

**A MULTIMODAL SPECTROMETER FOR
RAMAN SCATTERING AND NEAR-
INFRARED ABSORPTION
MEASUREMENT**

by

Wenbo Wang

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Biosystems Engineering

University of Manitoba

Winnipeg

Copyright © 2012 by Wenbo Wang

ABSTRACT

In food and pharmaceutical industries, the compositional complexity of raw materials and products promotes the use of multiple spectroscopic techniques. Availability of complementary methods, e.g., infrared absorption and Raman, is highly desirable. A multimodal Raman/NIR spectroscopic device was proposed. The device offered fast, non-destructive measurement, and remote sensing capability. A prototype was built. Spectral range covered 300-2600 cm^{-1} for Raman channel. Wavelength accuracy was better than 0.57 nm. Spectral resolution was better than 14.39 cm^{-1} (1 nm) for Raman (NIR) channel. Spectral response linearity was better than 0.998. Instrumental stability was better than 0.72% (0.39%) for Raman (NIR) channel.

The device was applied to quantify ethanol aqueous solutions (2%-10% (v/v)). Quantitation results gave a root mean squared error of prediction (RMSEP) of 0.45% (v/v). Calibration using NIR absorbance spectra produced a RMSEP value of 0.49% (v/v). Raman technique offered the most succinct calibration model using peak height at 882.29 cm^{-1} . Quantitation of ovalbumin (8%-16% (w/v)) in aqueous solutions and denatured states was also attempted. Stepwise multiple regression was performed using Raman peaks at 1243.35, 1448.68, 1662.66, and 1667.78 cm^{-1} , which produced RMSEPs of 1.05 % (w/v) for ovalbumin solutions and 0.74 % (w/v) for ovalbumin gels. Conformational change of ovalbumin upon thermal denaturation was studied using the height and area ratios between peaks at 1243.35 and 1662.66 cm^{-1} . Thermal denaturation of ovalbumin occurred at 70 °C and extensive formation of anti-parallel β -sheet structure was found at 90 °C.

The unique capability of this multimodal device was proven by first performing data fusion of Raman and NIR spectra for quantitation. Using the same ethanol spectra set gave a more complicated model and a higher RMSEP of 0.73% (v/v). However, important variables in fused spectra corresponded well with the spectral signatures of ethanol and water. Two-dimensional correlation spectra were calculated using concentration as the external perturbation. The autopower spectra revealed features pertinent to those of the analytes. Heterospectral correlation analysis was carried out on Raman/NIR spectra of ethanol aqueous solution. Raman peaks at 882.29 and 1454.00 cm^{-1} agreed with the NIR absorption feature around 906.18 nm.

ACKNOWLEDGEMENTS

Had it not been for the help from many individuals, this thesis work would not have been possible. First of all, I would like to express my sincere thanks to my advisor, Dr. Jitendra Paliwal, for his generous support, unfailing guidance, and inspiring optimism that helped me grow both professionally and academically throughout the years.

I am thankful to Dr. Arkady Major and Mr. Mark Hewko for serving on my advisory committee. Dr. Major's help has been instrumental to the completion of this thesis. I would like to thank him for allowing me access to his state-of-the-art biophotonics lab and offering me expert advices on optical designs. I also would like to thank Mark for enabling me a vision on the practical sides of spectroscopy and supervising my internship at the National Research Council (NRC). I wish to express my sincere gratitude towards my NRC internship supervisor, Dr. Lin-P'ing Choo-Smith, for introducing me into the wonderful world of Raman spectroscopy and helping me through the final stages of the research.

I would like to thank the Institute of Biodiagnostics (IBD-NRC) and Channel Systems, Inc. for offering me industrial internships respectively, which substantially enriched my technical skills and work experience. The Biosystems Engineering department at the University of Manitoba has been very supportive. I will always remember you all for your genuine smiles.

I would also like to acknowledge the financial assistance provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Government of Manitoba, and the University of Manitoba.

Finally, I would like to thank my parents and my family. You endured so much that I could not possibly bear.

TABLE OF CONCENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONCENTS	v
LIST OF FIGURES	viii
LIST OF TABLES.....	xi
1. INTRODUCTION	1
1.1 Raman Scattering, Infrared Absorption, and Complementarity.....	1
1.2 Theoretical Background	4
1.3 Practical Implications.....	6
1.4 Choice of Instrumentation and Justification.....	7
1.5 Feasibility of Proposed Instrumentation	12
1.6 Research Objectives	14
2. LITERATURE REVIEW	16
2.1 Basis for Vibrational Spectroscopy.....	16
2.1.1 Infrared and near-infrared spectroscopy.....	16
2.1.2 Molecular rotation and vibration	17
2.1.3 Molecular vibration	18
2.1.4 Anharmonicity, overtones, combination, and difference	20
2.1.5 Vibrational potential energy, quantum effect.....	22
2.1.6 Rotational transition and vibrational-rotational spectrum.....	24
2.1.7 Raman Spectroscopy	27
2.1.8 Polarizability.....	29
2.1.9 Raman cross-section and depolarization ratio	32
2.2 Recent Progress Relevant to Proposed Research	34
2.2.1 Comparative study on IR, NIR, and Raman spectroscopy.....	34
2.2.2 Multimodal spectroscopy	38
2.2.3 Protein analysis using NIR and Raman spectroscopy	45
2.2.4 Analysis of ethanol in aqueous solutions	51

2.2.5 Two-dimensional correlation spectroscopy.....	53
2.3 Summary	55
3. MATERIALS and METHODS.....	56
3.1 The Instrument	56
3.1.1 EMCCD camera and control software.....	58
3.1.2 The spectrograph	61
3.1.3 Fiber optics	62
3.1.4 Raman sampling probe	63
3.1.5 The light sources.....	65
3.1.6 Testing equipments and miscellaneous parts.....	65
3.1.7 System calibration and validation.....	66
3.2 Samples for Analysis.....	68
3.2.1 Chemicals	68
3.2.2 Albumin sample preparation	69
3.2.3 Ethanol sample preparation	70
3.3 Experimental design.....	70
3.4 Data Handling	72
3.4.1 Software.....	72
3.4.2 Spectral data extraction	73
3.4.3 Algorithms	73
4. RESULTS AND DISCUSSION	76
4.1 System Calibration	76
4.1.1 Camera and spectrograph alignment	76
4.1.2 Wavenumber (wavelength) calibration and validation.....	79
4.1.3 Spectral range estimation.....	89
4.1.4 Spectral resolution	90
4.1.5 Photometric linearity	92
4.1.6 Stability.....	93
4.2 Quantifying Ethanol in Aqueous Solutions.....	96
4.2.1 Raman spectroscopy	97
4.2.2 Near-infrared spectroscopy.....	111

4.2.3 Summary.....	116
4.3 Quantitative Analysis of Ovalubmin.....	118
4.3.1 Quantification using Raman spectroscopy	118
4.3.2 Quantification using Near-infrared spectroscopy.....	126
4.3.3 Summary.....	129
4.4 Ovalbumin Secondary Structure Changes upon Thermal Denaturation Using Raman Spectroscopy	130
4.5 Data Fusion	133
4.6 Two-dimensional Correlation Spectroscopy	138
4.6.1 Analysis of ovalbumin and ethanol solution using 2D correlation spectra	138
4.6.2 Raman/NIR 2D heterospectral correlation spectral analysis	142
4.6.3 Summary.....	144
5. CONCLUSIONS.....	145
6. RECOMMENDATIONS	149
REFERENCES	151
APPENDIX A. Data Acquisition Settings for Different Samples	166

LIST OF FIGURES

Number	Title	Page
Fig. 2.1	The absorption and scattering processes during vibration energy transitions	29
Fig. 3.1	Schematic layout of the multimodal spectroscopic system	57
Fig. 3.2	Assignment of the EMCCD sensor array	60
Fig. 3.3	Schematic layout of the three ends of the bifurcated fiber optic bundle.	62
Fig. 3.4	Schematic layout of the Raman sampling probe	64
Fig. 3.5	Experiment design diagram	71
Fig. 4.1	Spectral image of fluorescent light (0.1s integration time, 10 accumulation, and EM gain 100)	77
Fig. 4.2	Spectra at channel 5, 150, and 280	78
Fig. 4.3	Spectral images for wavenumber calibration	81
Fig. 4.4	Raman spectra of benzonitrile	82
Fig. 4.5	Raman spectra of 4-Acetamidophenol	83
Fig. 4.6	Variance versus polynomial order for different models	86
Fig. 4.7	Spectrum of Argon lamp	88
Fig. 4.8	Gaussian fitted spectral peaks	91
Fig. 4.9	Average spectra of reflectance standards	92
Fig. 4.10	Linear regression between reflectance measurement and certified reflectance levels	93

Number	Title	Page
Fig. 4.11	Percent SD plot	95
Fig. 4.12	Raman spectra of anhydrous ethanol and 10% (v/v) solution	98
Fig. 4.13	Calculation of peak height	100
Fig. 4.14	Univariate calibration between Raman signal at 882.16 cm ⁻¹ and ethanol concentration	102
Fig. 4.15	Prediction results using univariate calibration developed with (a) Raman peak height; (b) integrated peak area	103
Fig. 4.16	Prediction results using stepwise calibration developed with (a) Raman peak height; (b) integrated peak area	107
Fig. 4.17	Prediction results using PLSR calibration developed with (a) Raman peak height; (b) integrated peak area	108
Fig. 4.18	Absorption spectra of distilled water and anhydrous ethanol	111
Fig. 4.19	Predicted residual sum of squares values versus factor number	113
Fig. 4.20	Beta coefficients	114
Fig. 4.21	Prediction results using NIR calibrations	115
Fig. 4.22	Raman spectra of solid ovalbumin, 16% (w/v) ovalbumin solution, and water in quartz cell	119

Number	Title	Page
Fig. 4.23	Albumin concentrations in solutions predicted using calibration developed with (a) Raman peak height; (b) integrated peak area	122
Fig. 4.24	Denatured albumin gel concentration predicted using calibration developed with (a) Raman peak height; (b) integrated peak area	124
Fig. 4.25	Absorbance spectra for 15% (w/v) albumin solution, water, and solid albumin powder	126
Fig. 4.26	Beta coefficients for ovalbumin solution and denatured gel	128
Fig. 4.27	Raman spectra of 16% (w/v) ovalbumin solution and ovalbumin gels formed after heating at 70 °C and 90 °C for 1 h	131
Fig. 4.28	Fused spectra of Raman and NIR absorbance data for ethanol aqueous solution	134
Fig. 4.29	Predicted residual sum of squares values versus factor number	135
Fig. 4.30	Beta coefficients for fused Raman and NIR spectra	136
Fig. 4.31	Prediction results for fused Raman and NIR data	137
Fig. 4.32	2D correlation spectra of ovalbumin	139
Fig. 4.33	2D correlation spectra of ethanol in aqueous solution	141
Fig. 4.34	2D Raman/NIR heterospectral correlation spectra generated from Raman and NIR spectra of ethanol aqueous solution	143

LIST OF TABLES

Number	Title	Page
Table 1.1	Comparison between different instrumentation options	11
Table 2.1	Spectral absorption and chemical structure correlations	49
Table 2.2	Common vibrational modes used in protein structure study expressed in wavenumbers (cm^{-1})	49
Table 2.3	Physiochemical properties of major egg white proteins	50
Table 3.1	Chemicals purchased from Sigma Aldrich and their purposes	68
Table 3.2	Degrees of freedom for the rational experiment design	72
Table 4.1	Wavenumbers used for camera spectral axis calibration	85
Table 4.2	Variance calculated using second order polynomial	87
Table 4.3	Wavelengths used for verification	89
Table 4.4	Spectral resolutions at measured wavelengths	91
Table 4.5	Photometric linearity of the spectrographic system	93
Table 4.6	Assignment of Raman peaks and shift due to hydrogen bonding	98
Table 4.7	VIF values for peak height variables	105
Table 4.8	VIF values for integrated peak area variables	105

Number	Title	Page
Table 4.9	Stepwise regression statistics for peak height variables	106
Table 4.10	Stepwise regression statistics for peak area variables	106
Table 4.11	Model statistics for different calibrations on Raman spectra of ethanol aqueous solutions	110
Table 4.12	Model statistics for NIR calibrations of ethanol aqueous solutions	116
Table 4.13	Assignment of Raman modes	119
Table 4.14	Spectral regions selected for linear baseline and integrated peak area calculation (cm^{-1})	120
Table 4.15	Stepwise Regression coefficient statistics for peak height variables of albumin solution	121
Table 4.16	Stepwise Regression coefficient statistics for peak area variables of albumin solution	121
Table 4.17	Stepwise Regression coefficient statistics for peak height variables of denatured albumin gel	123
Table 4.18	Stepwise Regression coefficient statistics for peak area variables of denatured albumin gel	123

Number	Title	Page
Table 4.19	Model statistics for stepwise calibrations with Raman spectra of ovalbumin aqueous solutions	125
Table 4.20	Model statistics for NIR calibrations of ovalbumin aqueous solutions	128
Table 4.21	Type 3 Analysis of Variance (peak height)	131
Table 4.22	Type 3 Analysis of Variance (peak area)	131
Table 4.23	Differences of Least Squares Means (peak height)	132
Table 4.24	Differences of Least Squares Means (peak area)	132
Table 4.25	Model statistics for fused NIR/Raman data	137

1. INTRODUCTION

1.1 Raman Scattering, Infrared Absorption, and Complementarity

Life on earth depends on electromagnetic radiation, which interacts with electrons and nuclei of molecules through processes of absorption, emission, and scattering (Li-Chan 1996). Observing and recording the aforementioned processes helps human beings understand the molecular vibrations and rotations of complex molecules. Molecular vibrations manifest themselves as fundamental, overtones, and combination tones in a spectrum. According to quantum physics, for different molecules, spectral features corresponding to the energy difference between ground state and first excited state of vibrational modes are unique. Thus, spectral features could be used to identify molecules and functional groups. With vibrational spectroscopy, vibrations of molecules could be identified through processes such as absorption and inelastic scattering. Absorption of incident light due to transition between ground mode and the first excited mode of molecular vibration occurs in the infrared (IR) region. An IR spectrum of condensed-phase matter often consists of prominent absorption bands known as the fingerprint bands. These unique bands are correlated to certain functional groups or molecules. Inelastic scattering, i.e., Raman scattering, involves exactly the same vibrational and rotational energy levels. However, due to different selection rules of chemical bonds, spectra resulted from Raman scattering and IR absorption are not duplicates of each other, but are complementary. For instance, the C=C stretching band is weak in IR spectrum but has strong Raman scattering. On the other hand, Raman scattering due to

the C=O stretching band is relatively weak and IR absorption due to C=O stretch is strong.

Often, the IR absorption spectra are taken using Fourier transform IR (FT-IR) technique, which is popular and highly specific in analytical labs. However, the FT-IR technique has drawbacks as a routine analysis tool for *on-line* and *in situ* applications. First, the sampling requires non-trivial preparation, e.g., compressing a KBr pellet, thinning films, etc. Second, scanning an IR spectrum must deal with strong interference from CO₂ and gaseous H₂O in the atmosphere. Near-infrared (NIR) absorption technique, which requires little sample preparation and has less intense absorption bands, is used instead. Absorption in the NIR region occurs due to combinational tones and overtones of the fundamental vibrational modes. For organic samples, the first and second overtones of O-H, N-H, and C-H bonds, or combination tones involving these bonds could be detected in the NIR region. Compared to the fundamental modes, combination and overtone modes are generally weak and unintelligible due to the highly overlapping and broad spectral bands. With the assistance of chemometric methods, NIR spectra are amenable to both quantitative and qualitative analysis. Overall, NIR is chosen over FT-IR mainly for its sampling ease. For instance, NIR spectra of samples could be obtained directly through glass container and/or transmitted through fiber optics, which allows for convenient, fast, and non-destructive evaluation.

For the agricultural and food industry, NIR spectroscopy has been a widely accepted as a fast and convenient tool for analysis of food compounds. The rapidly growing interest in the use of NIR was mainly due to its sampling advantages and the development of chemometrics. Though not providing as much information as IR

technique, NIR spectroscopy has been used to assure identity of chemicals in products of industrial processes. However, it should be noted that there exist limitations to the NIR applications. As was previously mentioned, NIR bands only reflect overtones of C-H, O-H, and N-H stretching bonds or combinations of their vibration modes. Much spectral information is lost in NIR spectra compared to IR data and it is speculated that loss of structure specific information makes NIR spectroscopy less successful in identifying molecules, especially molecules with similar structures. On the other hand, it was also argued that NIR spectroscopy possessed slight advantage in analyzing structurally similar molecules because the effect the small structural difference of molecules has on first overtone is significant (Griffiths and Haseth 2007).

According to the classical polarizability treatment of Raman effect, a molecular vibration must be accompanied by a change in its polarizability to be Raman active. On the other hand, the absorption of IR radiation happens only when molecular vibration causes a change in the dipole moment of the molecule. Therefore, Raman spectrum reveals information on molecular structures other than that interpreted through absorption process in the IR region. Meanwhile, a thorough compilation of Raman spectra of various functional groups are readily available for spectral library search and fingerprinting. Besides the rich spectral information, samples for Raman measurement could be presented as gas, liquid, and solid with little sample preparation. Especially water, which is a strong absorbent in IR, is an exceptional solvent for Raman spectroscopy because of its weak Rayleigh scattering. Other than the sampling advantages, Raman provides additional measuring features such as resonance Raman, surface enhanced Raman, and polarization measurements to determine symmetrical vibrations.

1.2 Theoretical Background

Raman scattering and IR absorption are complimentary processes in elucidating molecular structures (Griffiths 2010). The frequency at which the absorption of IR radiation by molecules occurs depends on vibrational frequency of the molecules. The intensity of IR absorption depends on the strength of the change in dipole moment (Wartewig 2003). In Raman scattering, the frequency of incident photons is usually a lot greater than the frequency that is required in the IR absorption. According to the classical theory, the intensity of scattered Raman signal is proportional to the fourth power of the excitation frequency and to the square of the change in polarizability with the normal coordinates of the vibration (Long 2002). There exists another one fundamental difference between the two processes, i.e., the occurrence probability. The IR absorption is more likely to happen than Raman scattering. However, many symmetric vibrations which are forbidden in IR absorption are usually very prominent spectral features in Raman. Asymmetric vibrations in Raman, though much weaker than their counterparts in IR absorption, do appear as Raman signatures as well (McCreery 2000). In general, the dual process of both IR absorption and Raman scattering for the same molecular vibrations are identical in frequency, they are not duplicates of each other. Fingerprinting of molecular structure could be more complete with spectral data obtained from both techniques available. Though the NIR absorption spectroscopy provides less molecular information compared to IR spectroscopy, it is convenient and appropriate for routine analysis. Raman technique, especially Raman with NIR excitation, along with NIR spectroscopy could possibly reinforce each other to form a hybrid technique. Theoretically, the complimentary information that Raman scattering and NIR absorption provides may suggest a more versatile routine chemical analysis tool.

Mathematically, utilizing different types of indirect measurement data on the same subject matter could possibly reduce the size of the class of non-unique solutions. In that way, more stability is introduced into the calibration model. The term ‘joint inversion’, i.e., the joint interpretation of different types of measurement data, was cited from the research work done by Jupp and Vozoff (1975). In their paper, the authors reported the advantage of including both electric and magnetic induction data in the calculation of electrical conductivity structure. Anderssen et al. (2005) borrowed the concept and modified it with a method called leap frog calibration-and-prediction (LP-CAP). The authors applied the method to the joint inversion of Raman and NIR spectra of wheat flours. In terms of standard deviation of differences (SED) between reference values and predicted values, combined NIR and Raman data offered slightly improvement in predicting pentosan (0.13 vs 0.16) and gliadin/glutenin ratio (0.02 vs 0.02) when compared with results using NIR spectra alone. For predicting water absorption and total protein, NIR technique proved to perform better compared to joint NIR/Raman technique. The author believed that the lack of spectra features related to water and total protein in Raman spectra could explain the performance degradation when combining NIR and Raman data. Therefore, it could be stated that inclusion of both NIR and Raman data might improve the confidence level for prediction models. However, the quality of Raman and NIR absorption spectra are widely different due to the occurrence probability. Therefore, the significance of the spectra quality, i.e, signal-to-noise ratio (SNR), should not be underestimated in quantitative analysis. On the other hand, the qualitative analysis ability of NIR could be significantly improved in classifying samples with the library search capacity of Raman technique.

1.3 Practical Implications

The days when Raman spectroscopy for the ‘real world’ chemical analysis was considered merely an academic pursuit has long gone since Hirschfeld and Chase (1986) published their paper in *Applied Spectroscopy*. Nowadays, several spectroscopic equipment manufacturers such as Bruker and Perkin-Elmer offer commercial FT-Raman or dispersive spectrometers that enabled Raman spectroscopy a useful tool for many practical applications. The resurgence of interest in Raman spectroscopy since year 1986 has mainly been attributed to technical advances in spectrometric devices, in particular, instrumentation for NIR spectroscopy. Not surprisingly, dispersive instruments for NIR absorption and Raman scattering measurement share significant amount of similarities.

Using visible excitation, the detection unit of Raman spectrometer gets saturated easily by fluorescence background, which inundates the faint Raman signal. This is especially true for complex biological samples with high level of auto-fluorescence. During the post-1986 era, emerging diode lasers and yttrium-aluminum-garnet (Nd:YAG) laser were introduced as excitation light sources for Raman spectroscopy. These types of lasers emit laser light at longer wavelengths into the NIR region. Because NIR excitation produces much less fluorescence than the visible lasers, it drastically changes the situation for Raman technique as a routine analytical tool. To reduce the amount of excitation photons reaching the detector, a holographic notch filter needs to be placed in the receiving optical path to reject Rayleigh scattered light to a level from 10^{-5} to 10^{-10} or even less. For weak scatterers, a longer integration time, e.g., several minutes, is often required to acquire spectra with a decent SNR.

Nowadays, modern analytical chemistry labs tend to be equipped with both the Raman and FT-IR instruments for chemical structure elucidation. The combined analytical capacity of FT-IR and Raman enabled spectroscopy practitioners to investigate a wide range of samples, especially when non-trivial sample preparation and controlled environment are required. For industrial *on-line*, *in situ*, and *at-line* routine analysis, minimum sample preparation is expected and production environment could be harsh and dusty. In such a scenario, both Raman and NIR spectroscopy provide more flexible and robust solutions compared to FT-IR spectroscopy (Siesler 2008). The sampling ease of Raman is also perfectly matched to that of NIR technique, which allows placing the spectrometer away from the sampling point using fiber optic probes. Compared to the highly overlapping NIR absorption spectra, Raman spectra provide highly specific and complimentary structural information. Therefore, an alliance of the two techniques, i.e., Raman and NIR absorption, for practical chemical analysis seems not only favorable but also feasible. Nowadays, low cost NIR spectrometers are widely available. For many types of samples, NIR spectra with high SNR could be acquired within a fraction of second. On the other hand, the prohibitive capital costs and moderate quality of Raman spectra hindered the acceptance of Raman technique in industrial settings. Fast acquisition of Raman spectra with good SNR is becoming possible thanks to technological advances in photonics and electronics. Overall, instrumentation for Raman spectroscopy is becoming more sensitive, compact, and affordable (Sorak et al. 2012).

1.4 Choice of Instrumentation and Justification

Quality plays important roles in a wide range of industries from chemical and food production to process development in pharmaceutical field. Worldwide, there exist ever

increasing legislative and economical requirements on product quality along with demands on environment protection and production plant safety. Ensuring steady quality that adheres to the legal limits regarding product composition is paramount to the manufacturers. Therefore, analytical tools that offer rapid, non-destructive, robust, and economic quality monitoring solutions are highly desirable. The industrial processes also pose unique issues that should be addressed by the instrumentation. On a production line, multiple sampling points are often required to study the kinetics or chemical reactions and these signals are preferably collected by a single process analyzer. Often, the production plant environment is humid and dusty, has extreme temperatures, contains explosive vapor, and undergoes constant vibration (Slater et al. 2001). Therefore, the practical solution is to install the detection portion of the analyzer, i.e., the spectrograph and detector, in a safe and environmentally controlled room. Then the remote probe head can be connected through a window to the detection unit using fiber optics.

Vibrational spectroscopy techniques have been adopted for process analyses since 1980s (Tate et al. 2002). During the last decade of the 20th century, NIR spectroscopy has quickly surpassed FT-IR spectroscopy as the method-of-choice for process monitoring (Workman 1999). When it comes to remote sensing using long and low-cost fibers, both NIR and Raman techniques definitely have significant advantages (Marteau et al. 1995). The robust configuration of dispersive Raman/NIR spectrometer usually means excellent reliability. However, proper maintenance and repair work are required to ensure the proper functioning of process analyzers.

The purpose for undertaking this research is to look for an affordable and versatile spectroscopic tool for chemical analysis applications. Ultimately, the optical apparatus is

expected to be used for *in-situ* raw material inspection, *on-line* process control, and finished product verification. Therefore, the spectrometric system needs to be robust, compact, portable, and capable of working with fiber optic sampling accessories for remote sensing. A myriad of instrumentation options are available for realizing Raman scattering and/or NIR absorption measurement separately (Table 1.1). The FT systems offer excellent wavelength accuracy, spectral resolution, and high SNR. However, the FT systems fall short in Raman applications for problems such as low sensitivity and multiplexing due to Rayleigh line (Slater et al. 2001). The optical cavity of device requires regular purging of CO₂ and H₂O vapor and is susceptible to mechanical vibrations. It is also not amenable to long distance remote sensing even with the use of ATR probe and fiber optics. For Raman applications, if the fluorescence is not significant, it is advisable to go with other instrumentation options. After all, the cost of such system is prohibitive and it is best utilized in a lab environment. Electronic tunable filters (ETF) such as liquid crystal tunable filter (LCTF) and acousto-optic tunable filter (AOTF) are new alternatives to the conventional grating/prism based dispersing devices. These ETFs are fairly expensive and often utilized where both moderate to high spatial and spectral resolutions are demanded. There are LCTF based Raman imaging systems capable of achieving spatial and spectral resolutions of 0.25 μm and 9 cm⁻¹, respectively (Treado and Nelson 2002). Operation of AOTF requires rigid temperature regulation up to 0.001 °C to ensure wavelength stability (Tate et al. 2002). Therefore, for *in-situ*, *on-line*, and *in-line* process monitoring, factors such as cost and moderate spectral performance of ETF based spectrometers eluded themselves from further consideration. Grating and prism based spectrometers often offer good compromise between cost and

performance. Monochromators and spectrographs with grating as the dispersive element are compatible with highly sensitive multi-array detectors such as charge coupled device (CCD) and Indium gallium arsenide (InGaAs) cameras. Depending on the required performance characters, the price range could go from a couple of thousand to an amount exceeding \$200,000 for some state-of-the-art designs. Single stage and mutli-stage monochromators require mechanical rotation of the grating for wavelength tuning, which negatively affects its long-term reliability as required by process monitoring industries. In addition, monochromators are often bulky, suffer low light transmission, and become more expensive as design complexity increases. Spectrographs coupled with highly sensitive CCD and InGaAs cameras offer rugged design and acceptable performance characteristics at an affordable price in comparison to other instrumentation options. In addition, the spectrograph/multi-array detector configuration is amenable to multiplexing and multi-channel concepts, which enables the key design idea in an integrated multimodal spectrometer. However, at current stage, InGaAs cameras coupled with suitable spectrograph are fairly expensive compared to their CCD coupled counterparts, which have matured through the years.

For the Raman channel, it is also important to select the appropriate laser excitation source to work with the spectrograph/CCD detection unit. In choosing a laser for process monitoring, cost, power (tens to a few hundred milliwatts), wavelength stability, air-cooling, and operating life are some of the most important factors to consider. A wide array of laser sources for Raman channel is available. Argon ion lasers (488 nm), neodymium-doped yttrium aluminum garnet (Nd :YAG, 532 nm), and external-cavity diode lasers (ECDLs) have all been utilized for process monitoring

Table 1.1 Comparison between different instrumentation options

Instrumentation	Speed	Wavelength accuracy	Spectral resolution	Cost	Applications	
Fourier Transform	Medium-low	Excellent	Excellent	High	Analytical lab	
ETF	LCTF	High	Good	Good	High	Spectral imaging
	AOTF	High	Good	Moderate	High	
Monochromator	Low	Moderate	Good	Low-High	Laboratory	
Spectrograph	CCD	High	Good	Good	Moderate	Process control and imaging
	InGaAs	High	Good	Good	High-moderate	

Source: Chalmers and Griffiths (2002)

(McCreery 2000). For samples that fluoresce under visible excitation, there is the need to migrate the excitation wavelength into the red and near-infrared. It is also important to shift to sufficiently long wavelengths to suppress the fluorescence without severely weakening the Raman signal. Therefore, the commonly used 785 nm near-infrared excitation achieved using ECDL seems to be the optimum choice for the application. However, caution should be made that sensing ability of Raman signal above 1500 cm^{-1} is significantly reduced due to the drop in quantum efficiencies of CCD detector after 900 nm.

In summary, a spectrograph coupled with multi-array detectors, e.g., CCD or InGaAs camera, seems to be the instrument of choice to achieve integrated multimodal spectroscopic measurements. Because there are no moving parts, the instrument design is more robust than the FT systems. For NIR spectroscopy, a dispersive spectrograph based on fixed optics is more robust and affordable than FT systems for *in-situ* applications. For Raman excitation, near-infrared ECDL with emission wavelength at 785 nm is a good choice regarding the balance between fluorescence suppression and signal sensitivity. Fiber optics, which enable remote sensing, also work great with both NIR and Raman spectroscopy.

