/anti-xa

[119] M. J. Pelletier, “Quantitative analysis using Raman spectroscopy,” Appl. Spectrosc., 57(1), 20A–42A (2003).

[120] C. M. McGoverin, A. S. S. Clark, S. E. Holroyd, and K. C. Gordon, “Raman spectroscopic quantification of milk powder constituents,” Anal. Chim. Acta, 673(1), 26–32 (2010).

[121] R. M. El-Abassy, P. J. Eravuchira, P. Donfack, B. von der Kammer, and A. Materny, “Fast determination of milk fat content using Raman spectroscopy,” Vib. Spectrosc., 56(1), 3–8 (2011).

[122] C. A. Drumm and M. D. Morris, “Microscopic Raman line-imaging with principal component analysis,” Appl. Spectrosc., 49(9), 1331–1337 (1995).

[123] P. Matousek, “Subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy,” Appl. Spectrosc., 59(4), 393–400 (2005).

[124] G. V. Nogueira, L. Silveira, A. A. Martin, R. A. Zângaro, M. T. Pacheco, M. C. Chavantes, and C. A. Pasqualucci, “Raman spectroscopy study of atherosclerosis in human carotid artery,” J. Biomed. Opt., 10(3), 031117 (2005).

[125] H. Szelke, J. Harenberg, and R. Krämer, “Detection and neutralization of heparin by fluorescent ruthenium compound,” Thromb. Haemostasis, 102(5), 859–864 (2009).

[126] K. Gaus and E. Hall, “Evaluation of surface plasmon resonance (SPR) for heparin assay,” J. Colloid Interface Sci., 194(2), 364–372 (1997).

[127] N. Milovic, J. R. Behr, M. Godin, C. S. Hou, K. R. Payer, A. Chandrasekaran, P. R. Russo, R. Sasisekharan, and S. R. Manalis, “Monitoring of heparin and its low-molecular-weight analogs by silicon field effect,” Proc. National Acad. Sci., 103(36), 13374–13379 (2006).

[128] D. H. Atha, A. K. Gaigalas, and V. Reipa, “Structural analysis of heparin by Raman spectroscopy,” J. Pharm. Sci., 85(1), 52–56 (1996).

[129] D. L. Martin and K. Rimvall, “Regulation of gamma-aminobutyric acid synthesis in the brain,” J. Neurochem., 60(2), 395-407 (1993).

[130] S. R. Platt, “The role of glutamate in central nervous system health and disease - A review,” Vet. J., 173(2), 278-286 (2007).

[131] M. Chebib and G. A. R. Johnston, “The 'ABC' of GABA receptors: A brief review,” Clin. Exp. Pharmacol. Physiol., 26(11), 937-940 (1999).

[132] C. Advokat and A. I. Pellegrin, “Excitatory amino acids and memory: Evidence from research on Alzheimer’s disease and behavioral pharmacology,” Neurosci. Biobehav. Rev., 16(1), 13–24 (1992).

[133] F. Blandini, R. H. P. Porter, and J. T. Greenamyre, “Glutamate and Parkinson’s disease,” Mol. Neurobiol., 12(1), 73–94 (1996).

[134] B. S. Meldrum, “The role of glutamate in epilepsy and other CNS disorders,” Neurology, 44(11 Suppl 8), S14–S23 (1994).

[135] M. Podell and M. Hadjiconstantinou, “Low concentrations of cerebrospinal fluid GABA correlate to a reduced response to phenobarbital therapy in primary canine epilepsy,” J. Vet. Intern. Med., 13(2), 89–94 (1999).

[136] C. S. Jung, B. Lange, M. Zimmermann, and V. Seifert, “CSF and serum biomarkers focusing on cerebral vasospasm and ischemia after subarachnoid hemorrhage,” Stroke Res. Treat., 2013, Article ID 560305 (2013).

[137] Y. Qu, L. Arckens, E. Vandenbussche, S. Geeraerts, and F. Vandesandea, “Simultaneous determination of total and extracellular concentrations of the amino acid neurotransmitters in cat visual cortex by microbore liquid chromatography and electrochemical detection,” J. Chromatogr. A, 798(1-2), 19–26 (1998).