1.5 Feasibility of Proposed Instrumentation

A spectrograph coupled to a CCD is an optimum choice when a balance is sought between cost and performance. To enable the same detection system to acquire both signal channels, i.e., Raman and NIR, the array detector could be allocated into two separate imaging areas for each channel. Besides their similarities, optics for Raman study differs from that used for NIR absorption in the following aspects. Raman uses NIR

laser instead of halogen tungsten lamp as the excitation light source. The intense Rayleigh scattering in the Raman channel requires notch filter, band-pass filter, and possibly polarizers and polarization unscramblers to be placed in the optical path. Some drawbacks of the dispersive system with CCD detector are limited spectral range and moderate spectral resolution. With a 785 nm laser as excitation source, the observable Raman spectrum range could be from 100 cm^{-1} to 3240 cm^{-1} in wavenumbers, which covers most of the typical Raman shift range of interest (0 cm^{-1} to 3500 cm^{-1}). Thus, the impact of spectral range on Raman measurement is not significant. The spectral region from 785 nm to 1000 nm contains the second or third overtone of C-H, O-H, and N-H stretching bonds (Osborne et al. 1993). Compared to those in the short-wavelength IR range (1000 nm to 1700 nm), the NIR absorption features are relatively weak in this spectral region. Nevertheless, the short-wavelength NIR spectral region (780-1000 nm) is a good overlapping area for both techniques. Spectral resolution plays less important role in NIR absorption measurement because absorption features are broad and highly overlapping. For the Raman measurement, it could mean degradation in performance especially when frequency shift is concerned. In general, it is possible to achieve both Raman and NIR absorption measurement using the same detection system. For sampling ease, both techniques enjoy minimal sample preparation and are capable to work with fiber optics for remote sensing. The NIR reflectance spectroscopy and Raman spectroscopy share the same 180° back-scattering sampling geometry, which is experimentally convenient and common in commercial instruments.

1.6 Research Objectives

As previously discussed, there exists theoretical and experimental proof behind the idea of multi-modal spectroscopy by synthesizing NIR-Raman with NIR absorption techniques. Practicing both Raman and NIR absorption spectroscopy using a multi-modal spectrometer has its scientific foundation and technical merits. On the other hand, the prospect of allying NIR absorption and NIR-Raman techniques is challenging. The practical problems mainly come from the development of hardware system and handling the multimodal data it produces. As of this date, reported findings on multimodal spectral measurements were performed using separate NIR and Raman spectrometers. Therefore, a multimodal spectroscopic system for both Raman and NIR absorption measurement is desirable but seems to remain unavailable. The main objective of this research is to develop a multimodal spectroscopic system for Raman and NIR absorption measurements. The cost of the procured equipment placed certain restraints on our capability to carry out the project in an optimized manner. Therefore, compromises have to be made to sacrifice certain desirable features, e.g., sampling NIR absorption in the longer wavelength NIR region. This research project is mainly intended to serve as a proof-of-concept of a novel spectroscopic technique, which could potentially be used for at-line, on-line, and in-line analysis of food and pharmaceutical productions. The prototype device is the first of its kind and did not evolve from any affiliated research project. Therefore, completion of the device and its application in sampling analysis demonstrated the technique's viability and helped formulating the next stage of research and development. This research is not intended as an immediate solution to the needs as we observed in a range of industries. The following three research objectives are proposed.

1. To design, construct, calibrate, and test a prototype multimodal spectrometer capable of performing both NIR absorption and Raman measurements.
2. To apply the instrument to the qualitative and quantitative analysis of various samples, e.g., Acetaminophen, aqueous solutions of ovalbumin and ethyl alcohol.
3. To explore the fusion of multimodal spectra data in the analysis of target analyte and two-dimensional correlation spectroscopy.

The next few chapters of this dissertation are organized as follows: literature review, materials and methods, results and discussion, conclusion, and recommendations. Literature review covers the theoretical foundations for vibrational spectroscopy and an updated overview of applications relevant to analysis of ovalbumin and ethyl alcohol in aqueous solutions. Materials and methods section details the design and test procedures of the proposed device. The materials and method section also describes sample preparation and data analysis methods. Experimental data pertinent to the outlined objectives are presented in the results and discussion section. Interpretation of the results followed by discussion is also provided. The conclusion section lists the major findings in this research project. In the end, a few recommendations are given regarding the future direction of multimodal spectroscopy research.

2. LITERATURE REVIEW

2.1 Basis for Vibrational Spectroscopy

The infrared region is divided into near-infrared, mid-infrared, and far-infrared region for different wavelength ranges. The corresponding frequency (Hz) for wavelength in the infrared region, which is usually large, is commonly denoted as wavenumber (cm^{-1}). Near-infrared region covers the spectral range of 780-2500 nm ($12,800\text{-}4000\text{ cm}^{-1}$). Mid-infrared covers the region of 2500-50,000 nm ($4000\text{-}200\text{ cm}^{-1}$). Far-infrared covers the region of 0.05-1 mm ($200\text{-}10\text{ cm}^{-1}$).

2.1.1 Infrared and near-infrared spectroscopy

Mid-infrared spectroscopy, commonly known as FT-IR spectroscopy, is a popular analytical technique for laboratory uses in many fields such as pharmaceutical and food research. The molecular fingerprints in IR absorption spectra and the Beer's law on absorption intensities make FT-IR spectroscopy a good candidate for both qualitative and quantitative analysis. However, the strong absorption of the molecules in the IR region hampers its application in fields such as food science, where most samples are not susceptible to direct transmission measurement (Li-Chan 2010). For transmission measurement by IR spectroscopy, the sample commonly has to be prepared as thin film with a thickness of about 10 μm . The reason is that an IR spectrum of such thin film could yield absorption bands that are neither saturated nor too weak. In recent years, research activities in FT-IR spectroscopy picked up quickly due to the wide utilization of attenuated total reflection (ATR). Compared with transmission, the ATR technique requires less sample preparation in order to obtain consistent results. On the down side,

the internal reflection element (IRE) is susceptible to contamination by chemical agents such as hand lotion, silicon grease or even greasy fingerprints, etc. Therefore, the possibility of sampling real-world samples using ATR FT-IR seems to be restrained by practical conditions imposed on the ATR devices.

Absorption of light in the NIR region is comparatively weak which allows the light to penetrate deeper into the sample. The interference arising from the strong absorption of water in the IR region is greatly ameliorated in the NIR spectrum due to the low absorptivity in the NIR region. Thus, NIR spectroscopy seems to handle foodstuff better than mid-IR spectroscopy. However, some of its technical drawbacks, e.g., poorly understood prediction models, restrict its use in general food laboratories (Williams 2007). To further understand the fundamentals behind the difference in spectral characteristics of absorption in the NIR and IR region as well as Raman scattering, a brief review of the origin of vibration spectra is provided.

2.1.2 Molecular rotation and vibration

The molecular behaviors under external electromagnetic radiation are the keys to understanding the light absorption processes in the IR and NIR region. At room temperature, all molecules possess certain amount of energy. The energy of a molecule can be categorized into four different types. The four different energy types are: translational energy, rotational energy, vibrational energy, and electronic energy (Mortimer 2008). For a first degree approximation, the contribution of these different types of energies could be considered separately from each other. Their interaction with electromagnetic radiation could further segregate the distinction among the different energy types. For electronic energy transfer, the transition incurs absorption or emission

in the ultraviolet and visible regions. Energy transfer among the molecular vibration states contributes to the absorption lines and bands throughout most of the infrared region of the spectrum. Pure rotation gives rise to absorption in the microwave or the far-infrared region. The vibrational and rotational frequencies of molecules occupy the IR region. Such frequency features relating to the molecular structures could be studied using IR absorption spectroscopy and Raman spectroscopy. With different selection rules as mentioned in introduction, Raman and IR techniques reveal complimentary information of the molecular structures.

2.1.3 Molecular vibration

Supposing there are N atomic nuclei in a molecule, then there exists $3N$ degrees of freedom of motion in total for all the nuclear masses in that molecule. Considering the molecule as a whole, the centre of gravity of the molecule needs 3 coordinates to define its spatial position, which correspond to the three independent degrees of freedom of motion. A nonlinear molecule in its equilibrium configuration needs 3 rotational coordinates to define the orientation of the molecule towards the centre of gravity. Therefore, a nonlinear molecule possesses 3 independent rotational degrees of freedom while a linear molecule only has two. Taking out the total independent degrees of freedom from the total of $3N$ degrees of freedom for a molecule, we obtain $3N-6$ internal degrees of freedom for a nonlinear molecule and $3N-5$ internal degrees of freedom for a linear molecule. It was shown that the number of internal degrees of freedom corresponds to the number of independent normal modes of vibration for the molecule.

Considering the simple model for a diatomic molecule by connecting two masses m_1 and m_2 using a massless spring, the classical vibrational frequency formula could be

derived. Hooke's law has been used to determine such vibration frequency of the “ball and spring” model (Raghavachari 2001). It is a simple and straightforward way to look at the molecular vibration using such a mechanical model that describes the diatomic molecule as two masses attached to the two ends of a spring with a force constant (k). Any pair of bound atoms in a molecule exhibits a natural oscillation and vibrates at a specific frequency given by:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (2.1)$$

where k is the force constant ($\text{N}\cdot\text{m}^{-1}$), and μ (kg) is the reduced mass given by:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (2.2)$$

where m_1 and m_2 are two masses in a diatomic molecule (kg).

According to the mechanical model, the vibrating system only absorbs energy from externally applied forces that have an oscillating frequency matching that of the natural frequency of the “ball and spring” system. External forces at non-resonant frequencies do not increase the energy of the system. Similar to the external forces applied to the “ball and spring” system, the infrared radiation absorbed by the molecule must match specific frequencies to increase molecular vibration energy. Therefore, when irradiating samples with incident light having a continuous spectrum in the infrared, only radiation which matches the vibrational frequencies of molecules in the samples is absorbed. When a spectrometric system is used to receive the transmitted infrared photons, a spectrum with absorption features at specific frequencies is recorded. The

absorption intensities at the natural vibrational frequencies reflect the effectiveness of energy transfer from external IR radiation to the molecular vibration system. The term dipole moment is introduced to explain the effectiveness of energy transfer. The definition of dipole moment is the product of the magnitude of charge in the dipole and the charge spacing. In order to be infrared active, the dipole moment of a molecule must be changed by the molecular vibration simultaneously. The intensity of an infrared absorption band is proportional to the squared change in the dipole moment which is incurred by the molecular vibration causing such absorption band.

2.1.4 Anharmonicity, overtones, combination, and difference

Simplified diatom vibration model treat the molecular vibration as harmonic. In the real molecular system, the repulsive and attractive forces between atoms are not linearly proportional to the nuclear displacement coordinate. This introduces mechanical anharmonicity. Similarly, if the dipole moment changes in a nonlinear way with the change in nuclear displacement coordinates, electrical anharmonicity is introduced into the system. Either mechanical or electrical anharmonicity causes the dipole moment changes periodically with time. The periodic wave could be decomposed into simple sine terms with frequencies equal to integral multiples of the fundamental vibrational frequencies of the molecule. Therefore, the dipole moment oscillates at the fundamental frequency and its integral multiples for anharmonic vibration of the molecule. The integral multiples are referred to as the first overtone, second overtone, etc.

The combination and difference bands could be found in the spectra of polyatomic molecules because of $3N-6$ internal degrees of freedom they possess. In theory, if two or more fundamental bands interact, combination and difference bands

appear at frequencies that are the sums or differences of multiples of the fundamentals as defined in the following equation:

$$\nu_{comb,diff} = n_1\nu_1 + n_2\nu_2 + \dots \quad (2.3)$$

where $\nu_{comb,diff}$ is the combination or difference band frequency, and n_1, n_2, \dots are positive integers. The absorption of IR radiation involving the combination bands cause transition from the ground level to a combination energy level. Transition to combination level is a low probability phenomenon unless such transition is caused by no more than two vibrations with bonds that are either connected through a common atom or multiple bonds. Difference bands involve transition from an excited level of one vibration to an excited level of higher frequency vibration. Because difference bands do not involve transition from ground level, they are of even lower possibility and disappear at low temperatures.

Overtone, combination, and difference bands are weak spectral features compared to fundamentals. For example, the first overtones and combinations are usually 10 to 100 times weaker than the fundamentals. Higher order overtones and combinations are even weaker (Blanco and Villarroya 2002). Though NIR region is sometimes called the overtone region, many of the overtone and combination bands absorb in the IR region. Only the overtones of O-H, N-H, and C-H stretching modes along with the combination bands originating from stretching-bending modes of these bonds are found in the region above 4000 cm^{-1} , i.e., the NIR range (Kaye 1954).

2.1.5 Vibrational potential energy, quantum effect

According to the harmonic vibration model of a diatomic molecule, the potential energy is a function of the internuclear distance. The origin of the internuclear potential energy can be explained by the existence of electronic and nuclear repulsion energy. When two atoms move closer to each other, the internuclear distance becomes smaller resulting in internuclear repulsion. Vice versa, when two atoms move away from each other, the resulting larger internuclear distance leads to two atomic components attracting each other. At an equilibrium distance r_e , vibrational potential energy is at its minimum. The potential energy changes tend to level off when the internuclear distances become large. When the distance grows large enough, the force breaks the chemical bond and dissociates the molecule. For internuclear distances around the equilibrium distance r_e for a diatomic model, the harmonic oscillator of potential function provides a good first approximation. The energy can be expressed by:

$$E = \frac{1}{2}k(r - r_e)^2 \quad (2.4)$$

where E is the potential energy, k is a constant that depends on the strength of molecular force field holding two atoms, and r is the internuclear distance.

Further away from the equilibrium position, the coulombic repulsion between the two nuclei causes the potential energy to rise rather rapidly than the prediction from harmonic equation. On the other end, the potential energy levels off when the internuclear distance approaches dissociation spacing. An empirical function, due to Morse, offers better approximation:

$$E = E_D(1 - e^{-\beta(r-r_e)})^2 \quad (2.5)$$

where E_D is the dissociation energy, β is a constant.

Using Taylor series to expand Eq. 2.5, the following equation is derived.

$$E = E_e + \left(\frac{\partial E}{\partial S}\right)_e S + \frac{1}{2!} \left(\frac{\partial^2 E}{\partial S^2}\right)_e S^2 + \frac{1}{3!} \left(\frac{\partial^3 E}{\partial S^3}\right)_e S^3 + \dots \quad (2.6)$$

where S is the internal coordinate, i.e., $r-r_e$. To describe molecular vibration, term E_e , i.e., energy at equilibrium, can be set to zero. The linear change rate of potential energy at equilibrium, i.e., $(\partial E/\partial S)_e$, also equals zero. If starting from the term S^3 , the higher order terms in the expansion are neglected and the term $(\partial^2 E/\partial S^2)_e$ is replaced with a force constant k , the harmonic approximation equation in Eq. 2.4 is obtained.

From quantum mechanics theory, several conclusions are introduced to explain the harmonic molecular oscillation. Using Schrödinger's generalized wave equation and assuming constant potential energy, it is concluded that molecular energy only has discrete values where the quantum number can only allow integral numbers. That means the molecular energy is not continuous but quantized instead, which is a consequence of the wave-particle duality of matter. The same restriction, i.e., energy levels are quantized, applies to the molecular rotation as well. Another conclusion is that the momentum can never be zero for the oscillating particle in the model because wavelength cannot be infinite. Therefore, as part of molecular energy, vibrational energy can only have discrete values and the lowest possible value cannot be zero. The discrete energy levels are described in the following equation:

$$E = \left(v + \frac{1}{2}\right) h\nu \quad (2.7)$$

where E is the discrete energy level value, v is the quantum number, which is only allowed to have integer values (i.e., 0, 1, 2, ...), h is the Planck's constant (6.63×10^{-34} J s), and ν is the classical vibration frequency. When the quantum number equals zero, the oscillator possesses the lowest vibrational energy of $h\nu/2$.

Making the necessary adaptation to the anharmonic oscillator equations according to the quantum mechanics theory yield the following equations:

$$E = \left(v + \frac{1}{2}\right) h\nu - \left(v + \frac{1}{2}\right)^2 x h\nu + \dots \quad (2.8)$$

where x is called the anharmonicity constant, which is dimensionless and has a value between 0.001 and 0.02 (Griffiths 2010). One direct result of anharmonic oscillator model is that energy levels are not equally spaced and frequency is not completely independent of vibration amplitudes. Therefore, overtone transitions, e.g., from $v=0$ to $v=2$ or 3, which are allowed in anharmonic model, do not appear at exactly 2 or 3 times the fundamental frequency.

2.1.6 Rotational transition and vibrational-rotational spectrum

As stated earlier, pure rotational absorption happens in the microwave and far infrared region. However, for some molecules in gaseous states, their vibrational-rotational spectrum could be explained by the theory of rotational transitions and its relation with infrared absorption. For a linear molecule, the quantum mechanical rotational energy is at discrete levels given by:

$$\frac{E_{rot}}{hc} = BJ(J + 1) \quad (2.9)$$

where E_{rot} is the rotational energy, B is the rotational constant, and J is the rotational quantum number with only integer values 0, 1, 2,... allowed. The prerequisite for a molecule to absorb radiation energy to increase its rotational energy is that the rotating molecule must have a permanent dipole moment. If the rotating linear molecule is put into the electric field of a plane polarized electromagnetic wave, the electric field tries to make the portion of dipole moment along the electric field direction oscillate at the frequency of incoming electromagnetic radiation. Exposure of molecule with its dipole moment rotating at a frequency near that of radiation photons for sufficiently long time causes an energy transition in its rotational energy levels. Because the absorbed photon energy equals the increase in rotational energy, the frequency of the absorbed photons expressed in wavenumber is given by:

$$\bar{\nu}_p = 2B(J + 1) \quad (2.10)$$

where $\bar{\nu}_p$ is the frequency in cm^{-1} of the photon with the right energy to promote the rotational energy level from J to $J+1$. It is then clear from the equation that adjacent rotational absorption lines are equally spaced with a spacing of $2B$ in cm^{-1} . Because of Boltzmann distribution (Eq. 2.11), all the rotational absorption lines do not share the same intensity. The intensity is proportional to the population of molecules at the J_{th} rotational energy level.

$$\frac{n_i}{n_0} = \frac{e^{-E_i/kT}}{e^{-E_{i0}/kT}} \quad \text{or} \quad \frac{n_i}{n_0} = e^{-(E_i - E_0)/kT} \quad (2.11)$$

where n_i and n_0 denotes the total numbers of molecules existing in the i th and zero energy levels respectively, E_i and E_0 are the energy at the i th and zero energy levels, k is the Boltzmann constant ($1.380\ 6504 \times 10^{-23} \text{J}\cdot\text{K}^{-1}$), and T is the absolute temperature (K).

According to the Boltzmann function, the population of molecules n_j in the J_{th} level relative to n_0 , i.e., the number of molecules in the $J=0$ level, is calculated according to:

$$\frac{n_j}{n_0} = (2J + 1)e^{-J(J+1)Bhc/kT} \quad (2.12)$$

In case of observing the IR spectra of small molecules in the vapor state, lots of fine structures are revealed because of the simultaneous quantized vibrational and rotational transitions. Though similar phenomena are rare with larger molecules in vapor phase or any molecule in liquid state, it is worth studying the theory behind the vibrational-rotational spectrum because of possible presence of such molecules as interference in IR spectroscopy. In a classical mechanical model, the linear molecule vibrates along its axis and rotates about its center of gravity at the same time. In the presence of external radiation, the dipole moment along the electric field direction of the incoming photons oscillates with frequencies of $\nu_v + \nu_r$ and $\nu_v - \nu_r$, where ν_v is the vibrational frequency, and ν_r is the rotational frequency. Let's consider ν_v to be the same for numerous molecules in the vapor phase. The rotational frequency ν_r varies and has a distribution function defined by the Boltzmann distribution function. With a low resolution IR spectrum of an unsymmetrical diatomic molecule in its vapor phase recorded, its wing shape doublets can be observed. For a linear polyatomic molecule, the dipole moment can change along the molecular axis as well as perpendicular to the

molecular axis. The two types of dipole moment changes correspond to the antisymmetrical stretching and bending vibration of the molecule. Therefore, the molecular vibration results in two types of vibrational-rotational bands called parallel band and perpendicular band. The selection rule for a parallel band is $\Delta J = \pm 1$ and the selection rule is $\Delta J = 0, \pm 1$ for the perpendicular band. That means for the bending mode of molecular vibration, vibrational energy transition without a simultaneous rotational transition is allowed. When $\Delta J = 0$, a strong absorption line results in the spectrum, which is called the Q branch in the perpendicular band. When $\Delta J = 1$, corresponding rotational bands make up the P branch. When $\Delta J = -1$, the rotational bands constitute the P branch. For parallel bands of a linear molecule, they share the identical P and R branches with the perpendicular bands.

2.1.7 Raman Spectroscopy

The IR spectroscopy studies only one transition, which is the direct transition between two energy levels through the process of absorption. The Raman scattering process involves two transitions and three different vibrational states. To ensure that Raman effect occur, an intermediate level called the virtual state with a higher level of energy than the initial level and final level of transition must exist. Otherwise, the transition defined in Raman process is forbidden. In Fig.2.1, the absorption and scattering processes are illustrated. In Raman spectroscopy, the sample is illuminated with a monochromatic beam of intense radiation, which is usually a laser source. The frequency of the excitation radiation is much higher than the vibrational frequencies encountered in IR spectroscopy but is usually lower than the electronic frequencies. All incident photons that interact with the sample molecules increase the vibrational and rotational energy of the molecules

to the virtual state by the amount of energy equal to $h\nu_0$, where ν_0 is the frequency of incident photons. Immediately after the transition to the virtual state, most molecules involved in the interaction return to the ground state through the process of emission of photons at the same frequency of the incident photons. This process is called Rayleigh scattering and is an elastic collision process between the incident photon and the molecule. Energy of molecules that undergo Rayleigh scattering is unchanged and the scattered photon has identical frequency with the incident photon. Rayleigh scattered photons account for most of the scattered photons. Raman effect is an inelastic collision process between the incident photons and the sample molecule. During the inelastic scattering, the energy of the incident photons and the vibrational-rotational energy of molecule are altered. A small fraction of the incident photons loses energy as a result of a small proportion of molecules falling to the first excited energy level instead of the ground energy level. Correspondingly, the scattered photons with less energy appear in the Raman spectrum as a source of emitted radiation with increased wavelength. This process is known as Stokes Raman scattering and the spectral lines are called Stokes lines. Similarly, it is also possible for some of the molecules already in an excited vibrational state to be exposed to the laser radiation and promoted to a virtual level with unstable high energy and then return to the ground state. This type of inelastic collision is called anti-Stokes Raman scattering. In anti-Stokes Raman scattering, the molecules lose energy which gives rise to spectral lines of increased frequency compared to excitation radiation in the Raman spectrum. According to Boltzmann distribution, most molecules exist in the ground state instead of excited states at ordinary temperatures. Therefore,

Stokes lines have greater intensities than anti-Stokes lines because excited energy level has lower population compared to the ground state.

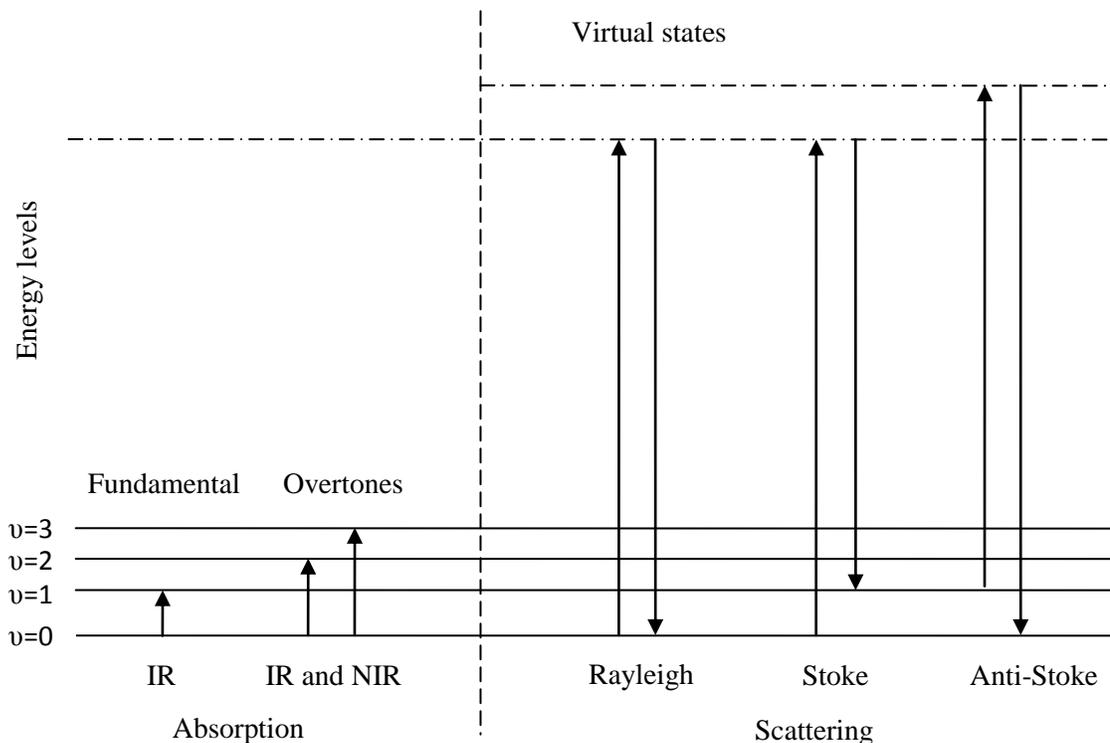


Fig. 2.1 The absorption and scattering processes during vibration energy transitions

Source: McCreery (2000)

2.1.8 Polarizability

A classical theoretical treatment of Raman and Rayleigh scattering involves the fact that scattered photons are the results of oscillating dipole moment of the molecule induced by the electric field of incident light wave. When a molecule is subject to the electric field of incident radiation, its protons and electrons receive electric forces in the opposite directions. Such oppositely directed forces cause a change in the charge spacing, which induce a dipole moment for the polarized molecule. Polarizability is used to describe the

relation between electric field of incident light wave and the induced molecular dipole moment. According to Eq. 2.13, polarizability is defined as the induced dipole moment divided by the electric field strength that leads to the induced dipole moment.

$$\alpha = \mu/E \quad (2.13)$$

where α is the polarizability ($\text{C m}^2 \text{ V}^{-1}$), μ is the induced dipole moment (C m^2), and E (V) is the strength of electric field for the incident radiation. The molecular polarizability is not a constant and its value depends on the displacements the electrons make relative to the protons during vibrations and rotations. The dependence of polarizability on normal coordinates Q could be expressed in an expanded form using Taylor series.

$$\alpha = \alpha_0 + \frac{\partial\alpha}{\partial Q} Q + \frac{1}{2!} \frac{\partial^2\alpha}{\partial Q^2} Q^2 \dots \quad (2.14)$$

where α_0 is the polarizability near equilibrium, Q is the normal coordinate, and $\partial\alpha/\partial Q$ is the molecular polarizability changing rate with respect to Q taken under the equilibrium condition. For small amplitude vibrations near the equilibrium, higher order terms in Eq. 2.14 are neglected and linear dependence of polarizability on Q is assumed.

$$\alpha = \alpha_0 + \frac{\partial\alpha}{\partial Q} Q \quad (2.15)$$

The normal vibrations are treated as harmonic and the normal coordinate Q varies periodically according to:

$$Q = Q_0 \cos 2\pi\nu_v t \quad (2.16)$$

where Q_0 is the maximum for Q , ν_v is the harmonic frequency for the normal coordinate vibration (Hz), and t is time (s). Substituting the Q in Eq.2.16 with that in Eq.2.15, the

polarizability is expressed as

$$\alpha = \alpha_0 + \frac{\partial \alpha}{\partial Q} Q_0 \cos 2\pi\nu_v t \quad (2.17)$$

Now consider the time domain. The induced dipole moment caused by the electric field of the incident radiation varies with time. The strength of electric field around the molecule varies with time and is given by:

$$E = E_0 \cos 2\pi\nu t \quad (2.18)$$

where E_0 is the maximum value of the electric field (V), ν is the frequency of the radiation (Hz), and t is time (s). Under the influence of electric field of the incident radiation, the time dependence of induced dipole moment is expressed as Eq. 2.19 by combining Eq. 2.17 and Eq. 2.18:

$$\mu = \alpha_0 E_0 \cos 2\pi\nu t + \frac{\partial \alpha}{\partial Q} Q_0 E_0 (\cos 2\pi\nu_v t)(\cos 2\pi\nu t) \quad (2.19)$$

Using trigonometric identity on the second term containing two cosine functions in Eq. 2.19, the following is obtained:

$$\mu = \alpha_0 E_0 \cos 2\pi\nu t + \frac{1}{2} \frac{\partial \alpha}{\partial Q} Q_0 E_0 [\cos 2\pi(\nu - \nu_v)t + \cos 2\pi(\nu + \nu_v)t] \quad (2.20)$$

From Eq. 2.20, there are three cosine functions with three different component frequencies, i.e., ν , $\nu + \nu_v$, and $\nu - \nu_v$. In the classical treatment, the polarized induced dipole radiates light at the frequency of its oscillation. Therefore, Eq. 2.20 predicts three distinct frequencies for the simultaneous scattering process. The first term describes Rayleigh scattering, which is observed at the frequency of incident radiation ν . The second and the third terms give rise to Stokes and anti-Stokes Raman scattering. It is noted that ν_v is the frequency that could be detected using IR absorption spectroscopy for the quantum

mechanical transition when $\Delta v = \pm 1$, if allowed by the molecular symmetry. The dependence of the second and third terms on $\partial\alpha/\partial Q$ explains the selection rules for molecular vibrations to be Raman active. Therefore, to yield Raman scattering at the Stokes and anti-Stokes frequencies, the molecular vibrations must change the polarizability, i.e., $\partial\alpha/\partial Q \neq 0$. Otherwise, the terms predicting Raman component frequencies in Eq. 2.20 have zero amplitude, which indicates no Raman scattering can be generated.

2.1.9 Raman cross-section and depolarization ratio

Like the molar absorptivity in the Beer-Lambert's law, Raman cross-section is defined to describe the efficiency of a Raman active scatterer. The integrated Raman cross-section, i.e., σ_j , is a number proportional to the probability of one incident photon being scattered as a Raman shifted photon at Raman shift wavenumber j . The importance of determining the empirical magnitude of Raman cross-section lies in the fact that experimenters have to deal with weaker scatters or with segregating scattering signals from multiple components. Besides its dependence on Raman shift band, the integrated Raman cross-section relies on parameters such as observation geometry, polarization, excitation wavelength, and the like (McCreery 2000). On the other hand, these parameters often remain unchanged and are pre-determined on a specific instrument. Therefore, estimates of single Raman cross-section for a specific laser wavelength and observation geometry help predict signal strength. The unit for cross-section is squared centimeters per molecule. Cross-section is related to the laser intensity and Raman scattering intensity in watts by:

$$I_R = I_0 \sigma_j D dz \quad (2.21)$$

where I_R (W) is the Raman intensity, I_0 (W) is the laser intensity, D is the number density of scatters (molecules per cubic centimeter), and dz is the penetration depth of laser into the sample (cm).

The integrated Raman cross-section implicates integration over two variables, i.e., over all directions, i.e., (4π steradians), around the sample and over all wavelengths within an entire Raman band. It is, however, impractical to collect scattering light over all directions around the sample and only a relatively small range of solid angle in the scattering directions from the sample is recorded. Therefore, the differential Raman cross-section β is often used instead of σ_j . Differential cross-section is given by:

$$\beta(\text{cm}^2\text{molecule}^{-1}\text{sr}^{-1}) = \frac{d\sigma_j}{d\Omega} \quad (2.22)$$

where sr refers to steradian and Ω refers to solid angle of observation (sr). Though, β are mostly reported instead of σ_j , it should be noted that β must be reported with a description of observation and excitation geometry. The reason is that β strongly depends on the angle of observation and polarization of excitation laser.

Assuming that the direction of propagation of the excitation laser light is the z -axis and plane perpendicular to the z -axis is the xy plane. Observation of Raman scattering happens in the xy plane. To calculate the depolarization ratio ρ , the intensity of Raman scattered light with its polarization direction perpendicular to the xy plane I_{\perp} and the intensity of scattered Raman light polarized parallel to the xy plane I_{\parallel} are required. The depolarization is given as:

$$\rho = \frac{I_{\perp}}{I_{\parallel}} \quad (2.23)$$

The molecular polarizability could be resolved into the isotropic and anisotropic parts. The value of ρ for a plane polarized excitation radiation, i.e., laser source, would fall between 0 and $\frac{3}{4}$. For an isotropic molecule, the value of ρ is zero for totally symmetric vibrations. It is also possible that the vibrating molecule is antisymmetric to any symmetry element of the molecule in its equilibrium configuration, and the value of ρ is $\frac{3}{4}$ in the Raman data. Therefore, the depolarization ratio ρ can be used to distinguish totally symmetrical vibrations from other kinds of vibrations.

2.2 Recent Progress Relevant to Proposed Research

The similarities and differences in the chemical information offered by the three branches of vibrational spectroscopy have led to comparative studies in various fields. The complementarity between inelastic scattering and absorption processes also intrigued some pioneering work to explore ways to utilize the synergy among different spectroscopic techniques. The availability of both Raman and NIR spectra generated by multimodal spectroscopic technique also provided a platform for two-dimensional correlation spectroscopy study. This section provided an updated overview of literatures pertaining to the aforementioned topics.

2.2.1 Comparative study on IR, NIR, and Raman spectroscopy

Schulz et al. (2003) explored the potential of vibrational spectroscopy techniques to identify and quantify substances in 11 different basil chemotypes. The NIR data of crushed samples were taken using rectangular cups and isolated basil oils were measured under transfection mode with quartz cuvettes. An ATR sample cell was used to carry the sample extract for FT-IR measurement. The ATR/FT-IR spectrometer operated within a spectral range of $650\text{-}3500\text{ cm}^{-1}$ with a spectral resolution of 2 cm^{-1} . Raman spectra were

collected using a NIR-FT Raman spectrometer with a Nd:YAG laser. Because different basil chemotypes possessed special chemical signature of main volatile components, all three vibrational spectroscopy methods were able to distinguish the 11 basil chemotypes. Due to the characteristic bands of individual volatiles present in IR and Raman spectra, discrimination of basil chemotypes was possible without subjecting IR and Raman data to full spectrum chemometric analysis. On the other hand, NIR data of various basil chemotypes could only be interpreted by coupling it with chemometrics algorithms. It was also demonstrated that NIR models calibrated using PLS algorithm were capable of quantitatively predicting the main valuable volatiles in the air-dried basil leaves. The authors finally recommended that NIR and Raman calibrations developed from essential oils could be used for on-line monitoring of the distillation process. Baranska et al. (2006) compared FT-Raman, ATR-IR, and NIR spectroscopy in determining the lycopene and β -carotene content in tomato fruits and products. Mashed tomato samples were presented to the sample stage for FT-Raman spectrometer, ATR cell of the IR instrument, and a spinning sample cup under the NIR fiber optic probe separately for spectra acquisition. The acquired spectral data were mean centered, and PLS models were subsequently calibrated. For qualitative analysis purpose, the results showed that FT-Raman spectroscopy could identify carotenoids in plant tissues and food products with no preliminary sample preparation. Water was a strong interferent for ATR/FT-IR and NIR measurements. In addition, no absorption bands characteristics for carotenoids are found in IR and NIR spectra. Interestingly, IR spectroscopy outperformed FT-Raman in quantifying the lycopene and β -carotene contents in tomato samples despite its lack of relevant spectral features. The quantitative prediction results using NIR spectroscopy

were the worst in their study with standard error of cross validation (SECV) of 74.34 and 0.34 for lycopene and β -carotene, respectively. Paradkar et al. (2009) examined the utility of FT-IR, FT-Raman, and NIR spectroscopy in detecting adulteration of maple syrup. A total of 54 samples of maple syrup adulterated with different quantities of corn syrup were prepared. The addition was expressed in percentage by the weight of maple syrup ranging from 0% to 27% in steps of 0.5%. The IR and Raman data were collected in the spectral range from 400 cm^{-1} to 4000 cm^{-1} . The NIR measurement was performed in the 600 nm -1700 nm region. It was reported that all three techniques were able to detect adulteration of corn syrup in the maple syrup. From quantitative analysis perspective, IR and Raman techniques were superior to NIR in predicting adulteration percentage. Qiao et al. (2008) compared three techniques, i.e., FT-IR, NIR, and Raman, for analyzing amino acid content in animal meals. Raw samples were presented to NIR and Raman instruments directly for scanning while finely ground samples were prepared for collecting FT-IR spectra. The authors found high level of noise in both Raman and IR data and determined that Raman technique was unsuitable for animal meal analysis. It was concluded that FT-IR and NIR are comparable in predicting amino acid content when both techniques use full spectral range. If the FT-IR wavenumber variables exhibiting high noise level were excluded from calibration, FT-IR offered improved calibration compared to NIR spectroscopy. The authors also noted that preparing samples for FT-IR was more time-consuming. Yang et al. (2005) conducted a qualitative analysis of edible oils and fats using FT-IR, FT-NIR and FT-Raman spectroscopy. Little sample preparation was needed for all three techniques. The authors employed linear discriminant analysis (LDA) and canonical variate analysis (CVA) to segregate different

oils based on their spectral features. It was found that FT-IR combined with chemometric analysis offered the highest successful classification rate (98%) followed by FT-Raman (94%) and FT-NIR (93%). The paper concluded that all three techniques were capable of rapidly discriminating edible oils and fats while FT-IR and Raman offer more exquisite structural information of the functional groups of oils than FT-NIR spectroscopy. Chung and Ku (2000) compared NIR, IR, and Raman for their utilization in analyzing American Petroleum Institute (API) gravity of atmospheric residue (AR) in crude oil. The authors found that Raman spectroscopy was unsuitable because of the strong fluorescence caused by asphaltenes. Although the employed FT-Raman system operated with a Nd:YAG excitation laser at 1064 nm and a radiation power of 5 mW, the fluorescence overwhelmed the Raman signal. Infrared spectra were acquired on a FT-IR spectrometer with an attenuated total reflection (ATR) probe. The poor reproducibility and high noise level made IR spectroscopy impractical for their quantitative application despite rich spectral features it provided. The authors claimed that NIR spectroscopy was most successful for determining the AR content in crude oil due to the high reproducibility of spectral data.

Gresham et al. (1999) compared NIR absorption spectroscopy versus Raman spectroscopy for industrial process monitoring purposes. In their experiment, two individual solid state instruments, i.e., one for NIR absorption and the other for Raman scattering spectroscopy, were used to simultaneously quantify components in xylene isomer mixtures. Their results demonstrated the utility of both NIR and Raman technique for quantitative analysis of ortho-, meta-, and para-xylene with superior absolute error at ± 0.05 , ± 0.12 , ± 0.09 (w/v) using NIR and ± 0.08 , ± 0.04 , ± 0.07 (w/v) using Raman. The

authors suggested that solid-state NIR and Raman spectrometers could be used for on-line raw material monitoring. The choice between two techniques was mainly determined by desired investment cost and actual monitoring environment. Ku and Chung (1999) compared the performance of NIR and FT-Raman in determining the chemical and physical properties of naphtha. The authors selected six different chemical compositions, i.e., total paraffin, total naphthene (cycloalkane), total aromatic, paraffin, benzene, and cyclopentane, and a physical parameter, i.e., specific gravity to compare NIR and FT-Raman methods. It was found that NIR calibration outperformed that of FT-Raman because of the high SNR and reproducibility of NIR data. However, the paper pointed out that Raman possessed great potential as an on-line quantitative analytical tool because of the richer chemical information it provided along with convenience in interfacing using fiber optics for remote sensing. Afseth et al. (2005) tested the capacity of NIR and Raman spectroscopy for determining fatty acid and main constituents in a complex food model system. The food model consisted of 70 different mixtures of protein, water, and oil blends to imitate fish and meat samples. For fatty acid determination, both NIR and Raman spectroscopy were able to predict content of saturated, monounsaturated, and polyunsaturated fatty acids in the samples with validation errors ranging from 2.4 to 6.1%. Raman spectroscopy provided best results for predicting the iodine values with a validation error of 2.8%. For predicting the main constituents, NIR technique demonstrated a substantial advantage over Raman spectroscopy.

2.2.2 Multimodal spectroscopy

When considering the right spectroscopic technique to choose, one important thing to note is the suitability of the three techniques for that application, which depends on

factors such as sample's physical state, measurement environment, and constituent of interest. Despite the strengths and weaknesses of the three vibrational spectroscopy techniques, all techniques are useful analytical tools depending on potential fields of applications. For on-line and on-site applications, NIR and Raman are more suitable than IR technique because of robust instrumentation and little need for sample preparation. De Beer et al. (2011) reviewed the use of NIR and Raman spectrometers as process analyzers for critical process monitoring and control. The authors specified that combination of complementary techniques for process control needed to be considered during implementation of process analyzers for real-time process measurements. The authors went through the fundamentals of NIR and Raman spectroscopy. It was pointed out that both spectroscopic techniques were suitable as process analyzers for their speed, non-destructiveness, and remote sensing capability. The complementary information provided by NIR and Raman techniques made them a natural alliance as process analyzers while also facing challenges in fields such as data handling. Recent application of NIR or Raman in critical processes such as blending, granulation, drying, and coating, etc. were thoroughly reviewed. In summary, both NIR and Raman spectroscopy will remain key players in the next step of realizing real-time monitoring of critical pharmaceutical production processes.

The complementary information provided by absorption and inelastic Raman scattering sparked an interest to combine the techniques to form a multimodal technique for improved performance. De Groot et al. (2003) conducted some pioneering work by combining Raman and IR data to quantitatively predict the property of yarn. In their work, the Raman spectra and IR spectra were concatenated to form new spectral data to be used

for partial least squares calibration. It was ruled out that such combination did not improve the performance of partial least squares regression (PLSR) model even with wavelength selection. The author therefore concluded that it was not necessary to combine complementary spectra for determination of the same compound. The same study also showed that Raman outperformed IR spectroscopy if no wavelength selection was carried out on IR spectra. In a different study, Anderssen et al. (2005) came up with research findings contradictory to that reported by De Groot et al. (2003) regarding combining spectral data from different spectroscopic methods. The LF-CAP method was proposed and used to jointly invert NIR and Raman spectra of wheat flours. Their experimental results verified that joint inversion of multi-modal data for effective information recovery improved model performance. The prediction models resulted from LF-CAP did not necessarily offer the lowest prediction error but its performance became more robust. There have also been some pioneering works on the fusion of NIR absorption and Raman technique (Clarke et al. 2001). In their paper, the authors emphasized the important roles that both spectroscopic imaging techniques played in pharmaceutical industry. In the meantime, the authors also pointed out that a single vibrational spectroscopy technique could not possibly identify all ingredients in the tablet formulation due to its wide range of materials. Therefore, combining two or more complementary techniques to visualize the entire pharmaceutical formulation and distribution of each component proves to be advantageous. However, for some practical considerations, e.g. poor transmittance performance of IR measurement, combination of IR with Raman spectroscopy was not an ideal method for pharmaceutical analysis. On the other hand, NIR and Raman spectroscopy possessed similar sampling requirements for

pharmaceutical samples. The two techniques were also complementary in detecting various components present in pharmaceutical formulations, such as inorganic dibasic calcium phosphate, carbohydrate species, and moisture content, etc. Despite the advantages for combining the two techniques, the authors stated that no instrument was available for simultaneous measurement of the NIR and Raman data. Therefore, their experiment was carried out on two separate instruments, i.e., a Raman and a FT-NIR microscopy system. Two batches of samples, A and B, were prepared. Sample A was used to evaluate chemical image fusion and sample B was used to solve a tablet sticking problem during production. To process the data, the authors adopted a method that used integrated area under unique peaks present in the second derivatives of the raw data to identify individual pharmaceutical components. In the first experiment, image results obtained from the fusion of Raman and NIR spectral images gave accurate mapping of individual pharmaceutical components. In the second experiment, the applicability of chemical image fusion in determining tablet sticking causes was demonstrated by providing proof to a newly proposed theory that is contrary to common belief. By combining image results from two spectroscopies using the most appropriate ingredient image, a complete visualization of pretableting blends was possible. The fused image suggested that the sticking was caused by the distribution of the inorganic binder within the blend rather than the distribution of lubricant, which was a common belief. In summary, synergy of NIR and Raman techniques removed ambiguity in collected image data and subsequently improved product quality through better understanding of manufacturing process.

Jérez Rozo et al. (2011) thoroughly discussed the complementary differences between NIR and Raman imaging in terms of spectral features, acquisition time, scale of scrutiny, and sampling volumes. A total of four hydroxypropyl methylcellulose (HPMC) film batches containing suspended griseofulvin particles were prepared. Both NIR and Raman mapping was performed on the polymeric film to study the distribution of griseofulvin particles. A Malvern SyNIRgi Near-Infrared Chemical Imaging System (Malvern, UK) was used for NIR mapping and Raman images were acquired using a LabRAM micro-Raman spectrometer (Horiba, Kyoto, Japan). The authors mentioned that such separate instrumental setup made the same region measurement impossible. Based on the highest intensity in the second derivative at 2080 nm, NIR binary images were formed, which identified the griseofulvin particle distribution. High intensity band at 1620 cm^{-1} , which is characteristic of griseofulvin, was used to predict distribution of both griseofulvin and HPMC within a sampling area of $100\text{ }\mu\text{m}^2$. According to the paper, the two methods complement each other with NIR mapping providing particle distribution within a few minutes and Raman imaging being used to monitor crystal form of griseofulvin and its size distribution.

With regards to instrumentation, there have been some previous efforts to construct instruments for multimodal spectroscopy measurements. Sostek and Wright (1998) proposed an innovative optical instrument for selective or simultaneous acquisition of IR or Raman data of the same sample. Prior to their invention, according to the patent, samples were measured on separate instruments and multiple types of spectroscopic data were required. With their invention, both microscopic IR and Raman data could be acquired on the same microscope system, which eliminated the need to

remove the sample once it was set up for analysis. In a more specific description, the optical apparatus included a sample stage, an IR source, a Raman excitation source, a Raman spectrometer, an IR detector, and an optical switch mechanism. The mechanism, which could be a mirror or a fixed dichroic beam splitter, was designed to selectively or simultaneously direct incident radiation, i.e., IR and Raman excitation light, and radiation signal to their respective detectors. The patent also included schematic diagrams and brief descriptions of several alternative embodiments of combined IR and Raman systems. Their design did not have a movable mirror or a fixed dichroic beam splitter in the common path to direct light. Rather, switching between IR and Raman mode was enabled by moving one type of detector unit out of the way while moving the other detector into the beam path. Though IR and Raman data could be acquired on the same microscope system, this design excluded the possibility to perform measurement of IR and Raman signal without system modification. The ability to perform IR and Raman measurement without hardware modification on other alternative designs was only possible when a suitable dichroic beam splitter instead of deflective mirror in the common path was placed to divert desired beam to the proper detector channel. In the patent, the authors mentioned a dichroic beam splitter capable of passing IR light and deflecting Raman signal but did not pursue the topic further.

Faupel et al. (2005) described in their patent an apparatus and method that combined more than one optical modality (spectroscopic method) to detect morphological and biochemical changes in tissues due to disease. The spectroscopic methods to be combined included but were not limited to absorption, fluorescence, reflectance, polarization anisotropy, and phase modulation. As the background of their

invention, the aforementioned spectroscopic methods were reviewed for their utilization in determining tissue characteristics. It was noted that absorption spectroscopy was more sensitive to morphological changes and fluorescence spectroscopy was more sensitive to biochemical changes. It was pointed out that none of the prior art technique alone was capable of accurately measuring tissue changes. The patent also claimed its unique advantage over prior spectroscopic designs by sampling a more generous number of points on the tissue. It was also stated that combined measurement using various methods could be taken simultaneously or one after another provided that critical timing window was controlled below a certain interval. However, simultaneous measurement on the same site of the tissue was preferred to ensure identical condition. The invention emphasized the importance of the design of optical fibers with predetermined patterns to make plurality of measurements on the target sample. The preferred embodiment consisted of a rotating core inside the endcap of optical fiber probe, which rotated between different measurement cycles. Technically speaking, the multi-modal spectroscopic system patented by Faupel et al. (2005) was a conceptual device without specific instrument designs, i.e., the entire system was composed of three main modules, which were a radiation source, a detector unit, and a processor. Although lots of details were provided on the design of an optical fiber probe to render more sampling area of target tissue, no specific information was available on how different spectroscopic techniques could be integrated in the three main modules.

In another patent, the requirement to combine two or more techniques to arrive at a more accurate ultimate determination was also confirmed by the findings of Samuels et

al. (1996), which used ratio of the fluorescence emission intensity to the scattering or reflected intensity of the lens to diagnose diabetic patients.

2.2.3 Protein analysis using NIR and Raman spectroscopy

Protein is a major component in food systems and essential for the functioning of living organisms. It is composed of amino acids connected by peptide bonds in a linear chain and folded into a globular form. There are 20 standard amino acids and most microorganisms and plants could biosynthesize them. On the other hand, human beings cannot synthesize all the amino acids and have to obtain some of the amino acids from diet. In food production, protein is an important additive to enhance the sensory properties of the food products.

In most applications of vibrational spectroscopy in food science and technology, NIR spectroscopy is frequently used to quantify crude protein, along with moisture, fat, and carbohydrate content in food systems. Due to its fingerprinting ability, Raman technique appears to be more suitable for elucidating the secondary structure of proteins and structural changes upon physical processing during heating or freezing. It leads to the assumption that the two complementary techniques may make a good team in investigating food protein content and its structural information. Based on this assumption, one of the research objectives aims to use the multimodal spectroscopy technique for quantify and qualitatively analyze aqueous samples of protein at different concentration levels. The protein samples may be processed by heating to alter their secondary or tertiary structure and detected using the proposed method. However, a review of literatures showed that NIR absorption spectroscopy might not prove to be

useful in protein analysis using the currently proposed multimodal spectroscopy system due to restricted wavelength range.

Ellepola et al. (2006) studied the conformation of rice globulin using Raman spectroscopy. The outcome of their experiment showed that α -helical structures predominated the native globulin conformation of rice. The protein structures were influenced by pH, ionic strength, and by some protein structure perturbants and heat treatments. It was also found that heat-induced aggregation of rice globulin was mainly influenced by disulfide bonds and hydrophobic interactions. Keller et al. (1993) reviewed the use of Raman spectroscopy for quality control of food. The paper went through the fundamentals of Raman theory and its instrumentation. The authors listed applications of NIR laser excited Raman spectroscopy in food analysis and categorized them according to main ingredients of food, i.e., lipids, peptides and protein, carbohydrates, sensory qualities and minor components. The assignment of Raman bands for amide groups and amino acids in protein was listed in the article.

As a more established analysis tool in food and agriculture, the use of NIR spectroscopy for analyzing protein content in food is widely accepted. It is one of the official methods for protein analysis in grain and its products approved by the American association of cereal chemists (AACC) (method 39-11.01, method 39-25.01, AACC, 2009). The NIR technique was used to analyze protein in semi-frozen beef on-line (Tøgersen et al. 2003), to detect animal protein by-products in farm feed (Pérez-Marín et al. 2009), to determine polymeric protein in wheat (Scholz et al. 2007), and to control the flour protein content (Osborne 1986). Therefore, it appeared feasible to apply both Raman spectroscopy and NIR spectroscopy for protein analysis. However, a careful

review of the use of near-infrared spectroscopy (NIRS) for protein analysis revealed that best wavelength region for protein determination was above 2000 nm. Previous studies utilizing NIRS only indicated the use of the combination and overtone bands of protein for its analysis. The spectral region from 2000 nm to 2500 nm corresponds to the combinational vibrations of C–H, O–H, or N–H stretching with bending vibrations of CH₂, N–H, or O–H (Shaw et al. 1996). Hall and Pollard (1993) noted that the spectral band around 2178 nm was especially useful for quantitative analysis of albumin. D áz-Carrillo et al. (1993) found the spectral bands around 2059 nm, 2170 nm, and 2300 nm correlated strongly with protein content in goat's milk because they were also found in the NIR spectra of purified albumin. The overtone bands for C–H, O–H, or N–H stretching are within 1400 nm and 2000 nm and are generally weaker than the combination bands. Spectral bands around 1560–1670 nm were usually assigned to characterize protein content in food (Chang 2010). Table 2.1 provides a more detailed correlation between molecular structure and absorption features specific to ovalbumin protein in the NIR region (Workman and Weyer 2008). Therefore, should conducting quantitative protein analysis fail on the proposed system (working spectral range from 752 to 988 nm), the result would not be unexpected.

The functionality of food protein is attributed to many physiochemical and structural properties of protein molecules, e.g., hydrophobicity, net charge, size, and secondary and tertiary structure. The primary molecular structure of protein, i.e., the amino acid sequence, ultimately determined the functional behavior of proteins (Culbertson 2006). However, the amino acid sequence only describes the polypeptide chain in two-dimensional space. The three-dimensional structure of protein molecules

exist on four levels of organization, i.e., primary, secondary, tertiary, and quaternary. The local conformation of the polypeptide chain, i.e., secondary structure, is determined by its primary structure. The basic motifs for secondary structure are designated as α -helix, β -sheets, and turns (random coils). Tertiary structure is based on secondary structure and gives a more complete picture of the entire polypeptide chain. Nevertheless, all levels of structures of protein are very important to the functionality of protein. Protein denaturation refers to the change in the tertiary fold structure and simultaneous unfolding of secondary structure. During food processing, denaturation induced by heat, pressure, and additives is unavoidable. Understanding how a protein behaves during food preparation and processing helps us manipulate its physiochemical properties in order to improve food quality. Vibrational spectroscopy is a great tool for determination of protein content in food, studying its molecular structural information, and evaluating the conformational changes (Jalkanen 2005; Jung 2000; Wen 2007). Evidence of structural changes in protein molecules could be sought in the intensity variations and frequency shifts of amide bands. Both Raman and NIR spectroscopy were suitable to study the conformational changes of protein during thermal denaturation. The NIR absorption bands associated with the secondary structures of protein molecules are mainly found at several discrete wavelengths above 2056 nm (Bruun et al. 2007a). Therefore, again, the proposed system might not be able to perform protein conformational analysis due to limited spectral detection range. The commonly used frequencies (cm^{-1}) for studying thermal denaturation of protein are listed in Table 2.2. Painter and Koenig (1976) noted that spectral region assigned to amide I ($1650\text{-}1670\text{ cm}^{-1}$) was less useful than the amide III mode in probing the protein secondary structures. The reason is that the amide I mode

was not well resolved into discrete spectral lines that could be assigned to different types of secondary structures.

Table 2.1 Spectral absorption and chemical structure correlations

Wavelength (nm)	Chemical structure assignment
1550	First overtone of an N–H stretching mode of the amide groups
2083	Combination band of the N–H bending and C–N stretching
2174	CONH ₂ as combination of amide B and amide II modes
2345	Combination band of first overtone of CH ₂ symmetric stretching and fundamental bending of CH ₂

Table 2.2 Common vibrational modes used in protein structure study

Spectroscopy	Wavenumbers related to secondary structure (cm ⁻¹)			Reference
	α -helix	β -sheets	Random coils	
Raman	1655 \pm 5, >1275	1670 \pm 3, 1235 \pm 5	1665 \pm 3, 1685, 1245 \pm 4	Li-Chan (1996)
NIR	4604, 4369	4535, 4416, 4323	4415	Bruun et al. (2007b)

Egg white or albumen is an important ingredient for many foods, e.g., bakery products (e.g., cakes and creams), cookies, and meat products (e.g., sausages). The functionalities of egg white protein include foaming, emulsifying, texturizing, heat setting, dough forming, water and flavor binding, etc. In its natural form, egg white is a

colloidal aqueous solution of different proteins. Its homogeneity, combination of functionalities, together with its ready availability makes egg white a most sought after ingredient in food preparation. The egg white proteins have been extensively studied by food scientists and biochemists. Understanding the physiochemical properties helps elucidate the relation between functionality and molecular structure, which is important for achieving the desirable sensory attributes of food products. The main constituents of egg white are 84-89% (w/w) water, 9.7-10.6% protein, 0.5-0.6% carbohydrate, and negligible amount of lipids. The characteristics of major egg white proteins are listed in Table 2.3.

Table 2.3 Physiochemical properties of major egg white proteins

Protein	Amount (% dry mass)	Molecular weight (kDa)	Denaturation temperature (°C)
Ovalbumin	54	44.5	84.0
Ovotransferrin	12.0	77.7	61.0
Ovomucoid	11.0	28.0	77.0
Ovomucin	3.5	5.5-8.3×10 ³	ND
Lysozyme	3.4	14.3	75.0
G2 globulin	4.0	49.0	92.5
G3 globulin	4.0	49.0	ND
Avidin	0.05	68.3	ND

ND - Not determined.

There are about 15 different types of proteins in egg white and the protein system can be described as an aqueous solution of ovomucin fibers and globular proteins (Li-Chan and Nakai 1989). Ovalbumin, a most studied protein, consists of more than half of the total proteins in egg white by weight. The purified ovalbumin is composed of three subclasses A1 (85%), A2 (12%), and A3 (3%). The proportions of ovalbumin with different secondary structures were reported to be 49% α -helix, 13% β -sheets, and 24% random coil respectively (Doi et al. 1987). Upon thermal denaturation, conformation changes and coagulation of egg white were reported using Raman spectroscopy (Painter and Koenig 1976). The results obtained using Raman spectroscopy suggested that the formation of stable β -sheets structures explained the conformation changes during thermal denaturation. A sharp and intense spectral line pertaining to amid III appears at 1236 cm^{-1} after heating the sample at $70\text{ }^{\circ}\text{C}$ for 10 mins. Also, the amid I peak at 1667 cm^{-1} shifted to 1672 cm^{-1} as a result of thermal denaturation. Similar results were obtained for 10% (w/w) ovalbumin solution with an amide III line emerging at 1239 cm^{-1} and amide I line shifted to 1672 cm^{-1} . The second most abundant protein, i.e., Ovotransferrin (conalbumin), was a weak Raman scatterer and therefore was not discussed.

2.2.4 Analysis of ethanol in aqueous solutions

Ethanol, also called pure alcohol or grain ethanol, is an important ingredient of alcoholic beverages (Wang et al. 2003). The chemical formula for ethanol is $\text{C}_2\text{H}_5\text{OH}$, though sometimes it is written as $\text{CH}_3\text{-CH}_2\text{-OH}$ to indicate its molecular structure. The average concentration range of ethanol in different types of alcoholic beverages are 7~21% (v/v) for wine, 20~50% (v/v) for liqueur, and 3~6% (v/v) for beer, respectively. Ethanol concentration is an important quality indicator and taxation factor for alcoholic beverages.

In biomedical field, ethanol is also a common substance of abuse found in human blood (Qiu and Shao 2001). As an alternative motor fuel, ethanol-blended gasoline is deemed a good substitute for its ability to reduce air contaminants and its role in water quality monitoring (Niven 2005).

Finding a fast, accurate, reagentless method to quantitatively determine ethanol's concentration certainly has its own significance. Both Raman and NIR spectroscopy have been used to determine ethanol concentration in various solutions. Nordon et al. (2005) utilized both NIR spectroscopy and Raman spectroscopy to assess the ethanol concentration inside glass bottles. The authors managed to obtain 32 spirit samples with ethanol concentration range of 19.9–63.3% (v/v). The prediction precision (average relative standard deviation (RSD)) of calibration models using NIR and Raman spectroscopy were found to be 0.4 and 0.5% respectively. The accuracy of NIR spectroscopy was slightly better than that of Raman technique. However, calibration built using Raman spectroscopy was easier to maintain because it is univariate. Cavinato et al. (1990) demonstrated that short-wavelength NIR spectroscopy could be used to quantitatively determine the ethanol concentrations during fermentation processes. They reported a standard error of prediction (SEP) of 0.2% (w/w) of ethanol in binary mixtures of ethanol and water with a concentration range of 0-15%. For water/yeast/ethanol solutions, the SEP value for succeeding fermentation processes was 0.42%. Both FT-NIR and FT Raman spectroscopy were applied to determine the ethanol content in ethanol fuel and beverages (Mendes et al. 2003). For alcoholic beverage analysis, twenty four samples were prepared with ethanol content ranging from 0 to 100% (v/v). Using independent test samples, SEP values of 0.324% and 0.638 were reported for

NIR and Raman PLS calibration models. Some real alcoholic beverages, e.g., beer, brandy, and vodka, were tested for their ethanol content. In terms of relative error, it was found that Raman spectroscopy performed better than NIR spectroscopy with an accuracy comparable to that of standard American Society for Testing and Materials (ASTM) D4052 method. Khijwania et al. (2007) reported a Raman sensor system coupled with fiber optic probe for ethanol and methanol detection in water and fuel. Binary solutions of aqueous solution of ethanol were prepared with concentrations at 0%, 20%, 40%, 60%, 80%, and 100% by volume. Results showed that the sensor was capable of distinguishing ethanol solutions of different concentration based on intensity of hydrocarbon peak. A linear relationship was also observed between integrated peak areas and hydrocarbon concentrations.

2.2.5 Two-dimensional correlation spectroscopy

The concept of two-dimensional (2D) correlation spectroscopy was first proposed by Isao Noda (Noda 1986). Later, the original idea of two-dimensional correlation spectroscopy was extended to a new form of 2D spectroscopy called generalized 2D correlation spectroscopy (Noda 1993). So far, the generalized 2D correlation spectroscopy has been successfully applied to a wide range of vibrational spectroscopic techniques such as IR, NIR, Raman, ultraviolet-visible (UV-Vis), and fluorescence spectroscopy (Jung et al. 2002). Generalized 2D correlation spectroscopy gained considerable popularity mainly because it enabled the analysis of spectral variations under external perturbations such as concentration, pressure, temperature, and time (Liu et al. 2000).