[138] R. T. Kenedy, C. J. Watson, W. E. Haskins, D. H. Powell, and R. E. Strecker, “In vivo neurochemical monitoring by microdialysis and capillary separations,” Curr. Opin. Chem. Biol., 6(5), 659–665 (2002).

[139] K. Buck, P. Voehringer, and B. ferger, “Rapid analysis of GABA and glutamate in microdialysis samples using high performance liquid chromatography and tandem mass spectrometry,” J. Neurosci. Meth., 182(1), 78–84 (2009).

[140] J. Kehr, “Determination of gamma-aminobutyric acid in microdialysis samples by microbore column liquid chromatography and fluorescence detection,” J. Chromatogr. B Biomed. Sci. Appl., 708(1-2), 49–54 (1998).

[141] J. Kehr, “Determination of glutamate and aspartate in microdialysis samples by reversed-phase column liquid chromatography with fluorescence and electrochemical detection,” J. Chromatogr. B Biomed. Sci. Appl., 708(1-2), 27–38 (1998).

[142] V. Sauvinet, S. Parrot, N. Benturquia, E. Bravo-Moratón, B. Renaud, and L. Denoroy, “In vivo simultaneous monitoring of aminobutyric acid, glutamate, and L - aspartate using brain microdialysis and capillary electrophoresis with laser-induced fluorescence detection: Analytical developments and in vitro/in vivo validations,” Electrophoresis, 24(18), 3187–3196 (2003).

[143] Y. Song, M. Shenwua, D. M. Dhosscheb, and Y. M. Liua, “A capillary liquid chromatographic-tandem mass spectrometric method for the quantification of GABA in human plasma and cerebrospinal fluid,” J. Chromatogr. B Analyt. Technol. Biomed. Life, 814(2), 295–302 (2005).

[144] C. Sanol, F. Artigas, J. M. Tusell, and E. Gelpi, “High-performance liquid chromatography-fluorescence detection method for endogenous GABA validated by mass spectrometric and gas chromatographic techniques,” Anal. Chem., 60(7), 649-651 (1988).

[145] G. S. Duchateau, W. M. Albers, and H. H. van Rooij, “Rapid and simple determination of alprenolol in serum,” J. Chromatogr., 383(1), 212-217 (1986).

[146] L. W. Cao, X. F. Tan, C. Li, C. Wu, Z. D. Zhang, T. Deng, and J. X. Meng, “Capillary electrophoresis-laser induced fluorescence detection of GABA and its analogs in human serum with solid-phase extraction and fluorescein-based probes,” Anal. Methods, 5(21), 6000–6008, (2013).

[147] C. A. Vyas, “Rapid detection of biogenic amines using capillary electrophoresis and gradient elution isotachophoresis,” Diss., Temple University, USA (2011).

[148] K. Kneipp, Y. Wang, R. R. Dasari, and M. S. Feld, “Near-infrared surface-enhanced Raman scattering (NIR-SERS) of neurotransmitters in colloidal silver solutions,” Spectrochim. Acta A Mol. Biomol. Spectrosc., 51(3), 481-487 (1995).

[149] R. J. Dijkstra, W. J. J. M. Scheenen, N. Dam, E. W. Roubos, and J. J. ter Meulen, “Monitoring neurotransmitter release using surface-enhanced Raman spectroscopy,” J. of Neurosci. Meth., 159(1), 43–50 (2007).

[150] N. S. Lee, Y. Z. Hsieh, R. F. Paisley, and M. D. Morris, “Surface-enhanced Raman spectroscopy of the catecholoamine neurotransmitters and related compounds,” Anal. Chem., 60(5), 442-446 (1988).

[151] N. Peica, C. Lehene, N. Leopold, S. Schlucker, and W. Kiefer, “Monosodium glutamate in its anhydrous and monohydrate form-Differentiation by Raman spectroscopies and density functional calculations,” Spectrochim. Acta A Mol. Biomol. Spectrosc., 66(3), 604–615 (2007).