The 2D correlation spectra are composed of two orthogonal components, i.e., synchronous and asynchronous spectra. Synchronous spectra provide information

representing the in-phase changes, i.e., simultaneous changes, of the two separate spectral variations at two independent variables under external perturbation. Asynchronous 2D correlation spectra describe out-of-phase spectral variations, i.e., sequential changes, measured at separate variables (Šašić and Ozaki 2001). Heterospectral correlation analysis investigates two completely different types of spectra obtained under the same external perturbation. In general, heterospectral correlation can be categorized into two types of analysis. One type of 2D heterospectral correlation compares closely related spectroscopic techniques, e.g., Raman/IR, NIR/IR, NIR/Raman. One purpose of this type of analysis is to assign highly overlapping spectral bands using better understood spectral features obtained from another spectroscopic probe. The other type of heterospectral correlation analyzes data from widely different types of physical techniques, e.g., Raman and X-ray scattering. This type of analysis helps elucidate physical properties of samples under perturbation applied to the system (Noda and Ozaki, 2004).

Matsushita et al. (2000) measured the FT-Raman and NIR spectra of poly-methyl-methacrylate (PMMA) and its immiscible blends with atactic polystyrene (PS) with PS contents from 1 to 10 % (w/w). Synchronous spectra derived from the 2D heterospectral correlation analysis of Raman and NIR spectra identified the NIR spectra bands of PS, which were otherwise indiscernible from the one-dimensional NIR spectra. McClure et al. (1996) studied the 2D correlation of NIR and Raman spectra for protein (lysozyme) and sugar (sucrose) mixture. The authors concluded that 2D heterospectral analysis helped identify NIR bands as a result of contribution of a component in the mixture by examining correlation slice of a Raman spectrum. The 2D correlation spectroscopic technique also made it possible to predict important wavenumber variables for calibration

using chemometrics. The paper also pointed out a drawback of the 2D correlation approach, viz., the difficulty to distinguish correlations between spectral bands originating from the same functional groups.

2.3 Summary

The chapter briefly reviewed the fundamentals of molecular vibrations and introduced basic concepts such as anharmonicity, overtones, polarization, and Raman cross-section as related to vibrational spectroscopy. The theoretical basis for the difference and complementarity between IR absorption and inelastic scattering were explained. An updated overview of research activities on the application of NIR, IR, and Raman spectroscopy in various industries was provided. It was shown that both NIR and Raman were promising techniques for routine analysis. Progresses in multimodal spectroscopy were discussed. Reported research findings focused on data fusion methods and data mining of different types of spectra. Based on previous findings, the question whether there existed advantages in combining NIR and Raman data to enhance the interpretation of chemical composition and structure was not conclusive. Approaches to fully utilize the synergy between NIR and Raman needed to be further explored. Several patents attacked the problem by proposing novel hardware designs that could accommodate multimodal spectroscopic measurements. These patents provided some general ideas for future work. Protein and ethyl alcohol are important food/beverage ingredients. The potential applications of NIR and Raman in protein and ethyl alcohol analysis were reviewed as they related to the experiments planned in this thesis. Finally, a brief review of 2D correlation spectroscopy was included to explain the fundamentals behind 2D heterospectral correlation NIR/Raman spectroscopy.

3. MATERIALS and METHODS

3.1 The Instrument

A schematic diagram of the design of the original multi-modal spectroscopic device is provided in Fig. 3.1. The prototype system was constructed on a sealed hole tabletop with tuned damping (model RS2000, Newport, Irvine, CA). An electron multiplication CCD (EMCCD) camera (model DV865A, Andor Technology plc., Belfast, Northern Ireland) was purchased as the detector unit. The CCD camera was connected to a personal computer (CPU 450MHz, RAM 128Mb) via a 14-bit peripheral component interconnect (PCI) interface card (model CCI-20, Andor Technology plc., Belfast, Northern Ireland). The camera sensor chip came in a format of 576×288 pixels and was exposed to radiation signal through a c-mount opening. A compatible Raman spectrograph (model ImSpector V10R, Spectral Imaging Ltd., Oulu, Finland) was attached in front of the Andor camera using a C-mount locking ring. Before tightening the locking screws, adjustment of the ImSpector spectrograph to properly align with the CCD camera sensor was carried out. A 3-axis (XYZ) linear translation manual micrometer stage was mounted in front of the entrance slit of the Raman spectrograph for fine alignment between fiber optics and the entrance slit. For the NIRS channel, a custom-made bifurcated fiber optic bundle was used. The excitation branch was coupled to the NIR light source, which was a halogen tungsten lamp (model HL-2000-LL, Ocean optics, Dunedin, FL). Near-infrared light exiting from the probe head was transflected back to the receiving optic fibers. The optic fibers at the receiving end of the bifurcated fiber optic bundle were arranged in a linear array for easy coupling into the entrance slit. A custom-made multimode optic fiber with a core diameter of $400 \mu\text{m}$ was used to couple the Raman signal into the spectrograph. An

external cavity diode laser (ECDL) (model PI-ECL-785-300-FS, Process Instruments Inc., Salt Lake City, UT) with an emission wavelength of 785 nm was used as the excitation light source for the Raman channel. The NIR laser was coupled into a 200 μm diameter optical fiber (model P200-2-UV-VIS, Ocean Optics, Dunedin, FL) via a subminiature version A (SMA) connector and delivered to the Raman probe that was built in-house. Sampling cell for Raman was a 3.5 mL quartz cuvette (catalogue number Z276901, Sigma Aldrich, St. Louis, MO). A glass cuvette with a working spectral range of 334-2500 nm (catalogue number Z276863, Sigma Aldrich, St. Louis, MO) was used to contain samples for NIR absorption measurement.

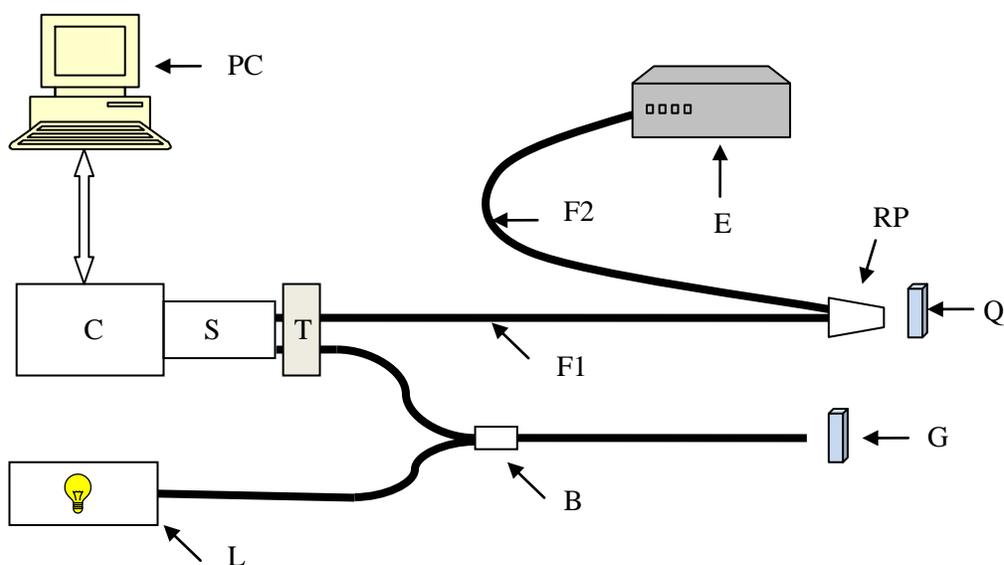


Fig. 3.1 Schematic layout of the multimodal spectroscopic system. B: bifurcated fiber optic bundle; C: EMCCD camera; E: Raman laser module; F1: 400 μm core diameter fiber optic; F2: 200 μm core diameter fiber optic; G: glass cuvette; L: halogen tungsten lamp; PC: personal computer; Q: quartz cuvette; RP: Raman probe; S: spectrograph; T: XYZ translation stage.

3.1.1 EMCCD camera and control software

The EMCCD technology was invented as an alternative to the existing intensified CCD (ICCD). Due to the introduction of EMCCD device, the low light level imaging has been revolutionized in recent years. The architecture of an EMCCD is different from that of a conventional CCD in an extended section of gain register. The solid state EM register was appended in between the normal serial register and the final detection nodes (Jerram et al. 2001). A concept, viz., impact ionization, in silicon was utilized to provide gain in the charge domain of an EMCCD (Madan et al. 1983). This was achieved by operating the gain register at high clock voltage to encourage avalanche multiplication. As the electron packets were passed along the elements in the gain register, a small number of additional electrons were generated after each transfer. This small amount of additional electrons created a statistical gain, i.e., mean gain per stage (A), which is usually in the range of 0.01. However, transferring the electron packets along elements in the serial register, the total gain could be huge (Eq. 3.1). Total gain adjustment was achieved by finely controlling the high amplitude clock pulse.

$$G = (1 + A)^n \quad (3.1)$$

where G is the total gain, A is the mean gain per stage, and n is the total number of elements in the serial register.

One obvious advantage of EMCCD over ICCD is its relatively smaller packaging. Because EMCCD does not need an intensifier to be placed in front of the CCD sensor, higher quantum efficiency (QE) is achieved with EMCCD. The peak QE of a front illuminated EMCCD is around 45%. If cost is not a factor, a possible QE of 95% could be achieved using back-thinned CCD with back-illumination (McWhirter 2008).

In this project, the detector array was a front illuminated type CCD sensor with a pixel size of $20 \times 30 \mu\text{m}$. The data acquisition board had 14-bit precision and was able to read out data at 200 ns per pixel. The control software was the Andor SOLIS for spectroscopy (version 4.3.0.0, Andor Technology plc., Belfast, Northern Ireland). Detector linearity was better than 1% and non-uniformity among pixels was below 0.01%. The base mean level, i.e., average DC offset for single pixel readout with minimum exposure time in the darkness with sensor at $-20 \text{ }^\circ\text{C}$, was 862 electrons with each pixel saturating at 193738 electrons. Under software setting, the camera was supposed to drive the sensor temperature down to $-70 \text{ }^\circ\text{C}$ using its thermoelectric cooler. However, the minimum achievable sensor temperature was $-58 \text{ }^\circ\text{C}$ with air cooling. Therefore, considering the minimum achievable temperature ($-58 \text{ }^\circ\text{C}$), the camera was set to cool to $-50 \text{ }^\circ\text{C}$ for all experiments in this research project. As indicated by the temperature reading in the software interface, the sensor temperature could be cooled down to $-50 \text{ }^\circ\text{C}$ and stabilized during measurements. The CCD head started to cool down as soon as the control software started to run. No data collection was performed before sensor temperature stabilized, which usually took 40~60 mins. There were three acquisition modes available, viz., single, accumulate, and kinetic. Accumulate mode was chosen to average out random noises and minimize effect of cosmic ray spikes. The acquired data could be read out either as images or multi-track signals. Image readout mode was chosen to prevent possible loss of data due to factors such as blurring and misalignment of optics. Because Raman signal was weak, the EM function of the camera was turned on for Raman measurement. This type of on-chip amplification allowed high sensitivity with high readout speed, and its performance was limited only by shot noise and spurious

charge. Choosing an appropriate EM setting could fully utilize the available dynamic range and improve the signal SNR. In the Raman experiments, an EM value of 150 was chosen. No EM function was applied for NIR measurement because of the relatively strong NIR signal as compared to Raman. The camera system was also capable of performing hardware binning of image data to improve SNR. Binning function was not used in this research because it resulted in loss of spectral resolution. The assignment of the sensor area is illustrated in Fig. 3.2.

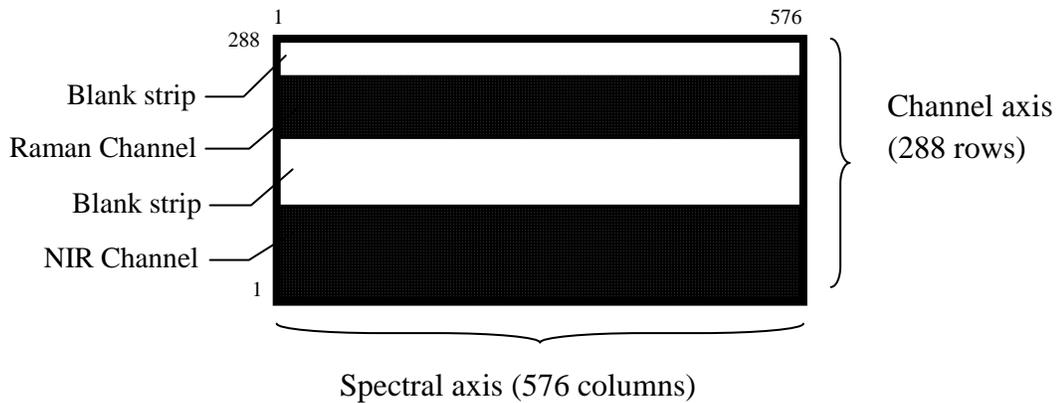


Fig. 3.2 Assignment of the EMCCD sensor array

The starting pixel is on the lower left corner of the imaging area. Because Raman peaks were sharp, the best achievable spectral resolution was preferred. Therefore, the horizontal dimension corresponding to 576 columns of pixels was chosen as the spectral axis. The vertical dimension (288 rows) was assigned as the channel axis. Possible ways to assign camera pixel rows to a channel was limited by the physical properties of optical components, e.g., core cladding diameter and entrance slit length. From the starting row 1 till around row 70, some of the rows were designated as the NIR channel. The optics of the custom made fiber optic probe made it impossible to use all the rows. This is discussed later in section 3.1.3. The row cluster corresponding to the Raman channel

started approximately from row 221 and continued till row 251. Rows in between the two channels and beyond the Raman channels were left as blank strips to prevent cross-talk.

3.1.2 The spectrograph

The key component inside the ImSpector spectrograph is believed to be a prism-grating-prism (PGP) component (Fisher et al. 1998). The PGP component consists of a volume transmission plane grating cemented between two symmetrical or closely symmetrical prisms. Spectral filters are often integrated into the PGP component to cut off radiation in unwanted wavelength range. Aperture stop is also an integral part of the PGP. The incident angle to the front prism is calculated as Bragg angle according to a certain wavelength within the spectral range. The purpose for achieving Bragg condition is to ensure maximum diffracted power.

$$\sin(\theta_B) = -\frac{\lambda_B}{2d} \quad (3.2)$$

where θ_B is the Bragg angle, λ_B is the Bragg wavelength, and d is the grating constant.

Inside the spectrograph, two specially design triplet lenses were installed both in front of the PGP component and after the PGP along the optical axis. The front lens collimates the incident light through the entrance slit and the rear lens focuses light onto the image sensing area of the CCD. The lenses are corrected for chromatic aberration and coated to minimize reflection loss (Herrala and Okkonen 1996).

The ImSpector spectrograph offered a nominal spectral range of 750-1000 nm. The actual spectral range according to the manufacturer's test report was 743-987 nm. At three wavelengths that were tested using a Mercury Argon (HgAr) calibration lamp, the spectral resolution expressed as full width at half maximum (FWHM) was below 0.76 nm.

The entrance slit had a width of 30 μm . The f-number of the spectrograph was 2.4, which corresponded to a numerical aperture (NA) of 0.2. The straylight level was below 1%.

3.1.3 Fiber optics

The bifurcated fiber optic bundle was custom designed in consultation with FiberTech Optica (Kitchner, ON) and built by the same company. The layout of the three ends of the fiber optic bundle is shown in Fig. 3.3.

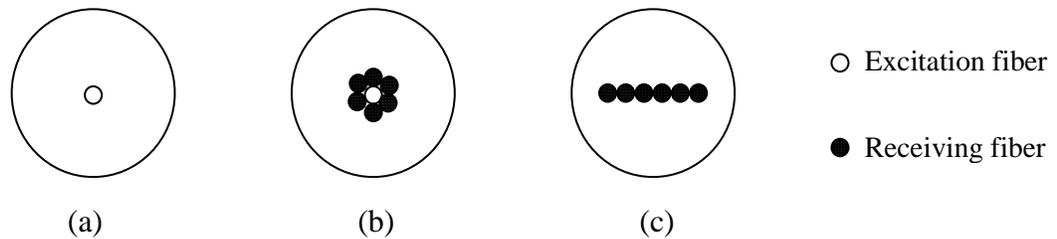


Fig. 3.3 Schematic layout of the three ends of the bifurcated fiber optic bundle. (a) source end (to light source); (b) probe end (to sample); (c) delivery end (to entrance slit of the spectrograph).

The source end was enclosed in a standard SMA connector for easy coupling to the halogen tungsten lamp (Fig. 3.3(a)). The probe end adopted a 6-around-1 design, i.e., six receiving fibers around one excitation fiber (Fig. 3.3(b)). The optic fibers at the signal delivery end were arranged in a linear array for easy coupling to the spectrograph entrance slit (Fig. 3.3(c)). Both the probe and the delivery end were enclosed in 50 mm long brass ferrule with a diameter of 6.35 mm. Both the pure silica core and fluorine doped silica clad employed ultra low hydroxide (OH) material. The diameter of the silica core was 440 μm . The numerical aperture (NA) of the fiber optic bundle was 0.22 and the bundle had a working spectral range of 350 to 2400 nm. The fiber optic bundle was

sheathed in black (Polyvinyl chloride) monocoil tubing. The overall length of the bifurcated fiber probe was 1 m and was split halfway. The delivery fiber optic bundle used in the optical path of Raman channel (F1 in Fig. 3.1) was provided by the same company and shared the same parameters as the bifurcated fiber probe. The use of low OH material as the fiber core helped lower the fluorescence background due to silicon in the presence of strong laser excitation. The source fiber used in the Raman channel (F2 in Fig. 3.1) was a 200 μm core diameter fiber for use in the visible to NIR region. This Raman source fiber allows for easy coupling to the laser light source as well as to the Raman sampling probe assembly.

3.1.4 Raman sampling probe

The optical setup of the in-house built Raman probe is illustrated in Fig. 3.4. In brief, the laser light came through the source fiber and was collimated using a lens. The collimated laser light then passed through a band-pass filter in order to remove unwanted side modes and fiber fluorescence. The filtered light was then reflected by the notch filter into the back of a $\times 10$ Nikon microscope objective. The sample cell was placed in front of the objective with the laser light focused at the centre of the sample cell cross-section. Sequentially, backscattered light was collected by the microscope objective before it reached the notch filter. At a tilting angle of 10° , the notch filter partially rejected the oncoming Rayleigh scattering from the sample cell. The filtered backscattered light was then collected by a collimator in front of the delivery fiber optic bundle.

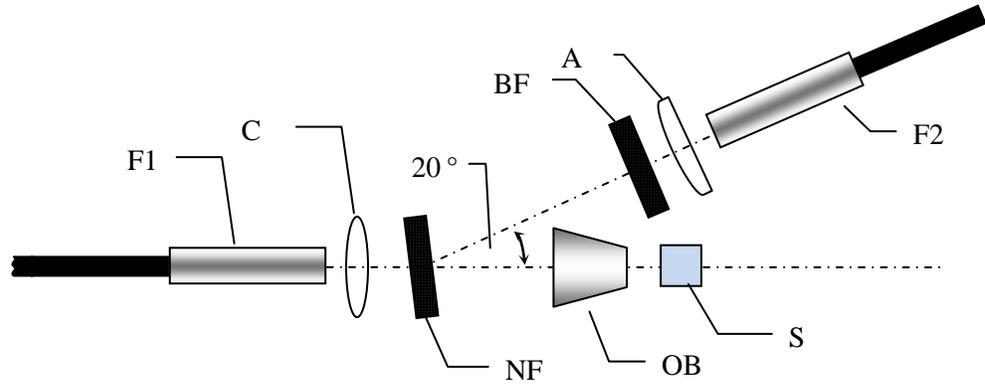


Fig. 3.4 Schematic layout of the Raman sampling probe. BF: bandpass filter; A: collimator for source fiber; C: collimator for collection fiber; F1: collection fiber; F2: source fiber; NF: notch filter; OB: microscope objective; S: sample cell.

The collimator (A) was a positive NIR achromat (part number AC080-010-B, Thorlabs, Newton, NJ). The first air-to-glass interface facing the point source, i.e., NIR laser emanating from the fiber optic, had a greater radius of curvature for optimal collimating performance. The achromat lens had a physical diameter of 8.0 mm and a focal length of 10 mm. The design spectral range was 650-1050 nm. Both the bandpass filter (BP) and notch filter (NF) had the same physical parameters and were provided by the same manufacturer (Iridian Spectral Technologies, Ottawa, ON). Both filters had an outer diameter of 12.5 mm, a clear aperture of 8 mm and a thickness of 3 mm. For the bandpass filter, the blocking wavelength ranges with an optical density (OD) of 5 were 706-781 nm and 789-981 nm. The FWHM was 2~3 nm with a maximum transmission efficiency of over 90% within the pass band. The notch filter had an OD of up to 6 within a bandwidth of 39 nm around the central wavelength 785 nm. The average transmission efficiency was 90% outside the notch band. All spectral specifications remained valid with an incident angle of 0°. Tilting of the filters with regards to the optical axis resulted in degraded

blocking performance. The microscope objective was a $\times 10$ Nikon imaging objective with a NA of 0.3. The collimator (C) in front of the collection fiber had a NA of 0.25 and a clear aperture of 10 mm.

3.1.5 The light sources

The HL-2000-LL tungsten halogen lamp had a nominal output power of 7 W. It provided continuous spectral radiation from 360 nm to 2000 nm. The lamp unit required 5 minutes to stabilize its output and offered an output stability of 0.5% while working continuously. The light bulb had an expected life of 10,000 h working under room temperature. The diode laser module generated Raman excitation at a wavelength of 785 nm with an output power of 300 mW. The output was a multimode (spatial), vertically polarized and collimated laser beam. The linewidth was 1.5 cm^{-1} (0.06 nm). The frequency and intensity stability during one week of operation were 0.2 cm^{-1} and 0.3% respectively. Emission wavelength was stabilized using an external cavity constructed by a diffraction grating. Wang et al. (2012) provided a detailed discussion on the principles and designs of ECDL device for Raman spectroscopy.

3.1.6 Testing equipments and miscellaneous parts

A hand-held optical power meter (Model 1918-C, Newport, Irvine, CA) coupled with a high-sensitivity thermopile sensor (Model 818P-001-12, Newport, Irvine, CA) was used for laser and optical design testing at the initial construction stage of the prototype device. An Argon calibration lamp (model 6030, Newport, Irvine, CA) was used for wavelength calibration and validation. After the system was set up and tested, a compact handheld laser power meter (stock No. NT54-018, Edmund Optics, Barrington, NJ) was used for routine output power check at the sample cell location. The purpose was to record the

laser power at the sample as experimental condition. A composite grating spectrometer working in the 200-1100 nm range (model HR4000CG-UV-NIR, Ocean Optics, Dunedin, FL) was used for spectra corroboration. Two XYZ micrometer translation stages were used for fine adjustment. The translation stage offers travel distance of 13 mm along each axis with 0.01 mm adjustment step. One translation stage is mounted in front of the spectrograph for alignment of fiber optic bundles with the entrance slit of spectrograph. The other stage is used for fine adjustment of location of the Raman sample cell with respect to the focal point of the laser beam. All other optical mechanical components, e.g., optical posts, post holders, table clamps, fiber chuck, NIR viewing cards, etc., were acquired from Thorlabs (Thorlabs, Newton, NJ).

3.1.7 System calibration and validation

After initial assembly and alignment of the optics of the spectroscopic system, hardware calibration and testing was conducted. The prototype device was tested for parameters such as spectral range, spectral resolution, linearity, and repeatability. Specific testing procedures for each parameter are described here.

Spectral range and wavelength (wavenumber) calibration: In order to calibrate the working spectral range of the spectrograph, NIR lasers with known emission wavelengths and a wavelength standard material were needed. Because the wavenumber versus pixel position relationship is not linear, two-point calibration is insufficient. In this experiment, the spectra of a 785 nm and an 830 nm laser source were first taken to roughly establish the corresponding relation between wavelength (wavenumber for Raman) and the camera pixel count. The spectra of standard wavelength materials, e.g., benzonitrile and 4-Acetamidophenol, were then taken. Based on the published standard Raman spectra of

the testing materials, assignment of wavenumbers to individual pixels was carried out one-by-one. Based on the assignment, a polynomial fit equation was developed to establish a more accurate wavenumber calibration of camera pixel count.

Spectral resolution: To test the spectral resolution, spectra of monochromatic source needs to be taken. Spectral emission lines available from the Argon calibration lamp are considered monochromatic. To estimate the spectral resolution of the prototype device, the lamp spectrum was used to calculate the FWHM of the instrument profile. The lamp spectral output was taken and then the spectrum was fitted into smoothed Gaussian curves centered at individual spectral peaks, and the half width could then be calculated.

Repeatability: During spectral measurement, the output signal of a CCD detector fluctuates. The temporal changes of signal could be an important factor in deciding the success of quantitative analysis using a single-beam technique. The signal fluctuation was caused by multiple instrumental factors such as intensity variation of light source, camera spectral response, slight changes in sampling geometry, etc. Wavelength of Raman shifts could also drift depending on the stability of excitation laser. As previously mentioned, the laser excitation used for this project has a superior wavelength stability of $0.2 \text{ cm}^{-1} / \text{week}$. Therefore, the intensity of the overall system was more of a concern for both Raman and NIR channel. To test the repeatability, sample spectra were acquired. The percent standard deviation, i.e., standard error divided by the mean, was used to estimate the overall variability of the signal.

Spectral response linearity: The linearity of the spectral camera was tested using a series of diffuse reflectance standards (Spectralon, Labsphere, North Sutton, NH). The

standards were available at four certified reflectance levels, i.e., 2%, 50%, 75%, and 99%.

The standards are spectrally flat within the visible and NIR region.

3.2 Samples for Analysis

Chemicals and sample preparation are described in details in this section. Experimental design method for quantitative analysis is also provided.

3.2.1 Chemicals

The following chemicals (Table 3.1) were purchased from Sigma Aldrich (St. Louis, MO) and used in the experiments.

Table 3.1 Chemicals purchased from Sigma Aldrich and their purposes

Description	Catalogue Number	Purpose
4-Acetamidophenol, 98%	A7302-100G-A	Raman wavenumber standard (solid)
Benzonitrile, Reagentplus, 99%	B8959-100ML	Raman wavenumber standard (liquid)
Melting point standard naphthalene	84679-250MG	Raman wavenumber standard (solid)
Albumin chicken egg grade II	A5253-500G	Qualitative and quantitative analysis

The grade II albumin powder is crude dried chicken white containing 62-88% percent egg albumin extracted using agarose gel electrophoresis. Food grade pure ethanol (ethyl alcohol) was obtained from Brenntag Canada (Toronto, ON).

3.2.2 Albumin sample preparation

Aqueous solutions of chicken egg albumin at different concentrations were prepared using procedures similar to those described in Nonaka et al (1993). The mass concentrations (w/v), i.e., mass of the constituent divided by volume of the mixture, of the prepared albumin solutions were from 8-16% at 1% intervals. Albumin powder was weighted using a laboratory precision balance (model PC 180, Mettler Toledo Canada, Mississauga, ON) and dissolved in distilled deionized water. For each sample, 30 ml solution was prepared. The albumin powder does not readily dissolve in water and mechanical stirring was used to achieve sample homogeneity. Powder aggregation in solution was especially apparent at higher concentration levels (>15% (w/v)). A magnetic stirrer was first placed in the flask containing the mixture of egg albumin powder and water. The flask was then placed on an unheated hot plate stirrer (model PC 351, Corning, Lowell, MA) and stirred for a minimum of 15 min or until no aggregation of albumin powder was visible. After stirring, the albumin solution was divided into three aliquots and stored in marked glass vials. The glass vials were then immediately transferred to a refrigerator and maintained at 5 °C until spectral measurements were carried out. After sample preparation was done, there were three batches of solutions at all concentration levels. One batch of the solution was used directly for spectral measurement. The other two batches were heated using a temperature controlled shaking Dubnoff metabolic water bath (Thermo Fisher Scientific, Waltham, MA). Each of the remaining two batches of the solution containing vials was heated in water bath at 70 °C and 90 °C for 1 h respectively. Immediately after removing from the hot water bath, the vials were cooled in an ice water bath for 5 min. After cooling, the vials containing albumin gels were transferred to the refrigerator maintained at 5 °C and kept overnight. Measurement of aqueous solutions of

albumin started 3-8 h after the solutions were prepared. Measurement of albumin gels started 18-24 h after the gels were prepared.

3.2.3 Ethanol sample preparation

Anhydrous ethanol was measured using adjustable-volume pipette (model Eppendorf research 500-5000 MI and Eppendorf reference 0.5-10 MI, Eppendorf Canada, Mississauga, ON) to prepare aqueous solution. The volume concentration levels were 2%, 3%, 3.5%, 4%, 5%, 6%, 6.5%, 7%, 8%, 9%, and 10%. Ethanol solution was prepared using the following steps. First, calculated amount of anhydrous ethanol was added into a 25 ml graduated cylinder. Second, distilled water was added to the calculated volume. The total volume of each ethanol solution was 20 ml. The prepared solution was immediately transferred into a plastic tube and sealed with an airtight cap. To minimize the effect of possible dissipation of ethanol, spectral measurements were taken within 2 h after the samples were prepared.

3.3 Experimental design

A rational experimental design was used in this research during preparation of aqueous solution of albumin and ethanol (Araujo and Brereton, 1995). Diagrams of the design are shown in Fig. 3.5. Measurements were taken at 5 evenly spaced concentration levels for calibration of linear models. For albumin aqueous solutions, the five levels were 8%, 10%, 12%, 14%, and 16%. For ethanol aqueous samples, the five levels were 2%, 4%, 6%, 8%, and 10%. Samples at level 2, and 4 were replicated twice and samples at level 3 were replicated 4 times. To account for instrumental errors, repeated measurements were performed on the same samples. According to the rational experimental design, four instrumental replicates were adequate to address errors due to instrumentation. Therefore,

for measurement duplicates, measurement of solutions at level 1 and 5 was repeated twice. Two of the solution replicates at level 3 were repeated twice as well. Therefore, a minimal total of 10 separate solution samples and 14 measurements were required. Thus, for a linear calibration, the degrees of freedom were broken down as follows (Table 3.2). For extra degree of freedoms in this experiment, measurement replicates were repeated more than the required number (2) at all levels. For albumin experiment, solutions at concentration levels 9%, 11%, 13%, and 15% were used as independent validation test sets. For ethanol aqueous solution, samples at 3%, 3.5%, 5%, 6.5%, 7%, and 9% concentration levels were used for validation of calibration models.

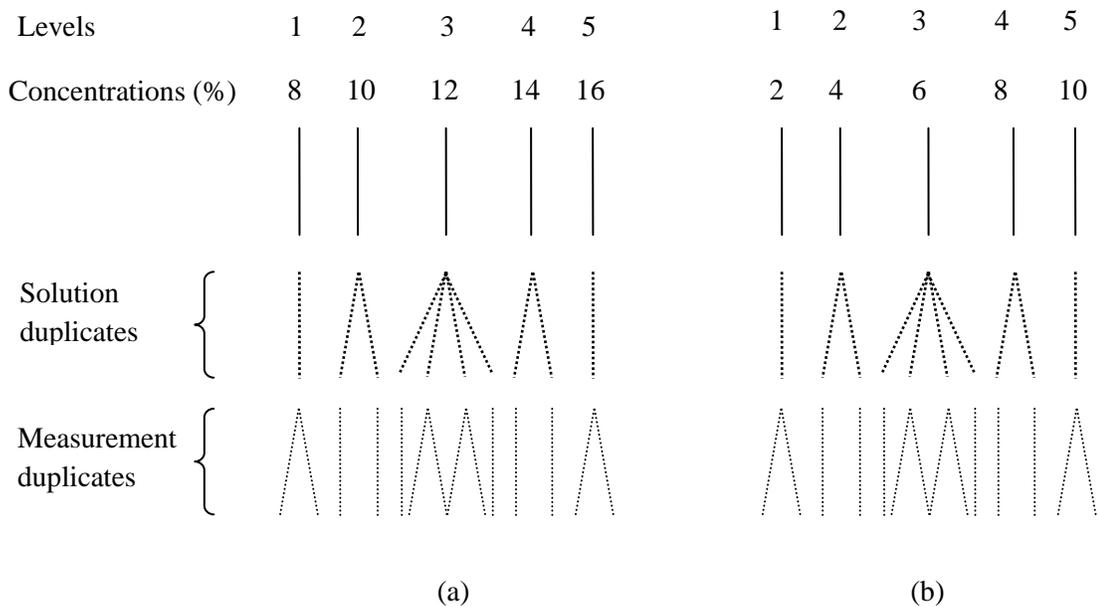


Fig. 3.5 Experiment design diagram. (a) albumin solution; (b) ethanol solution.

Table 3.2 Degrees of freedom for the rational experiment design

Description	Degrees of freedom
Total number of experiments	14
Model parameters	2
Model error	3
Sample error	5
Instrumental error	4

3.4 Data Handling

A list of software, algorithms, and data extraction methods is provided here. Most of the codes for spectra extraction Raman data processing were originally developed in MATLAB (version 7.2.0.232, Mathworks, Natick, MA). Other methods and algorithms were either available in commercial software or obtained from free research software packages.

3.4.1 Software

The original data were acquired using Andor SOLIS for spectroscopy software and saved as space delimited text file. Scientific computing software MATLAB (version 7.2.0.232, Mathworks, Natick, MA) was used for data processing, e.g., raw data handling, file conversion, curve fitting, regression analysis, etc. A set of MATLAB codes GSTools (version 0.4.2, Kris De Gussem, Ghent University) released under the general public licence (GPL) 3.0 was used for file conversion. Spectroscopy software GRAMS AI (version 9.0, Thermo Scientific, Waltham, MA) along with its chemometric software module GRAMS IQ and IQ Predict was used for spectra pre-processing, model calibration, and validation test. A Vancouver Raman algorithm software module (version

1.0.0, BC Cancer Agency and University of British Columbia) was used for fluorescence background removal in Raman spectra. The statistical analysis software SAS (version 9.2, SAS Canada, Toronto, ON) was used for analyzing Raman data for ovalbumin thermal denaturation. A free copy of generalized 2D correlation spectra software 2Dshige (version 1.3, Shigeaki Morita, Kwansei-Gakuin University, Japan) was used to generate 2D spectra of experimental data collected in this research.

3.4.2 Spectral data extraction

Raw data were saved as two dimensional array in space delimited text files (*.asc). The row of the data array denotes the spectral axis of the CCD camera and the column corresponds to the channels. The text files were then imported into MATLAB software first. Columns corresponding to the NIR and Raman channels were then binned along the channel dimension to produce average spectra. For NIR channel, column 12 to 23 and column 32 to 44 were binned along the channel axis to produce the NIR spectra. For Raman channel, columns 230 to 241 were binned along the channel axis to produce the Raman spectra. The averaged spectra along with assigned wavelength (wavenumber) were then exported as files in Thermo Scientific format (*.spc).

3.4.3 Algorithms

The proposed technique was based on diffuse reflectance for NIR spectroscopy. Light scattering noise correction was tried using methods such as standard normal variate (SNV) followed by detrending (DT) (Barnes et al. 1989) and multiplicative signal correction (MSC) method (Geladi et al. 1985). First and second derivatives were calculated using Savitzky-Golay method (Savitsky and Golay 1964), which can smooth the data and remove constant baseline. Raman spectra tend to ride on a strong fluorescence

background (Zhang et al. 2009). The baseline could be approximated as constant slope within a short wavenumber span. A linear piecewise interpolation method was used to correct for the fluorescence background in Raman spectra. An automated fluorescence background removal algorithm was also tested for its suitability to treat the Raman spectra for quantitative analysis (Zhao et al. 2007).

For calibration of Raman spectra, univariate calibration, i.e., classical least squares regression, stepwise multiple linear regression, and PLSR were tested. Because NIR spectra are unintelligible and highly overlapping, full spectra calibration was necessary. Thus, calibration of NIR spectra solely relied on PLSR. Leave-one-out cross validation was used to determine the optimum number of components to remain in the PLSR calibration model. Model performance was evaluated using the following statistics.

R_c^2 : coefficient of determination for calibration

R_t^2 : coefficient of determination for validation

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.3)$$

where \hat{y}_i is the predicted value for sample i , y_i is the actual value for sample i , and \bar{y} is the average of actual values.

RMSEC : Root mean standard error of calibration is defined as

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2}{n_c - f - 1}} \quad (3.4)$$

where n_c is the number of calibration samples, i is an integer with values from 1 to n_c , \hat{y}_i is the predicted value for calibration sample i using the calibration model, y_i is the corresponding actual value of the calibration sample i , and f is the number of factors in the model.

RMSEP : Root mean standard error of prediction is defined as

$$RMSEP = \sqrt{\frac{\sum_{j=1}^{n_p} (\hat{y}_j - y_j)^2}{n_p}} \quad (3.5)$$

where \hat{y}_i is the predicted value for calibration samples using the calibration model, y_j is the corresponding actual value of the validation sample j , n_p is the number of calibration samples, and j is an integer with values from 1 to n_p .

4. RESULTS AND DISCUSSION

4.1 System Calibration

4.1.1 Camera and spectrograph alignment

After preliminary assembly of the ImSpector spectrograph and the camera unit, alignment between the spatial axis of the spectrograph and the pixel lines of the camera was performed. Axis alignment was important to wavelength assignment to pixel columns. Correct axis alignment was also critical when taking advantage of the multiplexing feature of the CCD camera. Blurred spectral peaks meant loss of spectral resolution. Therefore, besides alignment, it was also necessary to adjust the back focal length of the spectrograph so that sharpest spectra peaks could be obtained for monochromatic radiation. Preliminary test spectral images were taken to assess the initial conditions after assembly.

Fig. 4.1 shows spectral images of a fluorescent light before and after the adjustment. In Fig. 4.1(a), the horizontal direction is the spectral axis and the vertical direction is the channel axis. Multiple groups of spectral lines appeared in the image. These tilted and blurred spectral lines showed that the dispersed light exiting the spectrograph was not properly focused onto the image sensor. Also, there existed an angle between the spatial axis of the spectrograph and the pixel line of the camera. To better assess the blurring and misalignment problem, the spectra from channels at both ends and the center, i.e., channel 5, 150, and 280, were extracted and compared (Fig. 4.2). In Fig. 4.2 (a), the spectra are purposely offset by 10,000 photon counts. The bandwidth broadening caused a significant degradation in spectral resolution. Around each peak, the intensity typically fell to half of the maximum values at the fifth or sixth pixel on both

sides of the maximum. Misalignment is evidenced by the three peaks marked by dashed lines. The same fluorescent light spectral peak was offset by 16 pixels between channel 5 and channel 150, and by 14 pixels between channel 150 and channel 280. By changing the position of the back focal lens, the sharpest possible spectral lines were able to be obtained. To align the spectrograph and the CCD camera, the spectrograph was carefully turned with respect to the CCD camera while observing the corresponding peak positions from signal channels on both ends.

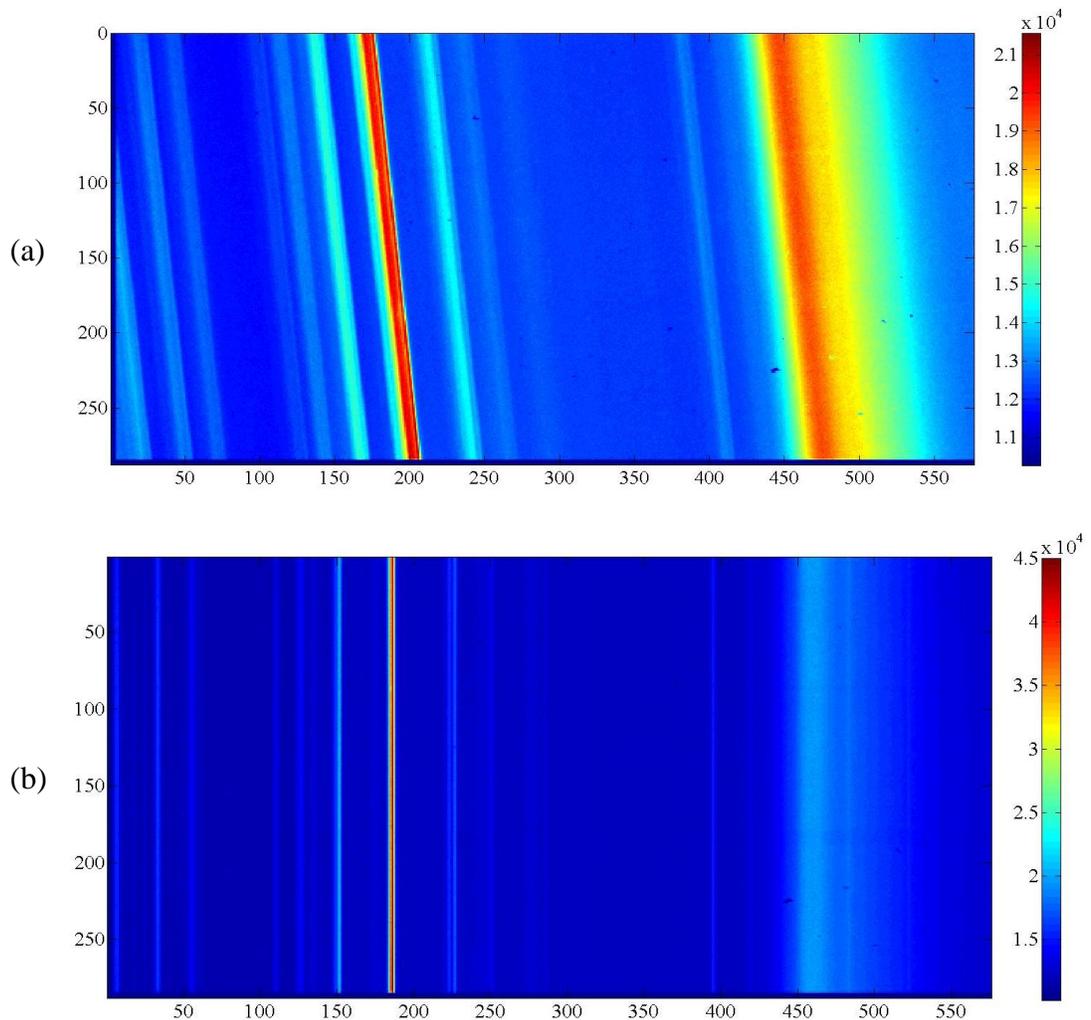


Fig. 4.1 Spectral image of fluorescent light (0.1s integration time, 10 accumulation, and EM gain 100). (a) before adjustment; (b) after adjustment.

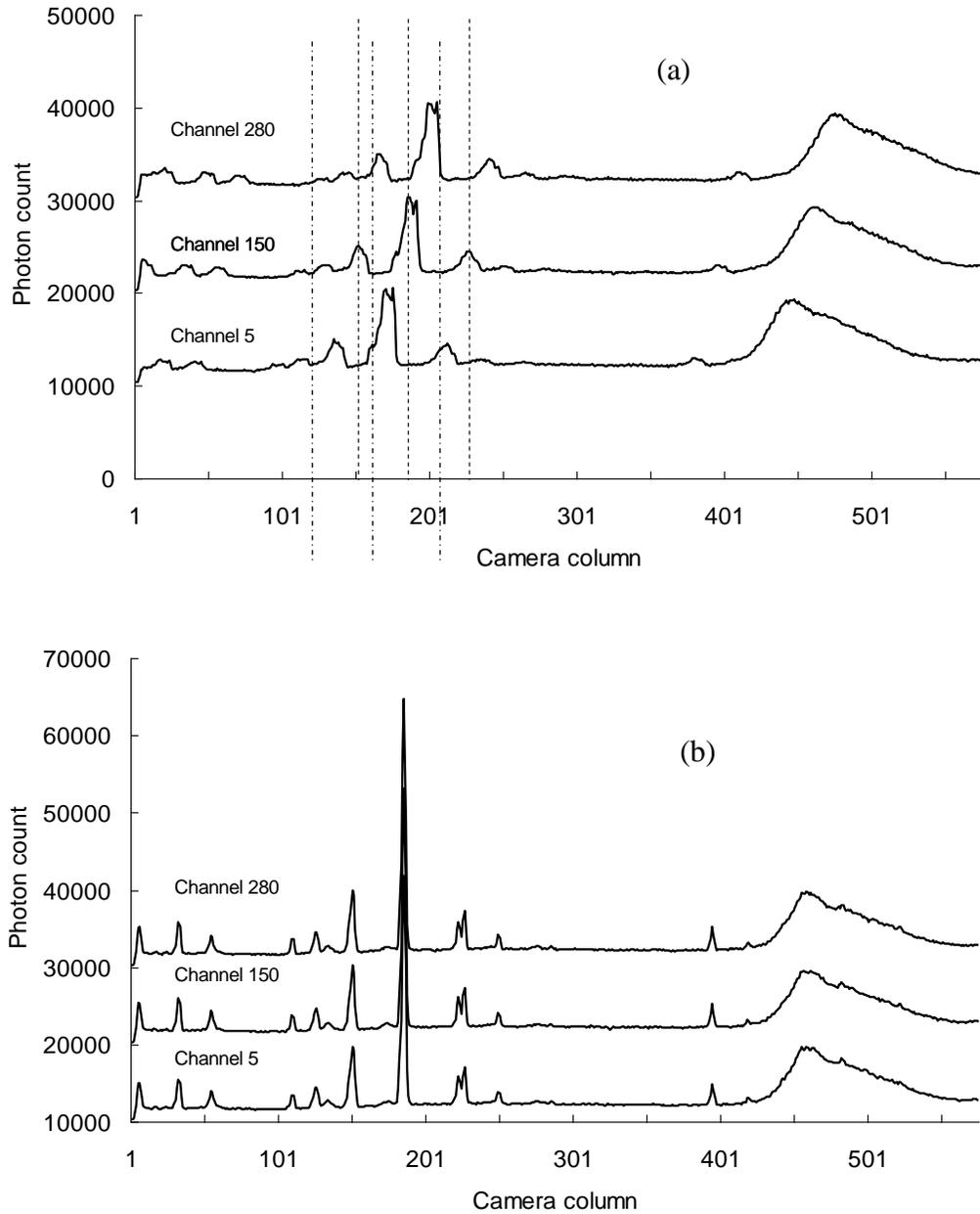


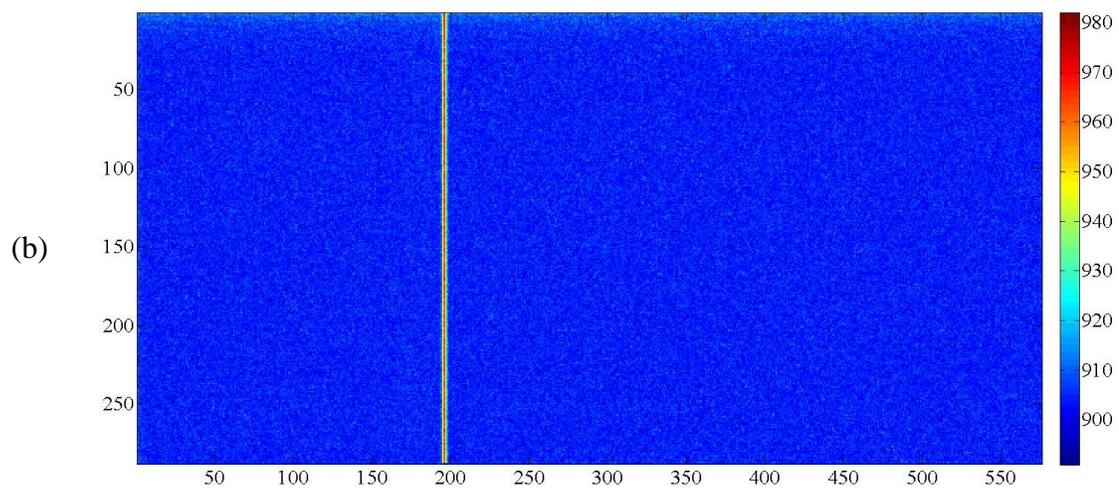
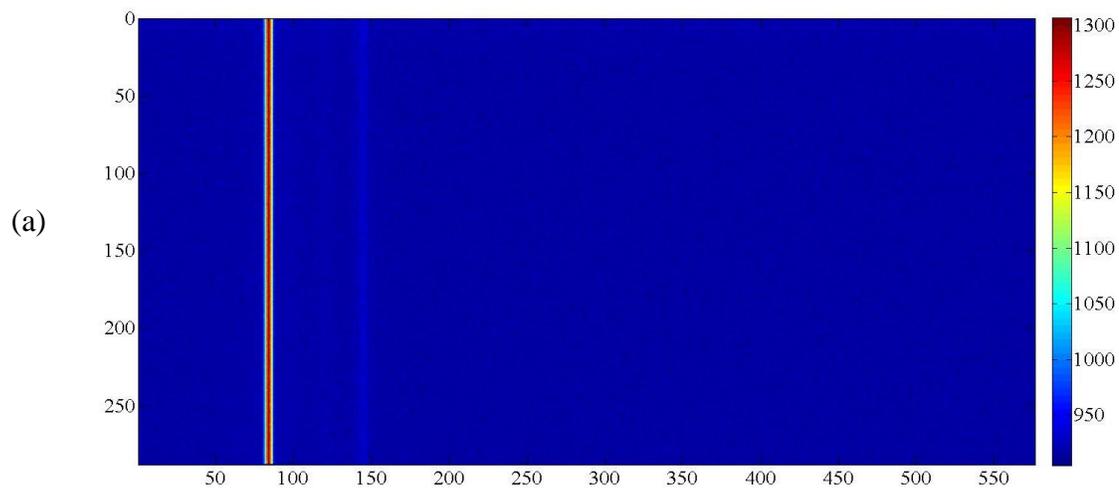
Fig. 4.2 Spectra at channel 5, 150, and 280. (a) before adjustment; (b) after adjustment.

After adjusting the hardware, excellent spectral alignment was achieved among all signal channels. Identifiable peaks aligned on the same pixel column of the camera. The peak intensities fell off to half of the peak value at the adjacent pixel on both sides, i.e., the FWHM was two pixels.

4.1.2 Wavenumber (wavelength) calibration and validation

Two possible ways to calibrate wavenumber (wavelength) versus pixel line were tried. One could either use an absolute wavelength standard, e.g., atomic emission from argon or neon lamps. Or one could calibrate the wavenumbers using relative Raman shift standards, e.g., chemicals with known Raman shifts. Wavenumber calibration of the spectrograph camera assembly was carried out using Raman shift standards in this project. The reason was that Raman shift standard materials were more readily available. On the other hand, wavelength calibration lamps are expensive, require power to operate, and have very short service life. Both benzonitrile (C_6H_5CN) and 4-Acetamidophenol ($C_8H_9NO_2$) were tried because both of these compounds were relatively stable under room temperature and possessed a fair number of Raman shifts within the wavenumber region of interest ($500\text{-}2000\text{ cm}^{-1}$). To establish a base point, the spectral images of a 785 nm laser and an 830 nm were acquired (Fig. 4.3). The output of the free space laser was 50 mW. The spectral images were acquired by carefully shining the collimated laser beam onto a diffuse reflectance board to create a Lambertian source in front of the entrance slit of the spectrograph. The camera ran a single scan with integration time of 0.1s. As is shown in Fig. 4.3(a) and Fig. 4.3(b), the 785 nm spectral line peaked at column 84 and the 830 nm spectral line peaked at column 196. The relative frequency shift between 785 nm and 830 nm is 690.66 cm^{-1} . Thus, column 196 was preliminary assigned as 690.66 cm^{-1} . When acquiring spectra for benzonitrile and 4-Acetamidophenol, the 785 nm laser delivered 104 mW of power at the sample. The camera EM gain was set at 150 and integration time was 2 s. Spectral images for benzonitrile and 4-Acetamidophenol are shown in Fig 4.3. As is evident from the images, both benzonitrile and 4-Acetamidophenol demonstrated several groups of spectral lines in columns beyond

84. Even with the notch filter in the optical path, the intense excitation laser managed to seep through and saturate the camera pixels surrounding column 84. The spectral image of benzonitrile showed less fluorescence background compared to that of 4-Acetamidophenol.



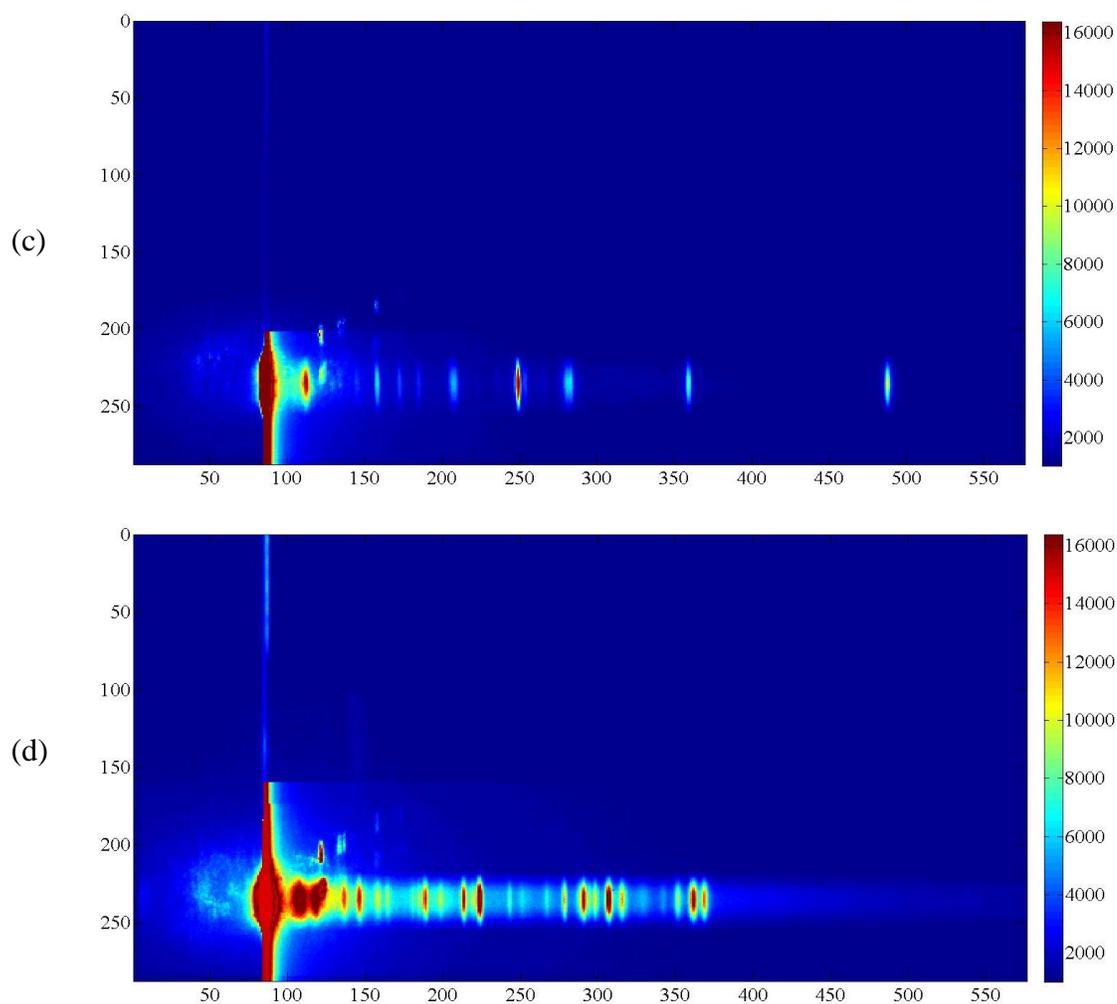


Fig. 4.3 Spectral images for wavenumber calibration. (a) 785 nm laser; (b) 830 nm laser; (c) benzonitrile; (d) 4-Acetamidophenol.

The Raman spectra of benzonitrile and 4-Acetamidophenol were then extracted from the spectral images (Fig. 4.4 and Fig. 4.5). The corresponding Raman spectra that were published by other groups are also shown in the figures for comparison. The Raman spectra of the wavenumber standard materials obtained in this section were not corrected for fluorescence background. The lower wavenumber region ($<300\text{ cm}^{-1}$) and part of the anti-Stoke Raman region are also shown in the Spectra (Fig. 4.4(a) and Fig. 4.5(a)).

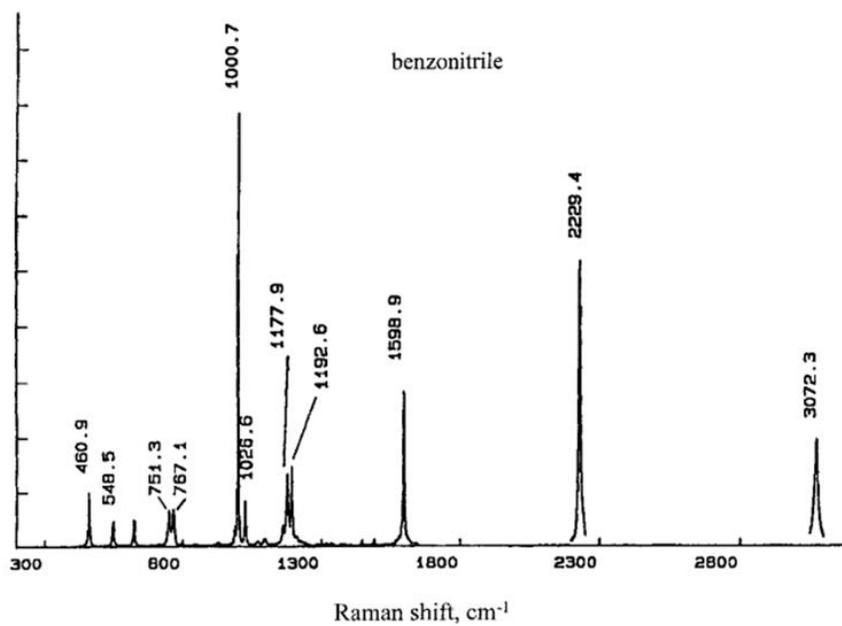
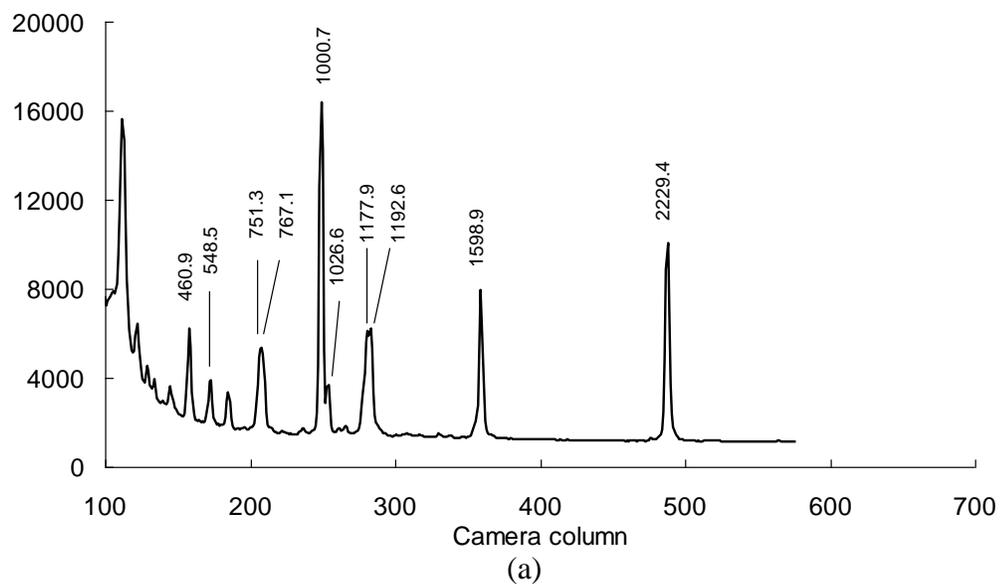
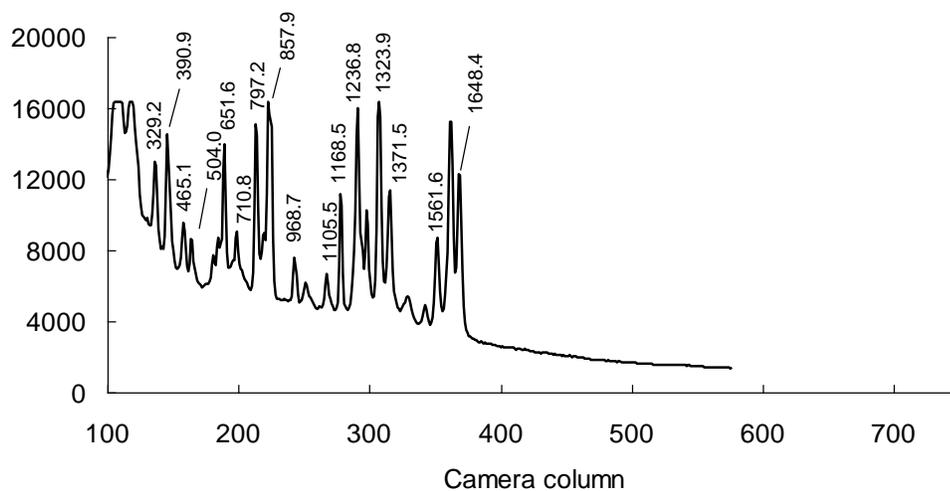
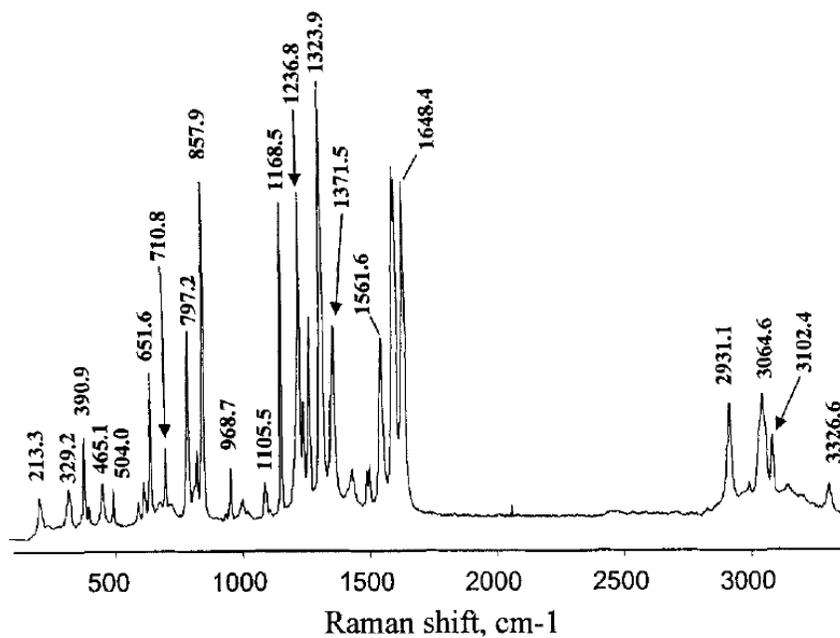


Fig. 4.4 Raman spectra of benzonitrile. (a) spectra collected on camera in this project; (b) published Raman spectra (source: McCreery 2000).