[152] J. L. Castro, S. Sanchez-cortes, J. V. G. Ramos, J. C. Otero, and J. I. Marcos, “SERS of γ-aminobutyric acid on silver colloid surfaces,” Biospectroscopy, 3(6), 449–455 (1997).

[153] P. D. O’Neal, M. Motamedi, W. C. Lin, J. Chen, and G. L. Cote, “Feasibility study using surface-enhanced Raman spectroscopy for the quantitative detection of excitatory amino acids,” J. Biomed. Opt., 8(1), 33–39 (2003).

[154] V. S. Tiwari, A. Khetani, A. Momenpour, B. Smith, and H. Anis, “Detection of amino acid neurotransmitters by surface enhanced,” Proc. SPIE, 8233, 82330Q (2012).

[155] N. Leopold and B. Lendl, “A new method for fast preparation of highly surface-enhanced Raman scattering (SERS) active silver colloids at room temperature by reduction of silver nitrate with hydroxylamine hydrochloride,” J. Phys. Chem. B, 107(24), 5723-5727 (2003).

[156] I. Shimizu, H. Okabayashi, K. Taga, and C. J. O’Connor, “Raman scattering study of polyaminopropylsiloxane and its compounds for characterization of 3-aminopropylsilane-modified silica gel. Utility of the CH2 rock and skeletal stretch modes,” Colloid Polym. Sci., 275(6), 555-560 (1997).

[157] D. M. Suresh, D. Sajan, K. P. Laladas, I. H. Joe, and V. S. Jayakumar, “Vibrational spectra of γ-aminobutyric acid,” AIP Conf. Proc., India, 1075(1), 95-97 (2008).

[158] J. S. Suh and M. Moskovits, “Surface-enhanced Raman spectroscopy of amino acids and nucleotide bases adsorbed on silver,” J. Am. Chem. Soc., 108916), 4711-4718 (1986).

[159] Y. Xu and C. Lu, “Raman spectroscopic study on structure of human immunodeficiency virus (HIV) and hypericin-induced photosensitive damage of HIV,” Sci. China C Life Sci., 48(2), 117-132 (2005).

[160] A. Khetani, J. Riordon, V. Tiwari, A. Momenpour, M. Godin, and H. Anis, “Hollow core photonic crystal fiber as a reusable Raman biosensor,” Opt. Express, 21(10), 12340–12350 (2013).

[161] D. K. Graham, D. B. Salzberg, J. Kurtzberg, S. Sather, G. K. Matsushima, A. K. Keating, X. Liang, M. A. Lovell, S. A. Williams, T. L. Dawson, M. J. Schell, A. A. Anwar, H. R. Snodgrass, and H.

S. Earp, “Ectopic expression of the proto-oncogene Mer in pediatric T-cell acute lymphoblastic leukemia,” Clin. Cancer Res., 12(9), 2662–2669 (2006).

[162] K. Zhang, T. Tan, J. J. Fu, T. Zheng, and J. J. Zhu, “A novel aptamer-based competition strategy for ultrasensitive electrochemical detection of leukemia cells,” Analyst, 138(21), 6323–6330 (2013).

[163] C. M. Brown, S. R. Larsen, H. J. Iland, D. E. Joshua, and J. Gibson, “Leukaemias into the 21st century: Part 1: The acute leukaemias,” Intern. Med. J., 42(11), 1179–1186 (2012).

[164] Z. Darzynkiewicz, and H. Zhao, “Cell cycle analysis by flow cytometry,” eLS, (2014).

[165]. H. G. Goh, M. Lin, T. Fukushima, G. Saglio, D. Kim, S. Y. Choi, S. H. Kim, J. Lee, Y. S. Lee, S. M. Oh, and D. W. Kim, “Sensitive quantitation of minimal residual disease in chronic myeloid leukemia using nanofluidic digital polymerase chain reaction assay,” Leukemia Lymphoma, 52(5), 896–904 (2011).

[166] R. J. Olsen, C. C. Chang, J. L. Herrick, Y. Zu, and A. Ehsan, “Acute leukemia immunohistochemistry: A systematic diagnostic approach,” Arch. Pathol. Lab. Med., 132(3), 462–475 (2008).

[167] C. Righeschi, T. Eichhorn, A. Karioti, A. R. Bilia, and T. Efferth, “Microarray-based mRNA expression profiling of leukemia cells treated with the flavonoid, casticin,” Cancer Genom. Proteom., 9(3), 143–151 (2012).

[168] R. Li, Y. Tan, X. Chen, F. Ren, Y. Zhang, Z. Xu, and H. Wang, “Fluorescence probe analysis of leukemia cells by modified graphene oxide,” New Carbon Mater., 29(6), 438-443 (2014).

[169] H. Yan, C. Gu, C. Yang, J. Liu, G. Jin, J. Zhang, L. Hou, and Y. Yao, “Hollow core photonic crystal fiber surface-enhanced Raman probe,” Appl. Phys. Lett., 89(20), 204101 (2006).

[170] V. S. Tiwari, A. Khetani, M. Naji, and H. Anis, “Study of surface-enhanced Raman scattering (SERS) within hollow core photonic crystal fiber,” IEEE Sensors Conference, 367–370 (2009).

[171]. V. S. Tiwari, A. Khetani, A. Momenpour, and H. Anis, “Optimum size and volume of nano particles within hollow core photonic crystal fiber,” IEEE J. Quantum Electron., 20(3), Article ID 7300608 (2014).

[172] T. Shimizu and Y. Pommier, “Camptothecin-induced apoptosis in p53-null human leukemia HL60 cells and their isolated nuclei: effects of the protease inhibitors Z-VAD-fmk and

dichloroisocoumarin suggest an involvement of both caspases and serine proteases," Leukemia 11(8), 1238-1244 (1997).

[173] M. Gupta, A. Fujimori, and Y. Pommier, "Eukaryotic DNA topoisomerases I," Biochim. Biophys. Acta (BBA)/Gene Structure and Expression, 1262(1), 1-14 (1995).

[174] Y. Pommier, "Eukaryotic DNA topoisomerase I: Genome gatekeeper and its intruders, camptothecins," Semin. Oncol., 23(1 Suppl 3), 3–10 (1996).

[175]. Y. Oshima, H. Shinzawa, T. Takenaka, C. Furihata, and H. Sato, "Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy," J. Biomed. Opt., 15(1), 017009 (2010).

[176] "Examples using the pls procedure,"
http://support.sas.com/rnd/app/stat/papers/plsex.pdf

[177] "The Unscrambler tutorials,"
http://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Tutorials.pdf

[178] A. Khetani, A. Momenpour, E. I. Alarcon, and H. Anis, "Hollow core photonic crystal fiber for monitoring leukemia cells using surface enhanced Raman scattering (SERS)," Biomed. Opt. Express, 6(11), 4599-4609 (2015).

[179] A. Momenpour, V. S. Tiwari, M. M. Tripathi, and H. Anis, "Raman spectroscopy for clinical-level detection of heparin in serum by partial least-squares analysis," J. Biomed. Opt., 18(2), 27010 (2013).

[180] "The PLS-Genetic algorithm toolbox for MATLAB(TM),"
http://www.models.life.ku.dk/GAPLS

[181] S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, and H. Zeng, "Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis," Biosens. Bioelectron., 25(11), 2414-2419 (2010).

[182] H. W. Han, X. L. Yan, R. X. Dong, G. Ban, and K. Li, "Analysis of serum from type II diabetes mellitus and diabetic complication using surface-enhanced Raman spectra (SERS)," Appl. Phys. B, 94(4), 667-672 (2009).

[183] D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, and R. Chen, “Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis,” Opt. Express, 19(14), 13565-13577 (2011).

[184] Z. Huang, A. McWilliams, H. Lui, D.I. McLean, S. Lam, and H. Zeng, “Near-infrared Raman spectroscopy for optical diagnosis of lung cancer,” Int. J. Cancer, 107(6), 1047-1052 (2003).