(a)



(b)

Fig. 4.5 Raman spectra of 4-Acetamidophenol. (a) spectra collected on camera in this project; (b) published Raman spectra (source: McCreery 2000).

Using the 830 nm (690.66 cm^{-1}) and published Raman spectra as the reference, wavenumbers were tentatively assigned to the spectral peaks in the acquired Raman spectra of benzonitrile and 4-Acetamidophenol. In terms of spectral features, a great level

of resemblance was shown between the acquired spectra and the published spectra. It was noted that the relationship between wavelength and column number was not linear (Wang and Paliwal 2006). Therefore, polynomial fitting was applied to find out the relationship between camera column numbers (spectral axis) and the designated wavenumbers using wavenumber standards. Because the spectral region between 500 cm^{-1} and 2000 cm^{-1} were of main interest to this project, spectral peaks within this spectral band were selected. Spectrally unresolved peaks, e.g., 751.3 cm^{-1} and 767.1 cm^{-1} , were intentionally left out. The following wavenumbers along with the corresponding column numbers were used for calibrating the camera spectral axis (Table 4.1).

Use of polynomial data fitting could lead to overfitting of calibration data, which results in poor fitting to data points not included in the training. Therefore, it is necessary to find out the proper polynomial order that gives the lowest fitting error to new data points. Here, the optimal order of polynomial fit was determined by calculating the variance defined by Eq. 4.1.

$$VAR = \frac{S_r}{n-(m+1)} \quad (4.1)$$

where VAR is the variance, S_r denotes the sum of squares of residuals, n is the number of data and m is the order of polynomial.

Table 4.1 Wavenumbers used for camera spectral axis calibration

Wavenumber standard material	Column number	Average wavenumber \pm standard deviation (cm^{-1})
Benzonitrile	172	548.5 \pm 0.82
	249	1000.7 \pm 0.98
	253	1026.6 \pm 0.81
	280	1177.9 \pm 0.82
	283	1192.6 \pm 0.56
	359	1598.9 \pm 0.70
	487	2229.4 \pm 0.39
4-Acetamidophenol	146	390.9 \pm 0.76
	158	465.1 \pm 0.30
	164	504.0 \pm 0.60
	189	651.6 \pm 0.50
	199	710.8 \pm 0.68
	213	797.2 \pm 0.48
	220	834.5 \pm 0.46
	223	857.9 \pm 0.50
	244	968.7 \pm 0.60
	267	1105.5 \pm 0.27
	278	1168.5 \pm 0.65
	291	1236.8 \pm 0.46
	298	1278.5 \pm 0.45
	307	1323.9 \pm 0.46
	316	1371.5 \pm 0.11
	342	1515.1 \pm 0.70
352	1561.5 \pm 0.52	
368	1648.4 \pm 0.50	

A maximum polynomial order of 6 was tested. However, it was found that both 5th and 6th order polynomial fit resulted in either large or infinite variance values. Therefore, variance values obtained using polynomial fitting with orders of up to 4 were used to decide the optimum order. Data for benzonitrile was used to calibrate the column versus wavenumber relationship and then tested with both data of benzonitrile and 4-Acetamidophenol. Same procedure was applied to 4-Acetamidophenol data. The variance values were plotted against the polynomial order number in Fig. 4.6. All plots confirmed

that variance either reached minimum or did not decrease significantly after the 2nd polynomial order. Therefore, second order polynomial fit was chosen as the best fit.

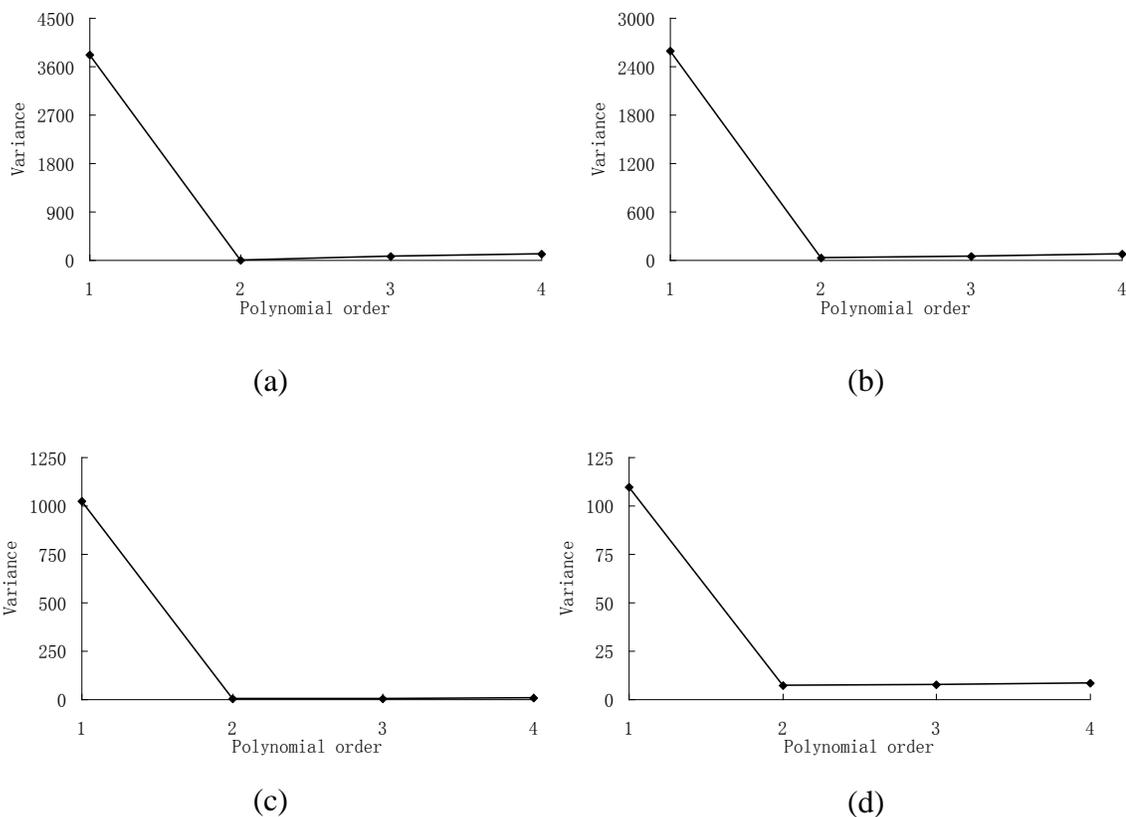


Fig. 4.6 Variance versus polynomial order for different models: (a) calibrated with benzonitrile and validated with benzonitrile; (b) calibrated with benzonitrile and validated with 4-Acetamidophenol; (c) calibrated with 4-Acetamidophenol and validated with benzonitrile; (d) calibrated with 4-Acetamidophenol and validated with 4-Acetamidophenol;

After deciding on the optimum polynomial order, it was necessary to find out which polynomial model to use to calibrate the wavenumbers. The variance values using a second order polynomial are summarized in Table 4.2.

Table 4.2 Variance calculated using second order polynomial

		Calibration	
		benzonitrile	4-Acetamidophenol
Validation	benzonitrile	5.36	3.10
	4-Acetamidophenol	31.75	7.22

As is shown in Table 4.2, wavenumber variance is constantly lower for calibration developed using 4-Acetamidophenol than those of benzonitrile. Therefore, the wavenumber calibration equation using 4-Acetamidophenol was employed (Eq. 4.2).

$$wn = -0.0023x^2 + 6.86x - 559.52 \quad (4.2)$$

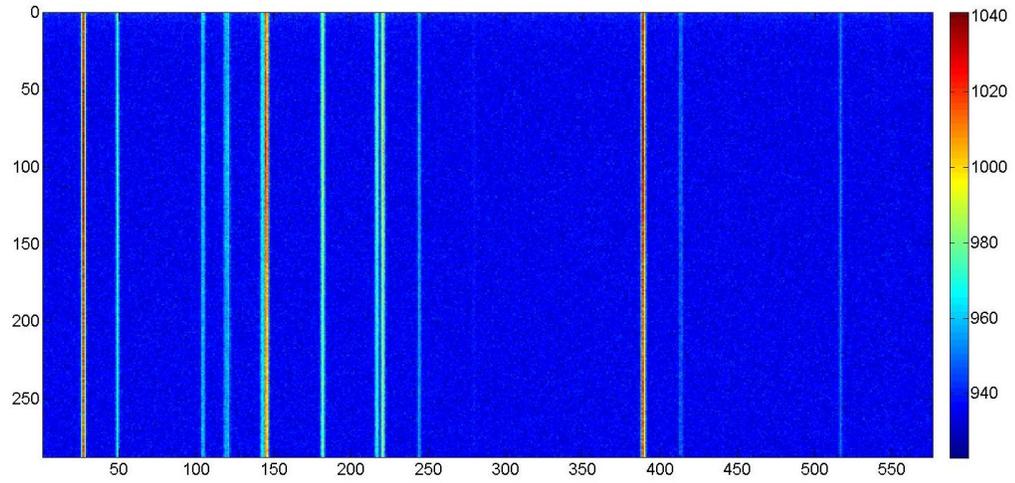
where wn is the calculated wavenumber (cm^{-1}) and x is the column number (unitless).

The excitation laser wavelength 785 nm equals 12738.85 cm^{-1} . The corresponding wavelength for each camera column was obtained using the following equation (Eq.4.3).

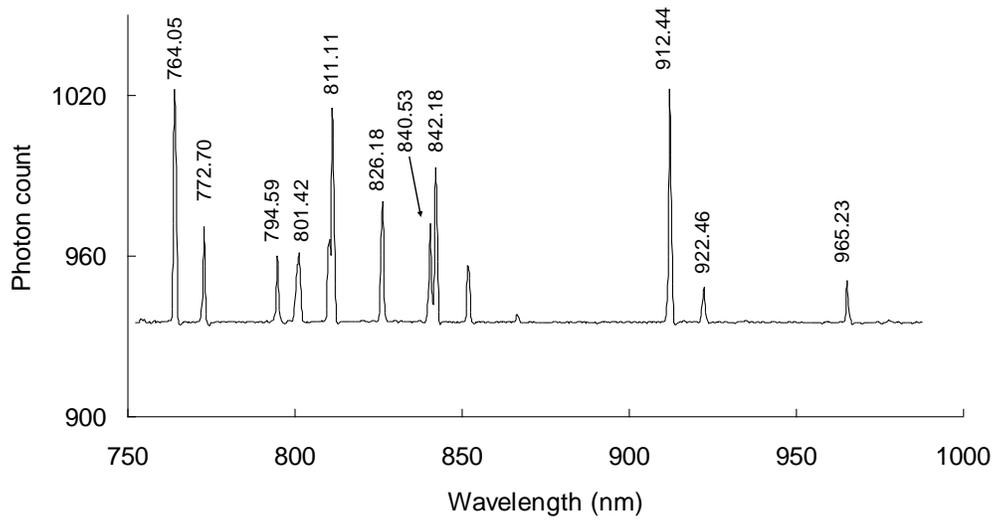
$$wl = 10000000 / (12738.85 + 0.0023x^2 - 6.86x + 559.52) \quad (4.3)$$

where wl is the wavelength (nm).

The assignment of wavelength was validated using a standard argon calibration lamp (model 6030, New Port, Irvine, CA). The spectral image of the argon lamp and the collected spectrum is shown in Fig. 4.7. The following wavelengths were verified (Table 4.3).



(a)



(b)

Fig. 4.7 Argon lamp. (a) Spectral image; (b) spectrum

Table 4.3 Wavelengths used for verification of argon lamp spectrum

Published wavelength (nm)*	Calculated wavelength (nm)	Difference (nm)
763.51	764.05	-0.54
772.38	772.7	-0.32
794.82	794.59	0.23
801.48	801.42	0.06
811.53	811.11	0.42
826.45	826.18	0.27
840.82	840.53	0.29
842.46	842.18	0.28
912.3	912.44	-0.14
922.4	922.46	-0.06
965.8	965.23	0.57

*Source: New Port product catalogue

4.1.3 Spectral range estimation

The spectrograph manufacturer used a different CCD camera with 1600×1200 pixel (model Imperx 2M30, Boca Raton, FL) for testing the spectrograph in this project. Therefore, it was possible that the nominal spectra range and resolution as stated in their test report could not be obtained with the Andor CCD camera used for this project. Though the relationship between wavelength and column number was nonlinear, a linear approximation offers fast estimation with reasonable accuracy. To give a first degree estimation of the available spectral range of the spectrograph and CCD camera system, the spectra of two lasers, i.e., the 785 nm and the 830 nm module, was used. The spectral

lines for 785 nm emission and 830 nm laser appeared at column 84 and 196, respectively. Using this corresponding relationship, a linear regression was performed and Eq. 4.4 was obtained.

$$\lambda = 0.40 * x + 751.25 \quad (4.4)$$

where λ is the wavelength and x is the column position.

According to Eq. 4.1, column 1 corresponded to 751.65 nm and column 576 corresponds to 981.65 nm. Thus the assembled spectrograph and CCD camera system could roughly measure a spectral range that spans from 751.65 nm to 981.65 nm. A more accurate estimate is calculated using Eq. 4.3. Column 1 corresponded to 752.36 nm and column 576 corresponded to 987.74 nm.

4.1.4 Spectral resolution

The spectral resolution was determined by fitting a Gaussian profile to the spectral peaks of monochromatic light and calculating the FWHM of the Gaussian profile. The original spectra along with the fitted spectrum using Gaussian fitting around peaks is shown (Figure 4.8). The calculated spectral resolution as wavelength and wavenumber at tested spectral peaks are listed (Table 4.4).

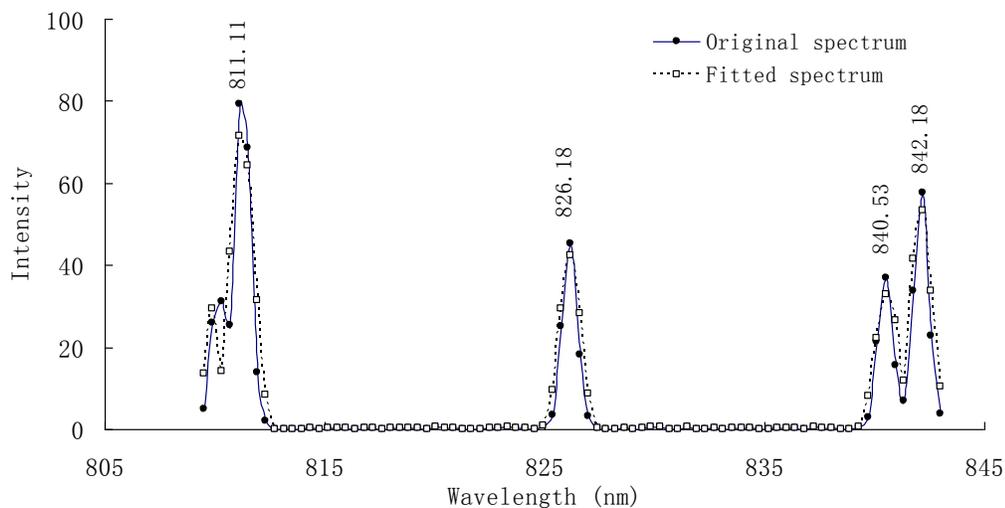


Fig. 4.8 Partial argon lamp spectrum with its spectral peaks fitted using Gaussian profile.

Table 4.4 Calculated spectral resolutions at measured argon lamp peaks

Relative wavenumber (cm^{-1})	Resolution (cm^{-1})	Wavelength (nm)	Resolution (nm)
-349.22	13.55	764.05	0.79
153.71	14.15	794.59	0.89
410.20	14.39	811.11	0.94
634.94	12.44	826.18	0.89
841.63	13.66	840.53	0.97
864.89	12.57	842.18	0.89
1774.24	11.02	912.44	0.92
2378.58	10.53	965.23	0.98

4.1.5 Photometric linearity

Gray scale standards are used to determine the linearity of detectors in a reflectance spectrometric system. The spectra of gray scales at four certified reflectance levels, i.e., 2%, 50%, 75%, and 99%, were acquired at 0.1s with 10 accumulations and no EM gain (Fig. 4.9). Baseline signal was subtracted from the spectra. As is shown in Fig. 4.9, the reflectance spectra are curved as opposed to the optically flat spectra of reflectance standards. This curvature was attributed to the instrumental response. Because the gray scale standards had been subject to prolonged prior usage, their surface appeared to be worn. To minimize spectral variance due to surface inhomogeneity, the reflectance standards were rotated 90° four times with three replicate spectra acquired at each position during spectra collection. To verify the photometric linearity, linear regression between average reflectance and reflectance levels was performed at 755 nm, 800 nm, 850 nm, 890 nm and 950 nm (Fig. 4.10).

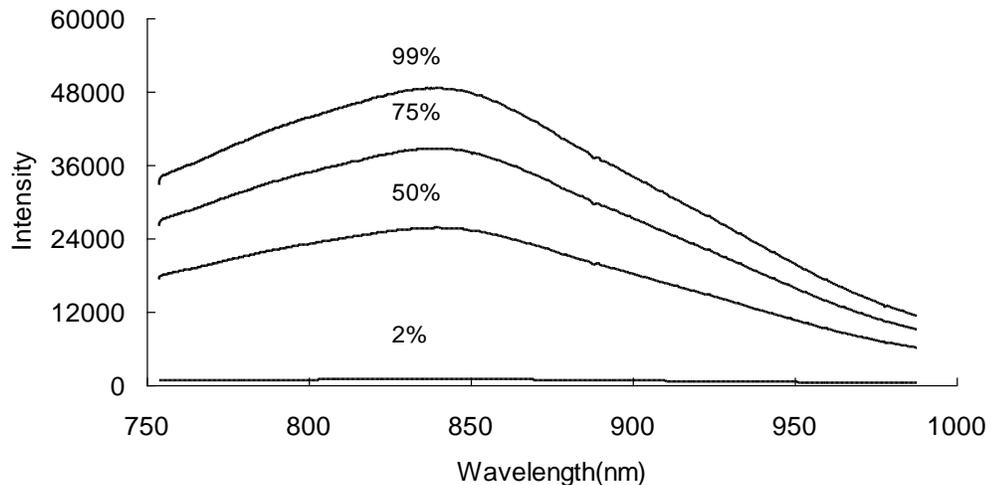


Fig. 4.9 Average spectra of reflectance standards at four certified reflectance levels.

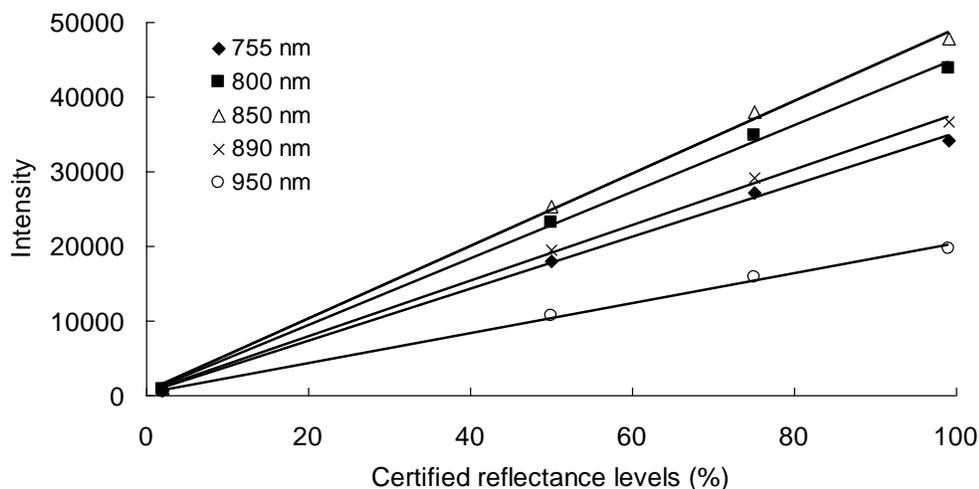


Fig. 4.10 Linear regression between measured reflectance and certified reflectance levels at five different wavelengths.

The coefficients of determination (R^2) for all five regression lines Fig. 4.10 were 0.998, which indicated good linearity of instrumental response (Noble et al. 2012).

Table 4.5 Photometric linearity of the spectrographic system

Wavelength (nm)	755	800	850	890	950
R^2	0.998	0.998	0.998	0.998	0.998

4.1.6 Stability

The instrument intensity stability was tested for both the Raman and NIR channel individually. For testing the instrumental stability of the Raman channel, spectra of 10% (v/v) ethanol aqueous solution hermetically sealed in a quartz sampling cell were acquired every 20 min for 3 h 20 min. Each acquisition was conducted with 10 s integration time, 6 accumulations, and an EM gain of 150. The NIR spectra were acquired with a 0.1 s integration time, 6 accumulations, and no EM gain. Because time

required to take Raman spectra was several orders of magnitude longer than that of NIR spectra acquisition, the waiting intervals for NIR spectra acquisition was set to 5 min instead of 20 min. The instrumental stability expressed as percent standard deviation (SD) (Eq. 4.5) for both Raman and NIR channel is shown in Fig. 4.11.

$$\text{Percent } SD = \frac{SD_w}{\bar{I}_w} \times 100 \quad (4.5)$$

where SD_w is the standard deviation of intensity readings at wavelength or wavenumber w (nm or cm^{-1}) for all repeated measurements and \bar{I}_w is the average intensity of repeated measurement at the same wavelength or wavenumber w .

For the Raman channel, the plot only showed percent SD starting from 500 cm^{-1} . As is seen in Fig 4.11, the average SD for Raman channel was 0.72%. If the spectral range proceeding 800 cm^{-1} was excluded during calculation, the average SD was below 0.67%. For NIR channel, the average SD was 0.39%.

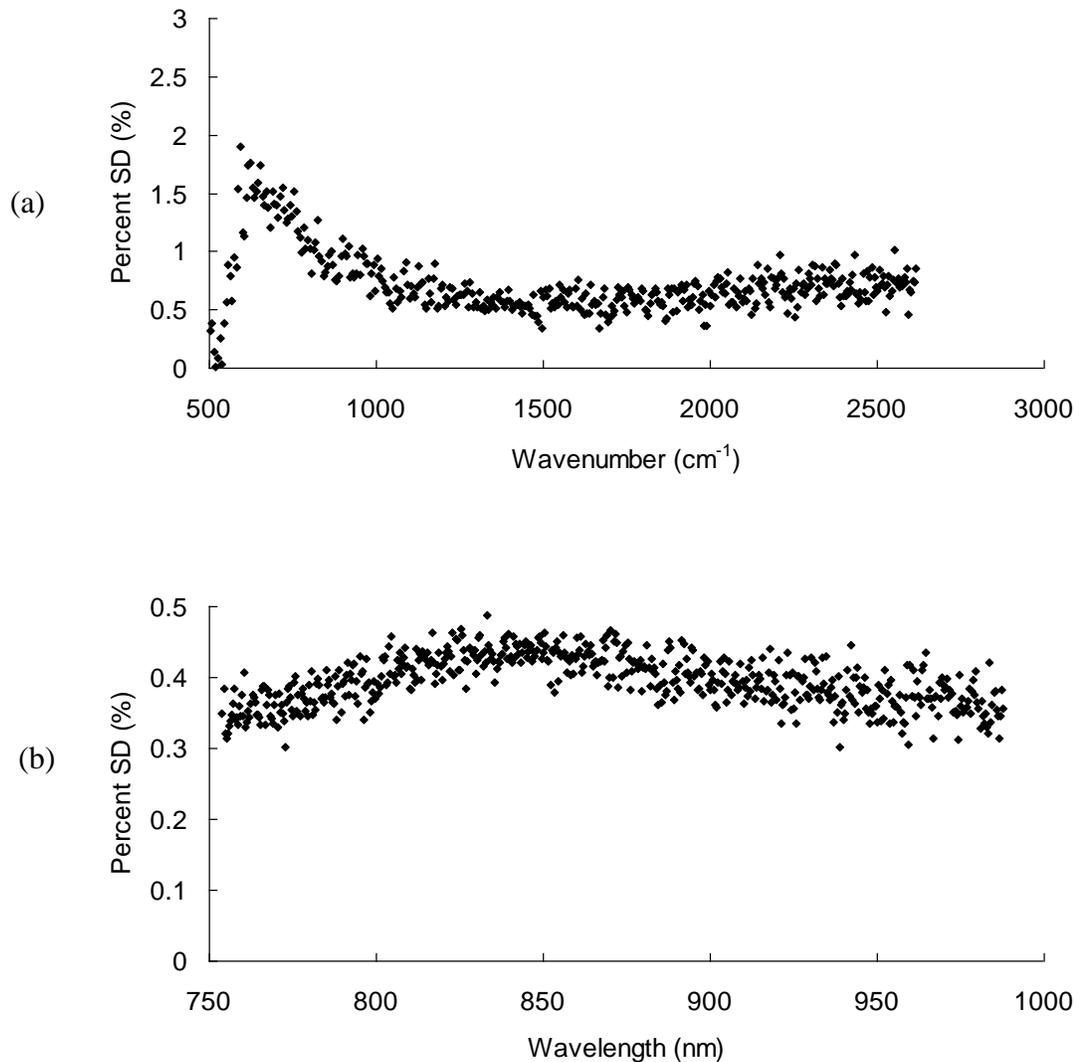


Fig. 4.11 Percent SD plot. (a) Raman channel; (b) NIR channel.

Because quantification experiment using Raman technique was planned on the system, the accuracy of the Raman scattering intensity depended on the accuracy of the abscissa of the spectra, i.e. the wavenumber. Therefore, it was necessary to characterize the abscissa accuracy over the long term. As was noted in an earlier study, the abscissa accuracy depended on factors such as slit width change, grating line spacing, and grating movements during fine tuning in a scanning monochromator (Fountain et al. 1998).

Among all factors, grating movement severely affect the abscissa accuracy (Berg and Nørbygaard 2006). In the currently described Raman system, all the aforementioned factors affecting abscissa accuracy were absent due to the fixed optics design. Alignment between camera and spectrograph and the back focal length are two adjustable settings that possibly correlated with the factors affecting abscissa accuracy. However, once calibration between camera and spectrograph was done, the hardware was tightly integrated without any further modification and the entire system was isolated from any possible mechanical disturbance. Possible abscissa changes in Raman shift could be caused by the wavelength drift in excitation light, i.e., the 785 nm laser. However, the frequency stability of the laser was 0.2 cm^{-1} during a week's testing as stated by the manufacturer. Such minute frequency change (0.2 cm^{-1}) could be barely shown as an abscissa error because the sampling intervals were greater than 4 cm^{-1} in the constructed multimodal Raman and NIR system. Routine check of the Raman spectra of pure ethanol showed peaks constantly at 882.29, 1053.92, 1093.36, and 1454.00 cm^{-1} according to the calibrated wavenumber abscissa.

4.2 Quantifying Ethanol in Aqueous Solutions

Both Raman and NIR absorption spectra of ethanol aqueous solutions were acquired and subject to quantitative analysis. For Raman data, peak heights and integrated peak areas were calibrated against corresponding ethanol concentration levels using univariate regression, stepwise multiple linear regression, and PLSR. For NIR absorption, full spectra were tested with different data pretreatment methods to compare their effect on PLSR calibrations. All calibration models were evaluated with independent test data.

4.2.1 Raman spectroscopy

In this section, Raman spectroscopy was utilized to study ethanol hydration and correspondence between ethanol concentration and Raman peak intensity.