[185] G. Basar, U. Parlatan, S. Seninak, T. Gunel, A. Benian, and I. Kalelioglu, “Investigation of preeclampsia using Raman spectroscopy,” Spectrosc. Int. J., 27(4), 239-252 (2012).

[186] J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez, and C. Medina-Gutiérrez, “Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients,” Lasers Med. Sci., 22(4), 229-236 (2007).

[187] B. H. Stuart and D. J. Ando, “Biological applications of infrared spectroscopy,” Chichester: John Wiley & Sons (1997).

[188] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Raman spectroscopy of biological tissues,” Appl. Spectrosc. Rev., 42(5), 493-541 (2007).

[189] R. Bansil, I. V. Yannas, and H. E. Stanley, “Raman spectroscopy: A structural probe of glycosaminoglycans,” Biochim. Biophys. Acta, 541(4), 535-542 (1978).

[190] K. Annapoorani, K. Maheshvaran, S. ArunKumar, N. S. Murthy, T. Soukka, and K. Marimuthu, “Structural and spectroscopic behavior of Er 3+ :Yb 3+ co-doped lithium telluroborate glasses,” Physica B, 457, 66-77 (2015).

[191] Q. Wang, J. Y. Kim, K. Xue, J. Y. Liu, A. Leader, and B. K. Tsang, “Chemerin, a novel regulator of follicular steroidogenesis and its potential involvement in polycystic ovarian syndrome,” Endocrinology, 153(11), 5600-5611 (2012).

[192] The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, “Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome,” Fertil. Steril., 81(1), 19-25 (2004).

[193] N. R. FARID and E. D. Kandarakis, “Diagnosis and management of polycystic ovary syndrome,” New York: Springer (2009).

[194] O. De Henau, G. N. Degroot, V. Imbault, V. Robert, C. De Porter, S. Mcheik, C. Galés, M. Parmentier, and J. Y. Springael, “Signaling properties of chemerin receptors CMKLR1, GPR1 and CCRL2,” PLoS One, 11(10), e0164179 (2016).

[195] K. Bozaoglu, K. Bolton, J. McMillan, P. Zimmet, J. Jowett, G. Collier, K. Walder, and D. Segal, “Chemerin is a novel adipokine associated with obesity and metabolic syndrome,” Endocrinology, 148(10), 4687–4694 (2007).

[196] H. Y. Shin, D. C. Lee, S. H. Chu, J. Y. Jeon, M. K. Lee, J. A. Im, and J. W. Lee, “Chemerin levels are positively correlated with abdominal visceral fat accumulation,” Clin. Endocrinol., 77(1), 47–50 (2012).

[197] A. Dunaif, “Insulin action in the polycystic ovary syndrome,” Endocrinol Metab. Clin. North Am., 28(2), 341-359 (1999).

[198] S. Sam, “Obesity and polycystic ovary syndrome,” Obes. Manag., 3(2), 69-73 (2007).

[199] D. H. Kort, A. Kostolias, C. Sullivan, and R. A. Lobo, “Chemerin as a marker of body fat and insulin resistance in women with polycystic ovary syndrome,” Gynecol. Endocrinol., 31(2), 152-155 (2015).

[200] B. K. Tan, J. Chen, S. Farhatullah, R. Adya, J. Kaur, D. Heutling, K. C. Lewandowski, J. P. O'Hare, H. Lehnert, and H. S. Randeva, “Insulin and metformin regulate circulating and adipose tissue chemerin,” Diabetes, 58(9), 1971-1977 (2009).

[201] S. S. Chang, D. Eisenberg, L. Zhao, C. Adams, R. Leib, J. Morser, and L. Leung, “Chemerin activation in human obesity,” Obesity, 24(7), 1522-1529 (2016).

[202] G. Zhu, X. Zhu, Q. Fan, and X. Wan, “Raman spectra of amino acids and their aqueous solutions,” Spectrochim. Acta A Mol. Biomol. Spectrosc., 78(3), 1187-1195 (2011).