4.2.1.1 Raman spectra of pure ethanol and hydrogen bonding effect

Within the available spectral range of the calibrated spectrometric system, the observable Raman shifts in Raman spectra of anhydrous ethanol and ethanol water mixtures are listed in Table 4.6. Due to the effect of hydrogen bonding between water molecules and ethanol in ethanol-water binary solutions, there exist vibrational frequency shifts at some of the Raman peaks (Numata et al. 2011). This type of hydrogen bonds' incurred frequency shift depends on ethanol concentration in aqueous solution. In this experiment, it was demonstrated that there exists a frequency shift for the C–O stretch at 1053.92 cm^{-1} of anhydrous ethanol sample to 1042.61 cm^{-1} in the 10% (v/v) ethanol aqueous solution. There was also an observable Raman shift for the rocking band of CH_3 that changed from 1093.36 cm^{-1} in anhydrous ethanol to 1082.16 cm^{-1} (Fig. 4.12). This result agreed qualitatively with the data published by Mizuno et al. (1995). Burikov et al. (2012) observed no frequency change for the C–O stretch around 1056 cm^{-1} and a frequency change from 1116 cm^{-1} for 90% (w/w) ethanol to 1110 cm^{-1} for diluted solution (<10% (w/w)). There was no frequency shift at all other Raman peaks on the current spectrometric system, which corroborated with Burikov and co-authors' findings.

Table 4.6 Assignment of Raman peaks and shift due to hydrogen bonding

Raman peaks in pure ethanol (cm^{-1})	Assignment	Shift due to hydrogen bonding in ethanol-water mixture (cm^{-1})
882.29	Stretching vibrations of C–C	No shift
1053.92	Stretching vibrations of C–O	1042.61
1093.36	Librations, rock vibrations of CH_3	1082.16
1276.22	Torsion and rotational vibrations of CH_2	No shift
1454.00	Bending vibrations of CH_3 and CH_2	No shift
1480.49	Bending vibrations of CH_3	No shift

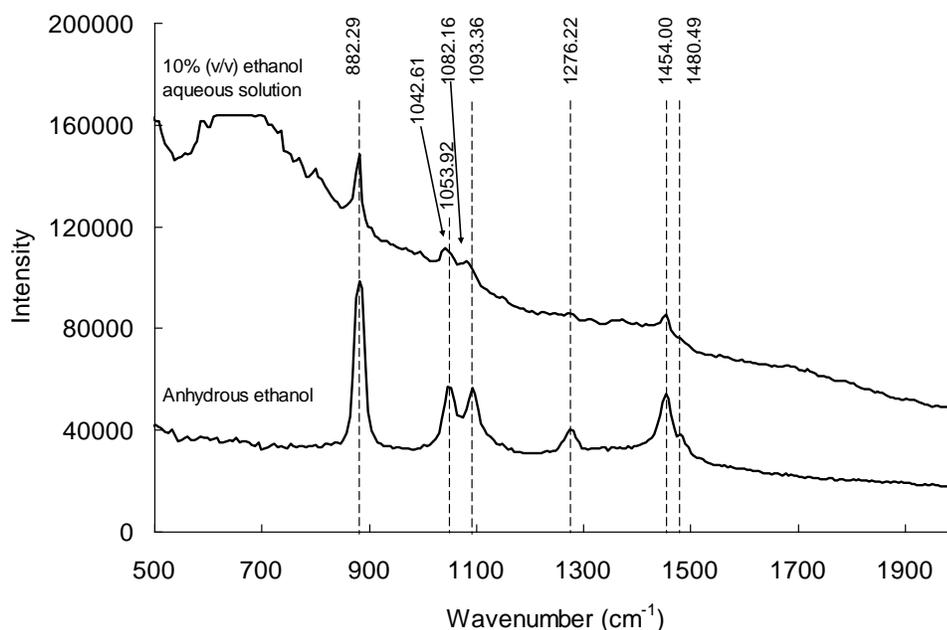


Fig. 4.12 Raman spectra of anhydrous ethanol and 10% (v/v) solution.

4.2.1.2 Quantitative determination of ethanol concentration

Unlike the highly overlapping and unintelligible NIR spectra, the Raman peaks for ethanol were observable at well-defined wavenumbers. Therefore, it was sensible to

apply a univariate regression approach to calibrate the peak intensity to the concentrations of ethanol aqueous solution. Because the peak at 882.16 cm^{-1} was the most intense among all observable Raman shifts for ethanol, baseline corrected peak height at 882.16 cm^{-1} were used for univariate calibration. The peaks for C–O stretching and molecular vibrations of CH_2 and CH_3 were also available. It was of interest to find out if adding other wavenumber variables to the model could possibly improve the prediction accuracy. Therefore, stepwise multiple linear regression was applied to select a subset of wavenumber variables most correlated with the ethanol concentration. The tested wavenumber set included 882.16 , 1042.61 , 1082.16 , and 1454.00 cm^{-1} . To calibrate a linear model, Raman peak heights were regressed against their corresponding concentration levels at 2%, 4%, 6%, 8%, and 10%.

For Raman peak intensity, two calculation methods were tested. One method involved calculating the peak height (length h in Fig. 4.13) against the baseline using Eq. 4.6.

$$h = I_d - \frac{w_d - w_a}{w_b - w_a} (I_b - I_a) - I_a \quad (4.6)$$

where I_d , I_b , and I_a are the spectral intensity readings at point d (Raman peak), b , and a , w_d , w_a , and w_b are the wavenumber at points d , a , and b (cm^{-1}).

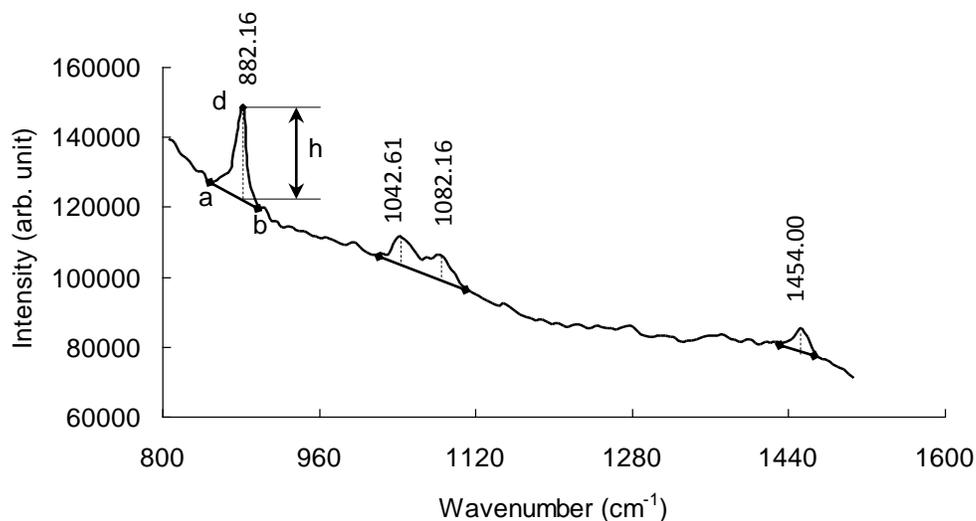


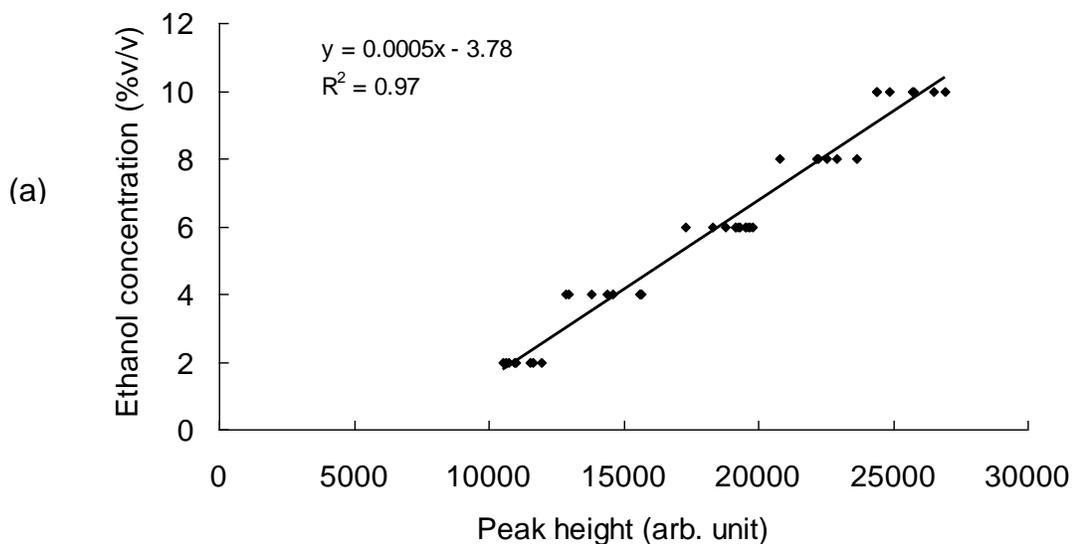
Fig. 4.13 Calculation of peak height.

The other method involved the calculation of the integrated area under a Raman peak using trapezoidal numerical integration in MATLAB. The spectral regions chosen for computing the integrated areas were 864.89-893.86 cm^{-1} for 882.16 cm^{-1} peak, 1031.28-1110.20 cm^{-1} for peaks at 1042.61 and 1082.16 cm^{-1} , and 1432.-1469.91 cm^{-1} for 1454.00 cm^{-1} peak.

Univariate calibration For univariate calibration, both the peak height calculated using Eq. 4.6 and integrated peak area at 882.16 cm^{-1} were regressed against the ethanol concentration in aqueous solution (Fig. 4.14). The following observation could be made.

1. There exist strong linear correlation between Raman intensity for the C–C stretch and ethanol concentration evidenced by coefficient of determination ($R^2=0.97$).
2. Both the peak height and integral area under the Raman peak increased as the ethanol concentration goes from 2% (v/v) to 10% (v/v).
3. Both peak height and integrated area under the peak appeared good estimators for ethanol concentration prediction.

As previously discussed, baseline removal was necessary before integration to offset the spectral variation in the intense fluoresce background. Two baseline correction algorithms, i.e., the Vancouver Raman correction (Zhao et al. 2007) and linear baseline correction (Vickers et al. 1996), were tested. Linear regression was performed on ethanol concentration and integrated peak intensities at 882 cm^{-1} calculated using baseline corrected Raman spectra. The Vancouver Raman algorithm introduced spectral intensity variance unrelated to the ethanol concentration and resulted in a coefficient of determination (R^2) of 0.20 (Fig. 4.14(c)). With linear baseline correction, the R^2 value was 0.97. Therefore, the Vancouver Raman method might not be suitable for quantitative analysis and user discretion was advised.



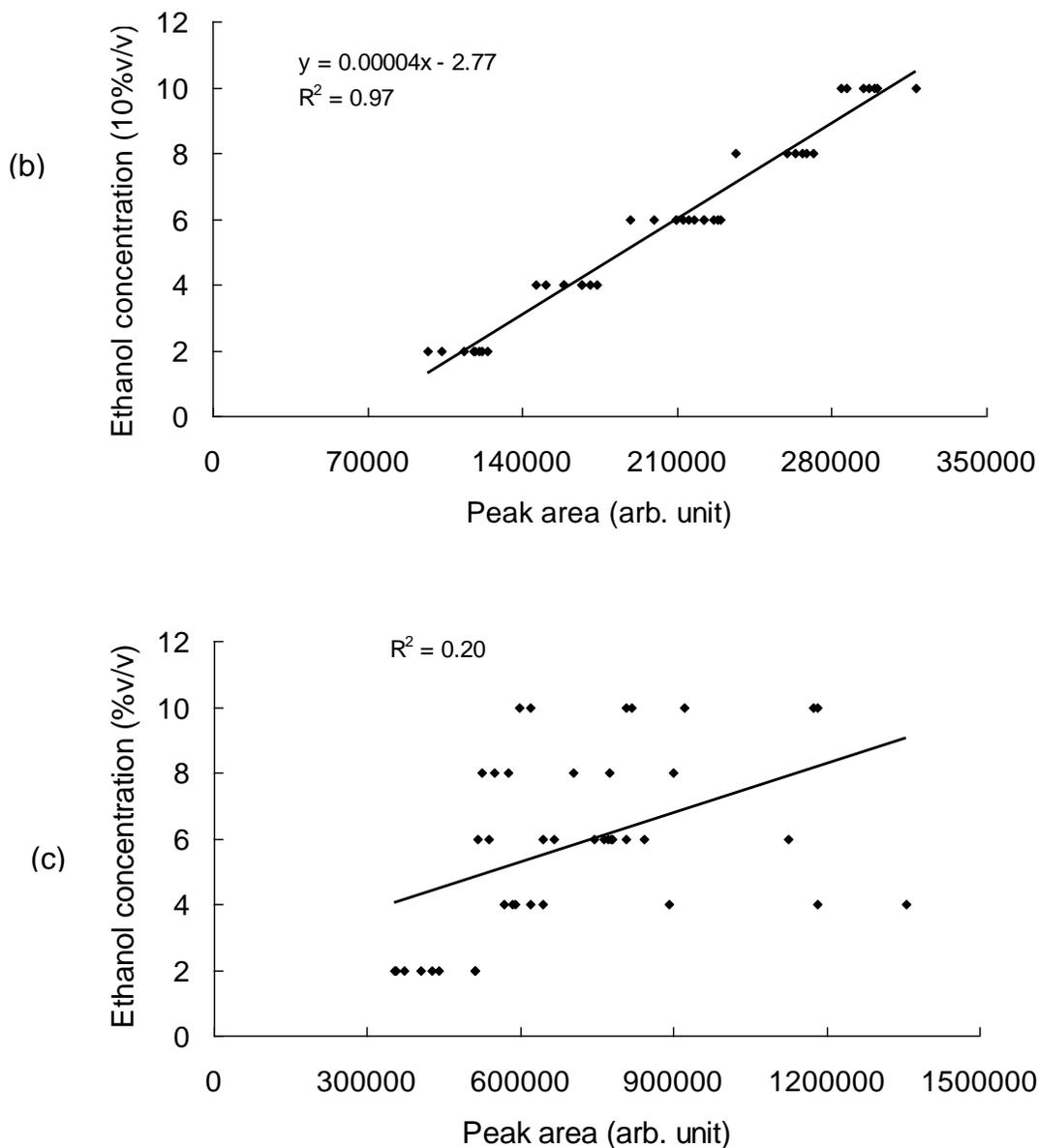


Fig. 4.14 Univariate calibration between Raman signal at 882.16 cm^{-1} and ethanol concentration. (a) peak height versus ethanol concentration; (b) integrated peak area versus ethanol concentration; (c) integrated peak area using Vancouver Raman corrected spectra versus ethanol concentration.

The calibrated regression equations were then used to predict ethanol concentrations of the test set using either peak height or integrated peak area. The

prediction results are shown in Fig. 4.15. The Working-Hotelling 95 percent confidence bands ($\alpha=0.05$) were calculated for the regression lines. The confidence bands denoted a region where the regression lines lay. The Working-Hotelling confidence bands given in Fig. 4.15 were boundaries used to assess the appropriateness of regression equation with a 95 percent confidence interval.

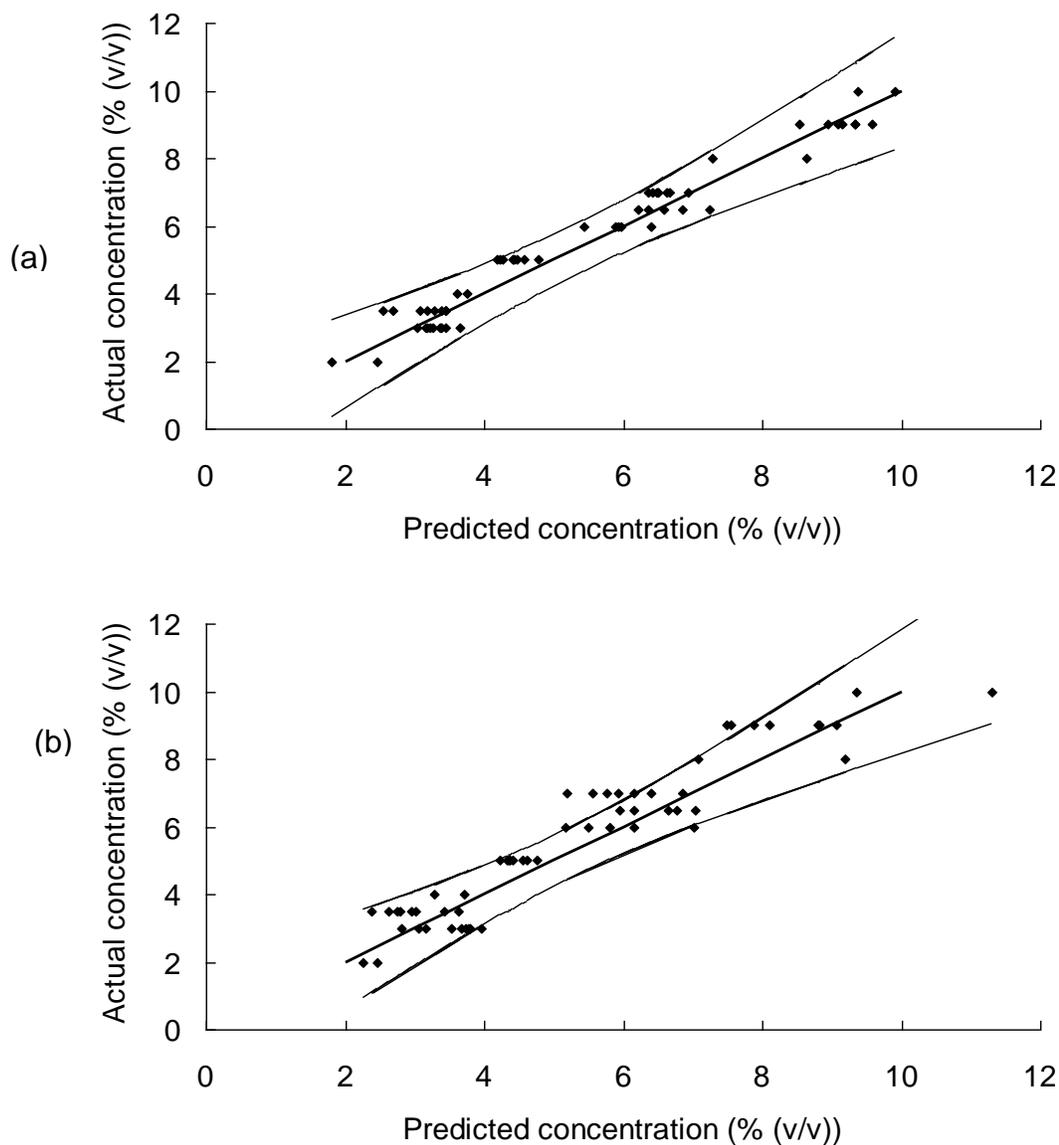


Fig. 4.15 Prediction results using univariate calibration developed with (a) Raman peak height; (b) integrated peak area.

Stepwise regression and partial least squares regression As previously discussed, predictor variables at 882.16, 1042.61, 1082.16, and 1454.00 cm^{-1} could be used in combination so that better estimates of the predictand might be obtained. Use of standard multiple linear regression was not recommended when the problem of multicollinearity exist (Geladi and Kowalski 1986). Multicollinearity defined the case when a strong linear relationship was demonstrated among predictor variables. To diagnose multicollinearity, the statistic measure, variance inflation factor (VIF), was used here (Eq.4.7).

$$VIF_j = \frac{1}{1-R_j^2} \quad (4.7)$$

where VIF_j is the variance inflation factor for the j th predictor, and R_j^2 denotes coefficient of determination in a linear regression of the j th predictor on all other predictors.

The predictor VIFs were calculated for both peak height and integrated peak area (Table. 4.7 and 4.8). If any predictor VIF value was greater than 10, it indicated that multicollinearity existed among the predictor variables (Haan 2001). The multicollinearity posed several issues with the regression equation including yielding models with no statistically significant variables but strong overall regression equation (Belsley et al., 1980). Therefore, the collinearity problem should be dealt with by compressing the data, e.g, deleting one or more variables or constructing new variables using PLSR. Although data compression helped improve robustness of the models, it was also noted that large VIF values did not necessarily indicate poor regression analysis (O'Brien 2007). Stepwise regression and PLSR were tested on the peak height data. Inclusion of full Raman spectra in PLSR models as predictor variables known to contain only noise could

severely degrade the resulting model performance (McShane et al. 1999). Therefore, full Raman spectrum calibration using PLSR was not further pursued.

In Table 4.9, predictor variables at 882.16, 1082.16, and 1454.00 cm^{-1} are selected by stepwise regression to remain in the model for the peak height data. The P-value was the statistical significance with a value between zero and 1. It was the probability of the difference between random sampling and population mean greater than the difference observed in the experiment (Sullivan 2012). For peak area data (Table 4.10), predictor variable at 882.16 and 1454.00 cm^{-1} were selected by stepwise regression to enter the calibration model. Prediction results using stepwise regression are shown in Fig. 4.16.

Table 4.7 VIF values for peak height variables

Predictor variables at Raman peaks (cm^{-1})				
	882.16	1042.61	1082.16	1454.00
VIF	20.03	4.3	6.84	20.12

Table 4.8 VIF values for integrated peak area variables

Predictor variables at Raman peaks (cm^{-1})			
	882.16	1042.61 and 1082.16	1454.00
VIF	14.83	0.71	15.11

Table 4.9 Stepwise regression statistics for peak height variables

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
882.16	2.56×10^{-4}	3.96	3.27×10^{-4}
1042.61	-1.63×10^{-5}	-0.127	0.90
1082.16	2.57×10^{-4}	2.20	3.44×10^{-2}
1454.00	7.77×10^{-4}	3.06	4.07×10^{-3}

Table 4.10 Stepwise regression statistics for peak area variables

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
882.16	2.62×10^{-5}	6.54	0
1042.61&1082.16	4.50×10^{-6}	1.54	0.13
1454.00	4.12×10^{-5}	3.98	0.30×10^{-4}

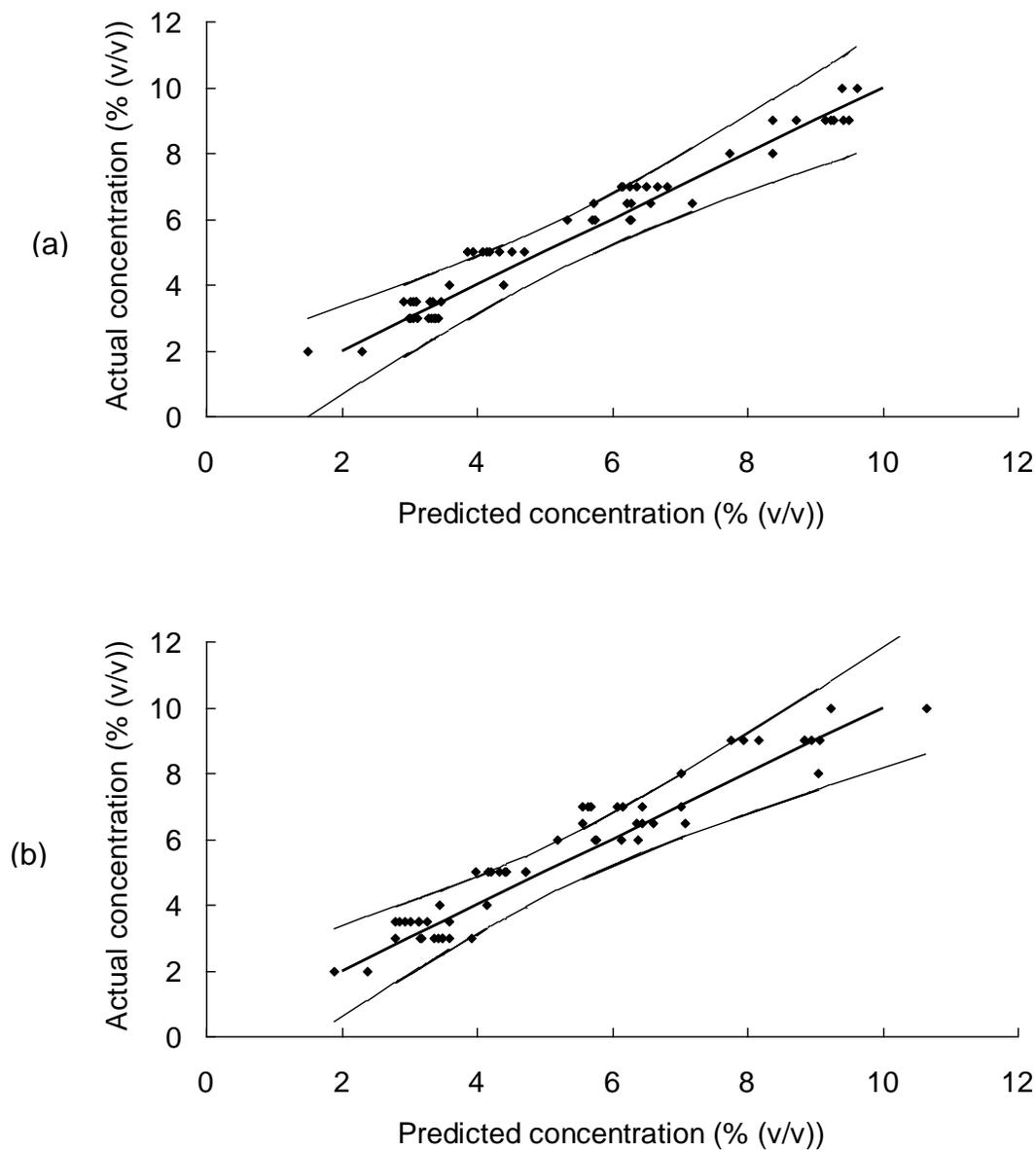


Fig. 4.16 Prediction results using stepwise calibration developed with (a) Raman peak height; (b) integrated peak area

With PLSR method, both the peak height and peak area were regressed to the ethanol concentration. Because the original Raman data were stripped of irrelevant variables other than those at the spectral peaks, the spectral data fed into PLSR was a truncated version of the full spectra. No data pretreatment was applied to the truncated

data. Leave one out cross-validation was applied during calibration. No outlier was detected by calculating the Mahalanobis distance with a threshold of 3. Prediction results using PLSR equations are shown in Fig. 4.17.

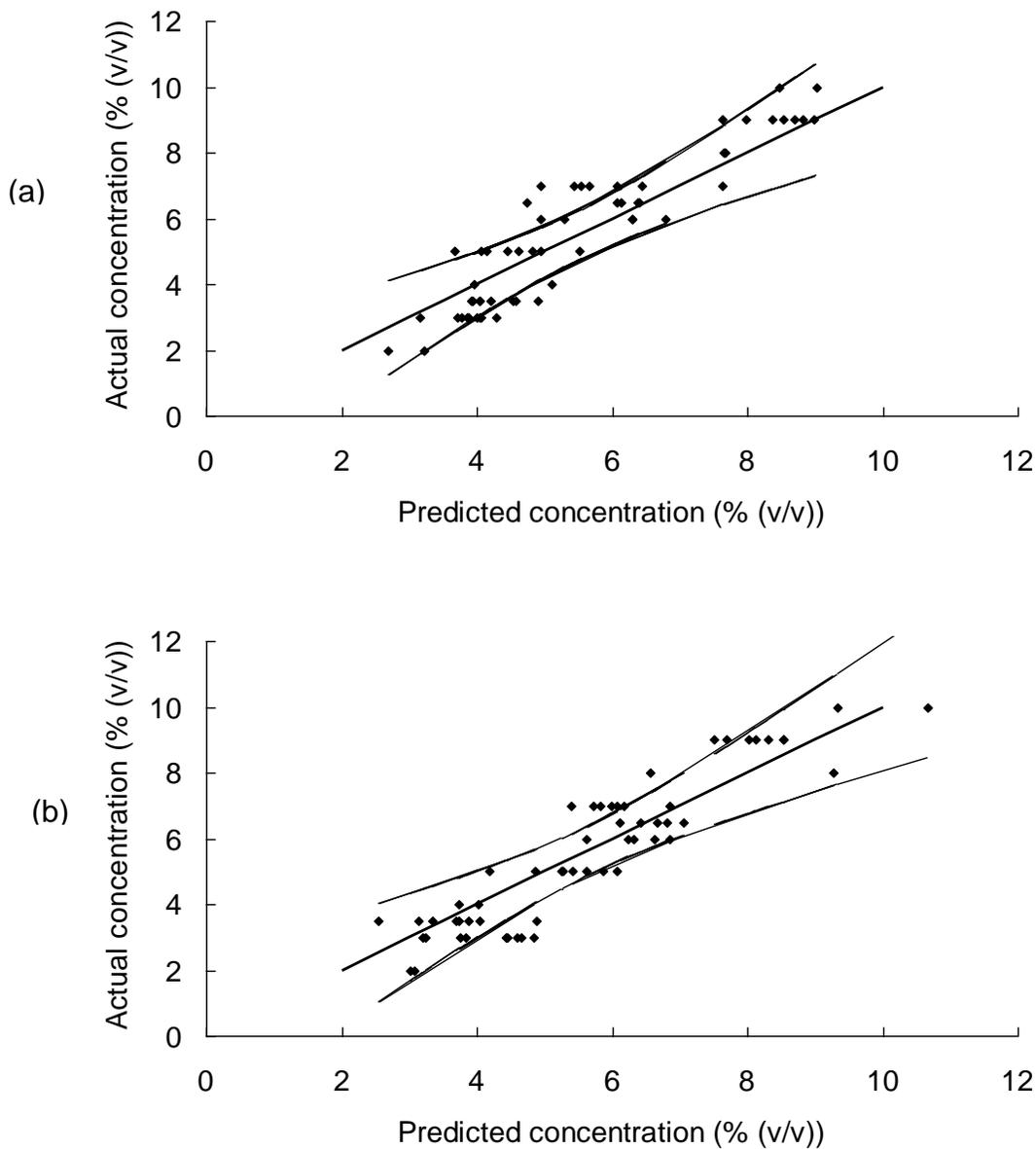


Fig. 4.17 Prediction results using PLSR calibration developed with (a) Raman peak height; (b) integrated peak area.

Model statistics for univariate calibration, stepwise regression and PLSR models are listed in Table. 4.11. As shown in the table, peak height at 882.16 cm^{-1} served as a good predictor of ethanol concentration for both univariate and stepwise regression models compared to their counterparts that used peak area. Stepwise regression offered good calibration statistics but relatively poor prediction results in both categories, i.e., peak height and peak area. This meant that there was overfitting of the calibration data. The model parameters for PLSR models were inferior to both univariate and stepwise regression. The use of peak area seemed to improve the model performance of PLSR as compared to the PLSR model that used peak height data. In terms of model complexity, univariate models only involved one predictor variable while the other two methods tended to involve more variables. In brief, the univariate model using peak height at 882.16 cm^{-1} turned out to be the most succinct model with the best prediction parameters.