[203] S. Stewart and P. M. Fredericks, “Surface-enhanced Raman spectroscopy of amino acids adsorbed on an electrochemically prepared silver surface,” Spectrochim. Acta A Mol. Biomol. Spectrosc., 55 (7-8), 1641-1660 (1999).

[204] L. Simons, A. Pohjavirta, O. Nevanlinna, P. Suomela, “Laser Raman spectroscopy of amino acids, oligopeptides, polypeptides and enzymes,” Helsinki: Societas Scientiarum Fennica (1972).

[205] S. A. Asher, M. Ludwig, and C. R. Johnson, “UV resonance Raman excitation profiles of the aromatic amino acids,” J. Am. Chem. Soc., 108(12), 3186-3197 (1986).

[206] P. F. Facanha Filho, P. T. C. Freire, K. C. V. Lima, J. Mendes Filho, and F. E. A. Melo, “High temperature Raman spectra of l-leucine crystals,” Braz. J. Phys., 38(1), 131-137 (2008).

[207] J. Jehlicka, P. Vítek, and H. G. M. Edwards, “Raman spectra of organic acids obtained using a portable instrument at −5 ◦C in a mountain area at 2000 m above sea level,” J. Raman Spectrosc., 41, 440–444 (2010).

[208] K. A. Esmonde-White, G. S. Mandair, F. W. L. Esmonde-White, F. Raaii, B. J. Roessler, and M. D. Morris, “Osteoarthritis screening using Raman spectroscopy of dried human synovial fluid drops,” Proc. SPIE, 7166, 71660J-1 (2009).

[209] Y. Li, J. Pan, G. Chen, C. Li, S. Lin, Y. Shao, S. Feng, Z. Huang, S. Xie, H. Zeng, and R. Chen, “Micro-Raman spectroscopy study of cancerous and normal nasopharyngeal tissues,” J. Biomed. Opt., 18(2), 27003 (2013).

[210] S. Olsztynska, N. Dupuy, L. Vrielynck, and M. Komorowska, “Water evaporation analysis of L-phenylalanine from initial aqueous solutions to powder state by vibrational spectroscopy,” Appl. Spectrosc., 60(9), 1040-1053 (2006).

[211] J. R. Govani, W. G. Durrer, M. Manciu, C. Botez, and F. S. Manciu, “Spectroscopic study of L-arginine interactions with potassium dihydrogen phosphate crystals,” J. Mater. Res., 24(7), 2316-2320 (2009).

[212] J. Guicheteau, L. Argue, A. Hyre, M. Jacobson, and S. D. Christesen, “Raman and surface-enhanced Raman spectroscopy of amino acids and nucleotide bases for target bacterial vibrational mode identification,” Proc. SPIE, 6218, 62180O (2006).

[213] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, “Reference database of Raman spectra of biological molecules,” J. Raman Spectrosc., 38(9), 1133-1147 (2007).

[214] J. T. L. Navarrete, V. Hernández, and F. J. Ramírez, “Vibrational study of aspartic acid and glutamic acid dipeptides,” J. Mol. Struct., 348, 249-252 (1995).

[215] K. A. Esmonde-White, S. V. Le Clair, B. J. Roessler, and M. D. Morris, “Effect of conformation and drop properties on surface-enhanced Raman spectroscopy of dried biopolymer drops,” Appl. Spectrosc., 62(5), 503-511 (2008).

[216] K. A. Esmonde-White, G. S. Mandair, F. Raaii, J. A. Jacobson, B. S. Miller, A. G. Urquhart, B. J. Roessler, and M. D. Morris, “Raman spectroscopy of synovial fluid as a tool for diagnosing osteoarthritis,” J. Biomed. Opt., 14(3), 034013 (2009).

[217] K. A. Esmonde-White, J. Sottnik, M. Morris, and E. Keller, “Raman spectroscopy of bone metastasis,” Proc. SPIE, 8207, 82076P-1 (2012).