Table 4.11 Model statistics for different calibrations on Raman spectra of ethanol aqueous solutions

	Peak height			Peak area		
	Univariate	Stepwise	PLSR	Univariate	Stepwise	PLSR
R_c^2	0.97	0.98	0.93	0.97	0.98	0.96
$RMSEC(\%(v/v))$	0.50	0.40	0.79	0.49	0.41	0.58
Number of predictors/latent variables	1	3	3	1	1	3
R_t^2	0.96	0.96	0.86	0.90	0.93	0.88
$RMSEP(\%(v/v))$	0.45	0.50	0.90	0.77	0.65	0.78

4.2.2 Near-infrared spectroscopy

For its sampling ease, the transmittance measurement mode was selected. The absorbance spectra of anhydrous ethanol, water, and ethanol-water mixtures were acquired. In Fig. 4.18, the absorbance spectra of anhydrous ethanol and water are shown. The broad band peaks near 970 nm in water spectrum at room temperature are attributed to the combination transition, $2\nu_1+\nu_3$, where ν_1 refers to the symmetric O–H stretching and ν_3 the antisymmetric O–H stretch, (Weyer and Lo 2001). In the absorption spectra of ethanol, the peak measured around 908 nm was assigned to the third overtone of the C–H stretch on the methyl group (Cavinato et al. 1999).

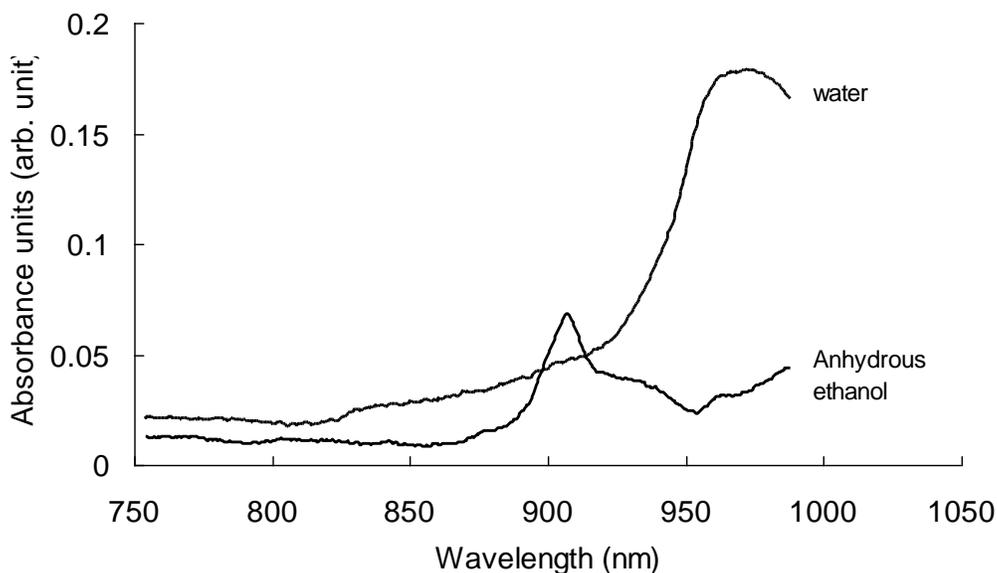


Fig. 4.18 Absorption spectra of distilled water and anhydrous ethanol.

The concentrations of water-ethanol mixtures ranged from 2% to 10 % (v/v). The absorption spectra were preprocessed with mean centering followed by 5-point first-order Savitzky–Golay derivative (Calibration 1). Cavinato et al. (1999) suggested the use of second-order Savitzky–Golay derivative to remove baseline offset due to cuvette

placement and to enhance spectral features. For comparison, second-order Savitzky–Golay derivative with 11-point smoothing was also applied for the second calibration model (Calibration 2). In this experiment, glass cuvette stayed in the optical path. Therefore, spectral variation due to cuvette was minimized. An outlier sample was detected when its calculated Mahalanobis distance was greater than the threshold value of 3. Outlier sample was subsequently removed from training set and the PLSR model was retrained. The minimum predicted residual sum of squares (PRESS) was used to determine how many factors should be kept in the calibration model (Fig. 4.19). The optimal numbers of factors to be kept in the model recommended by the GRAMS IQ software were 3 for Calibration 1 and 6 for Calibration 2. A closer look at the numerical values of PRESS suggested that PRESS values did not decrease significantly after reaching the recommended factor number.

The regression coefficient, i.e., beta coefficients, are shown in Fig. 4.20. The original beta coefficient plot was smoothed using a first order 11-point Savitzky–Golay smoothing filter to remove the noise spikes and reveal the underlying pattern. As seen in Fig. 4.20(a), wavelength variables around 895 nm, between 910 nm and 950 nm, as well as the ones between 965 nm and 982 nm were significant terms for the determination of ethanol content in aqueous solution. The important predictor variables became even more evident as the beta coefficient plot for Calibration 2 was examined. In Fig. 4.20(b), two wavelengths, i.e., 906 and 960 nm, were significant terms compared to other variables. Not surprisingly, these two wavelength variables were closely associated with third overtone of the C–H stretch for ethanol (907 nm) and the broad water band around 970 nm.

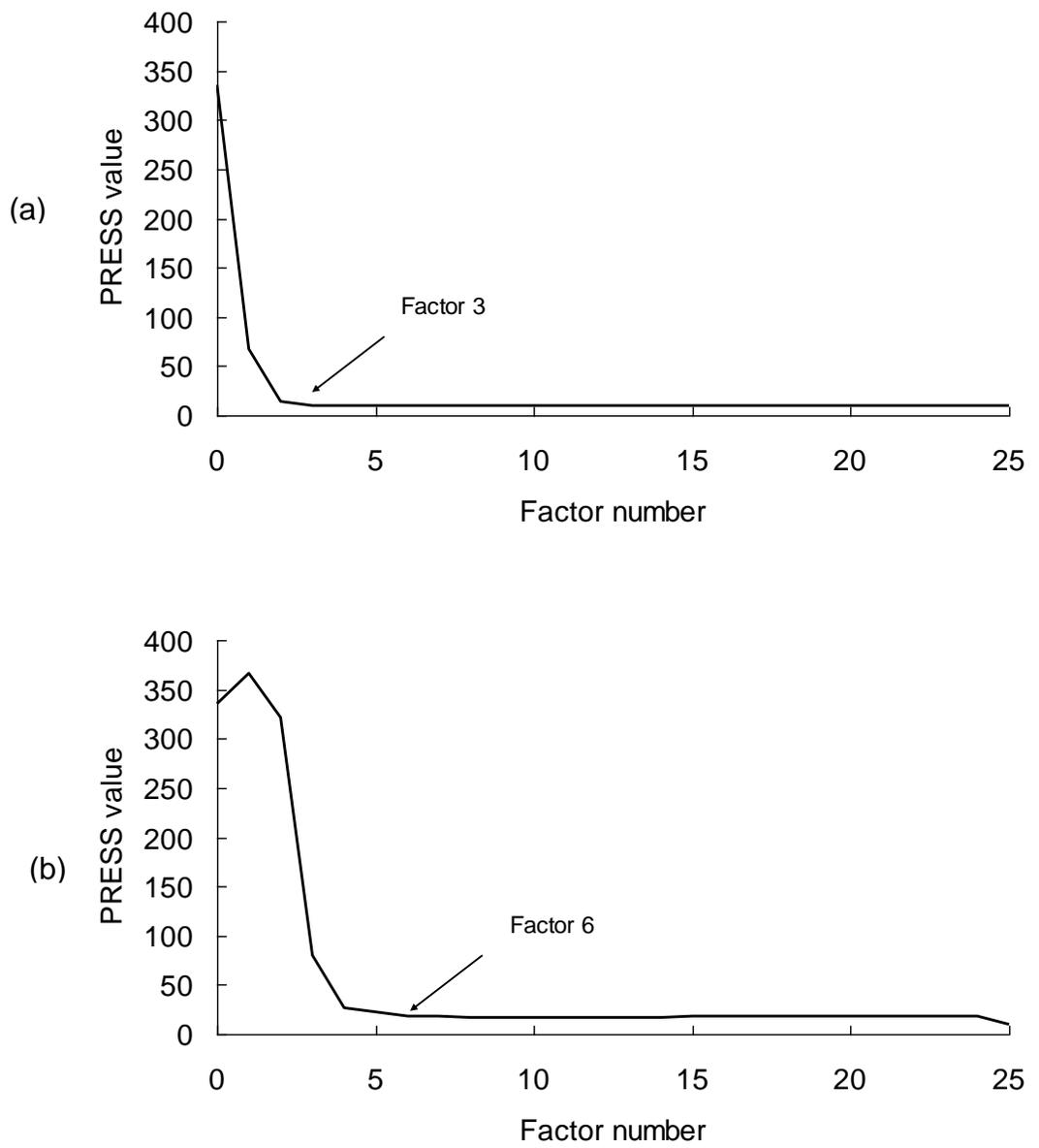


Fig. 4.19 Predicted residual sum of squares values versus factor number. (a) calibration 1; (b) calibration 2.

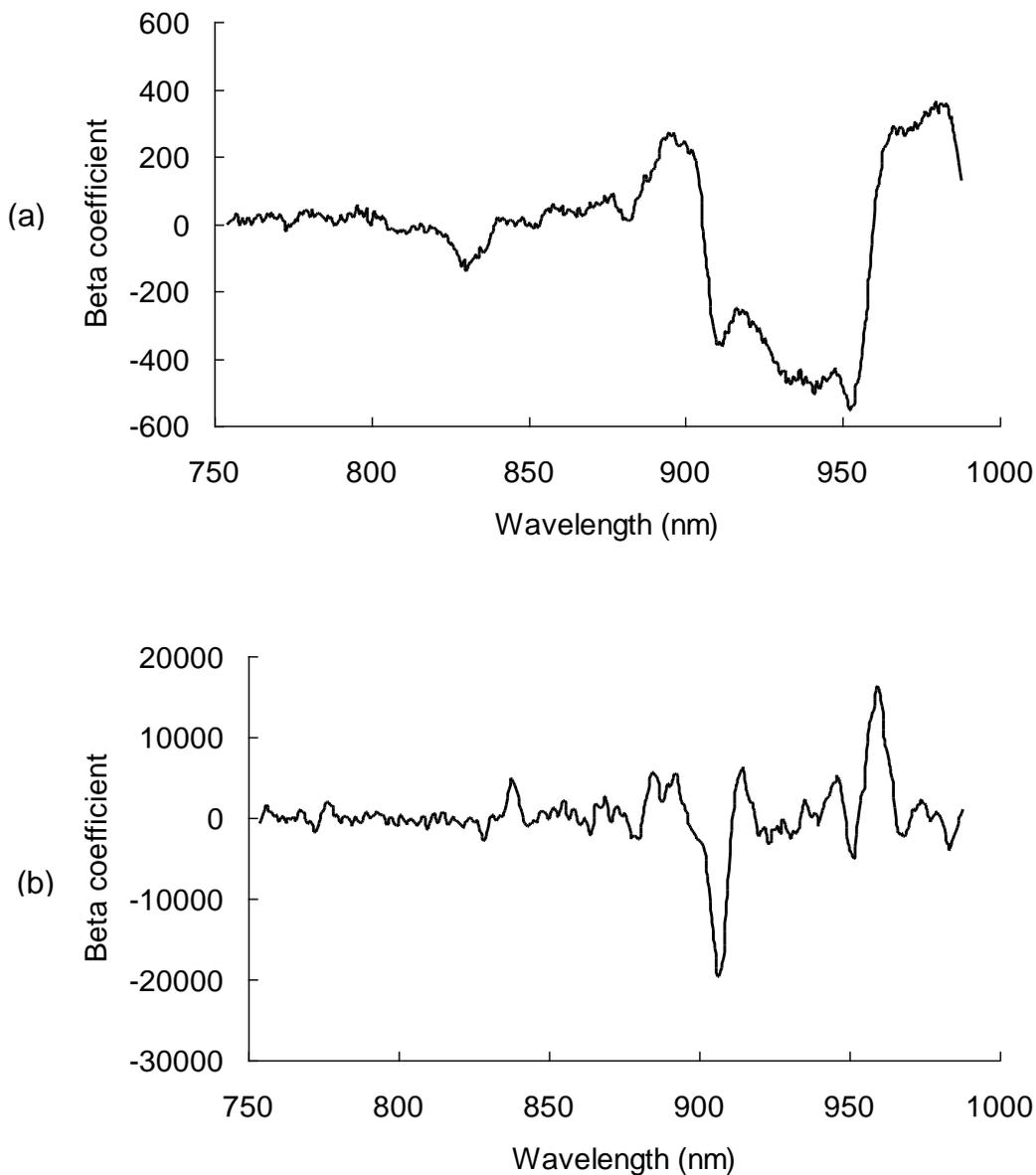


Fig. 4.20 Beta coefficients. (a) calibration 1; (b) calibration 2.

The predicted ethanol concentrations versus actual concentrations are plotted in Fig. 21. Model statistics are summarized in Table 4.12. As seen from the table, calibration 1 had similar training statistics to those of calibration 2. However, calibration 1 was much simpler and outperformed calibration 2 in terms of prediction statistics.

Therefore, mean-centering followed by 5-point first-order Savitzky–Golay derivative seemed to be more appropriate in preprocessing the NIR spectra in this experiment.

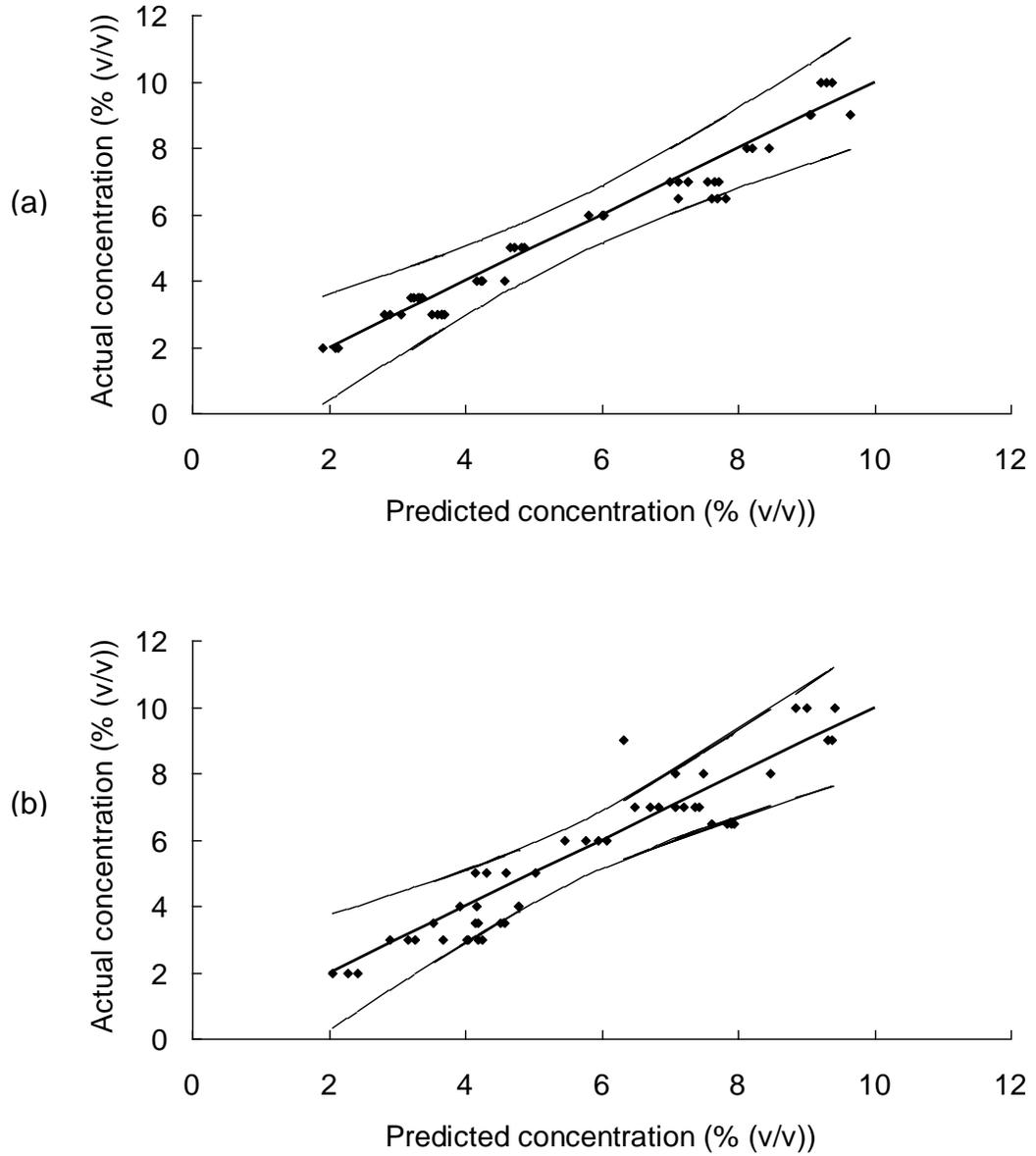


Fig. 4.21 Prediction results using NIR calibrations. (a) calibration 1; (b) calibration

2.

Table 4.12 Model statistics for NIR calibrations of ethanol aqueous solutions

	Calibration 1	Calibration 2
R_c^2	0.96	0.96
<i>RMSEC</i> (%(v/v))	0.52	0.61
Total number of latent variables	3	6
R_t^2	0.96	0.89
<i>RMSEP</i> (%(v/v))	0.49	0.81

4.2.3 Summary

The use of either Raman or NIR spectroscopy to quantitatively determine ethanol concentration in aqueous solutions proved to be feasible. Univariate calibration using single Raman peak at 882.29 cm^{-1} produced the most succinct model with the lowest RMSEP value of 0.45% (v/v). Use of PLSR with full NIR absorbance spectra between 750 nm and 980 nm produced comparable prediction results with a RMSEP value at 0.49% (v/v). On the practical side, NIR technique seemed to be more appropriate as a routine analysis tool. In this experiment, acquisition of a single Raman spectrum required 42 s while it only took 0.5 s to acquire an NIR transmittance spectrum of the same sample. In addition, the hardware, viz., optics and excitation mechanism, were more complicated and less robust for Raman channel compared to that of NIR channel. However, a unique strength about Raman technique was its well-understood spectral features, which enabled reliable qualitative interpretation of molecular structures and most succinct quantitative analysis models.

In review of the previously reported results on ethanol, the ethanol results obtained on this Raman/NIR multimodal spectrometer are comparable or slightly inferior to what other groups obtained on commercial equipment. Ysacc Sato-Berrú et al. (2004) performed quantitative NIR Raman analysis of ethanol in various liquid mixtures. In their experiment, they utilized a Renishaw Micro-Raman spectrometer (Model 1000B, Renishaw, Gloucestershire, U.K.) with an 830 nm Renishaw Laser System (HPNIR Series, Renishaw, Gloucestershire, U.K.). Utilizing the Raman peak at 882 cm^{-1} , they managed to achieve a SEP value of 0.72 (v/v) for binary mixtures. A comparison between NIR and Raman spectroscopy in determining ethanol content in spirit was made by Nordon et al. (2005) The group used a Zeiss Corona 45 NIR reflectance spectrometer (Clairet Scientific, Northampton, UK) with a working spectral range of 939–1708 nm and a Kaiser HoloProbe system with 785 nm excitation ((Kaiser Optical Systems, Ann Arbor, MI, USA) for their experiment. They also found the sharp peak around 880 cm^{-1} to be useful for ethanol concentration determination. Their quantitative results indicated the average accuracy was 2.1 and 2.9%, respectively, for NIR and Raman spectroscopy on spirit samples with ethanol concentrations of 19.9–61.7% (v/v). In a paper by Cavinato et al. (1990), the authors explored the utilization of fiber-optic based NIR spectroscopy for monitoring ethanol during fermentation. They reported a SEP of 0.25% (v/v) in predicting the ethanol concentrations in ethanol/water mixtures over a range of 0-15%. Their instrument was a NIR photodiode array spectrophotometer with bifurcated fiber optic bundle probes (Hewlett-Packard 8452A, Mississauga, Ontario), which was similar to the design in this research project. Mendes et al. (2003) managed to determine the ethanol in fuel and beverages using both FT-NIR and FT-Raman spectroscopy. Both NIR

and Raman spectra were obtained on a Bruker interferometer (Equinox 55, Billerica, MA). Using PLS models, they obtained SEP values for ethanol content in fuel of 0.04 and 0.2% (v/v) for FT-NIR and FT-Raman respectively. For beverage samples with ethanol concentrations over 4.5% - 44% (v/v), the SEP values were around 0.3% (v/v) and 0.6% (v/v) for NIR and Raman, respectively. The author attributed the significant increase in SEP for NIR measurements to the increased complexity of mixture. Also note that the FT-NIR employed the spectral regions of 1394-1542 nm and 1843-2164 nm, which have significantly richer information compared to the short-wavelength NIR region (780-1000 nm).

4.3 Quantitative Analysis of Ovalbumin

The possibility to quantify ovalbumin content in aqueous solutions was explored in this section using both Raman and NIR spectroscopy.

4.3.1 Quantification using Raman spectroscopy

The Raman spectra of solid albumin from chicken egg white (62%-88% pure), 10% (v/v) ovalbumin aqueous solution and distilled water are shown in Fig. 4.22. Due to the limited spectral resolution of the system, Raman peaks were not as well resolved as those obtained on commercial FT spectrometers. However, the strong Raman peaks shown in the solid power of albumin corresponded well with previous findings (Cîțu et al. 2001). There were a few Raman peaks related to protein structure and they were assigned to vibrational modes of protein molecules (Table 4.13). The broad band around 1640 cm^{-1} in the water spectrum was due to ν_2 O–H bending mode of water molecule (Carey 1998).

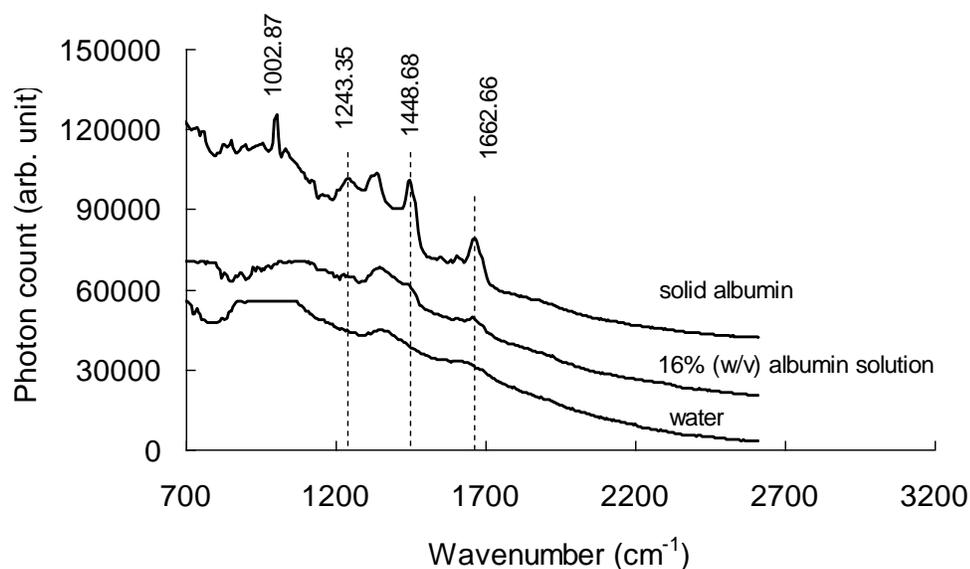


Fig. 4.22 Raman spectra of solid ovalbumin, 16% (w/v) ovalbumin solution, and water in quartz cell.

Table 4.13 Assignment of Raman modes*

Wavenumber (cm ⁻¹)	Origin and assignment	Structural information
1002.87	Phenylalanine Ring breathe	Conformation insensitive
1243.35	Amid III, N-H in-plane bend, C-N stretch	Disordered structure
1448.68	Aliphatic residues, C-H bending	Microenvironment
1662.66	Amide I, C=O stretch, N-H wag	Disordered structure (solvated)

*Souce: Li-Chan 1996.

Raman peaks at 1243.35, 1448.68, and 1662.66 cm⁻¹ were chosen for quantitative analysis of albumin concentration in aqueous solution. For denatured gels, peaks at

1243.35, 1448.68, and 1667.78 cm^{-1} were used for concentration prediction. Use of 1667.78 cm^{-1} instead of 1662.66 cm^{-1} for predicting gel concentration was due to thermal conformational changes of protein structure, which would be shown later. The peak at 1002.87 cm^{-1} was strong in the spectrum of solid powder albumin sample but was unavailable in the spectrum of albumin solution due to saturation of CCD camera in that spectral region. Peak height and peak area were calculated using the same methods as described in section 4.2.1.2. The spectral regions used for linear baseline interpolation and for calculating integrated peak area are listed in Table 4.14. Because intensities of the three peaks appeared to be mild compared to each other, stepwise regression was used. Multivariate outlier detection was performed before regression and no outlier was found at a significance level of 0.05 (Wilks 1963).

Table 4.14 Spectral regions selected for linear baseline and integrated peak area calculation

Peak position (cm^{-1})	Baseline (cm^{-1})	Peak area integration (cm^{-1})
1243.35	1215.83 – 1281.69	1215.83 – 1281.69
1448.68	1281.69 – 1496.33	1416.71 – 1496.33
1662.66, 1667.78	1574.89 – 1733.93	1611.20 – 1713.66

**Table 4.15 Stepwise Regression coefficient statistics for peak height variables of
albumin solution**

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
1243.35	1.7×10^{-3}	2.32	0.027
1448.68	3.6×10^{-3}	6.50	3.6×10^{-7}
1662.66	0	0.35	0.73

**Table 4.16 Stepwise Regression coefficient statistics for peak area variables of
albumin solution**

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
1243.35	0	0.23	0.82
1448.68	5.8×10^{-5}	2.52	0.017
1662.66	5.1×10^{-5}	3.44	0.0017

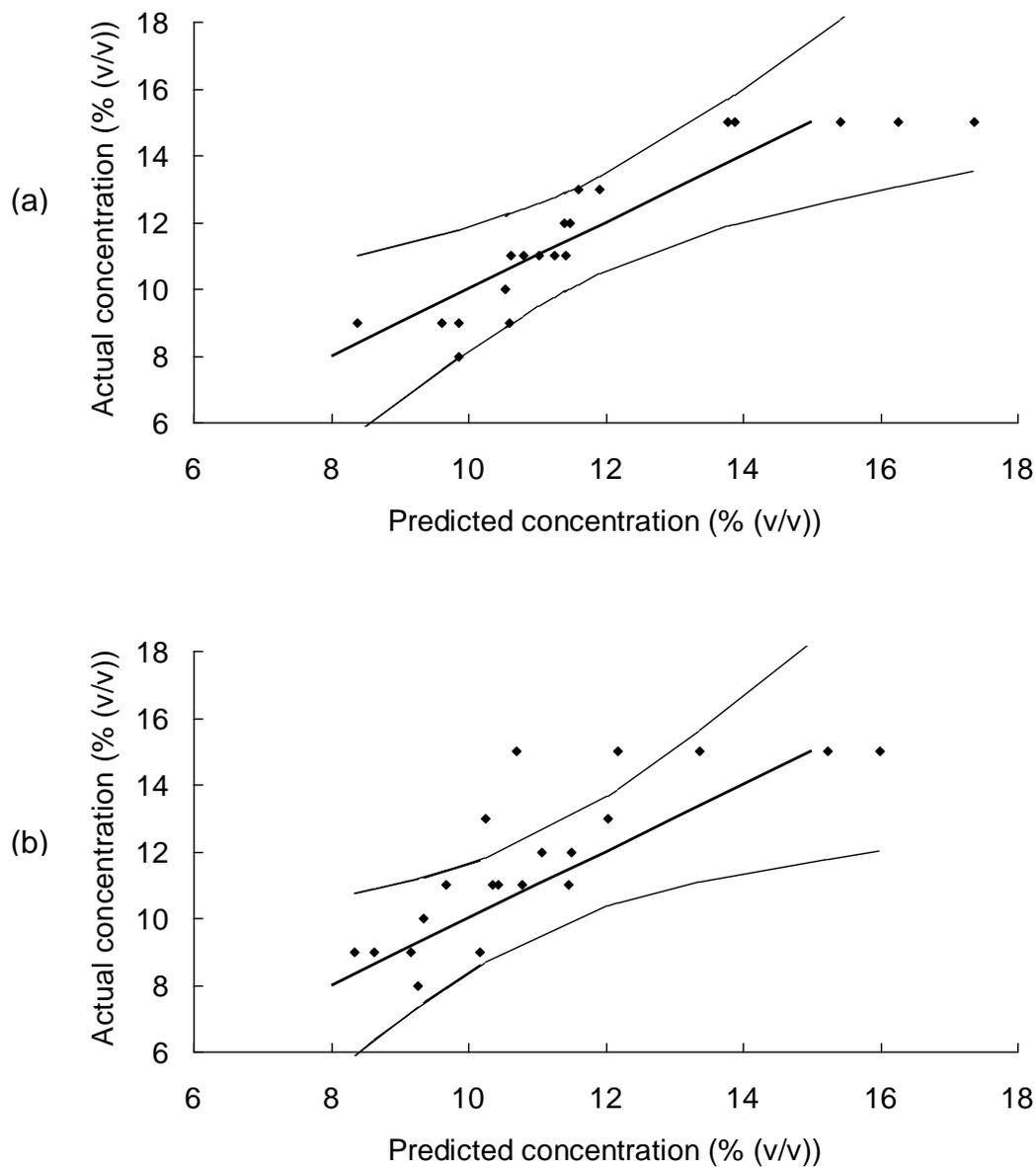


Fig. 4.23 Albumin concentrations in solutions predicted using calibration developed with (a) Raman peak height; (b) integrated peak area.

Table 4.17 Stepwise Regression coefficient statistics for peak height variables of denatured albumin gel

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
1243.35	-4.0×10^{-3}	-2.63	0.015
1448.68	4.9×10^{-3}	7.34	0
1667.78	0	0.67	0.62

Table 4.18 Stepwise Regression coefficient statistics for peak area variables of denatured albumin gel

Wavenumber (cm ⁻¹)	Coefficient	T-statistics	p-value
1243.35	-1.3×10^{-4}	3.10	5.1×10^{-3}
1448.68	7.2×10^{-5}	2.95	7.2×10^{-3}
1667.78	4.3×10^{-5}	2.24	0.035

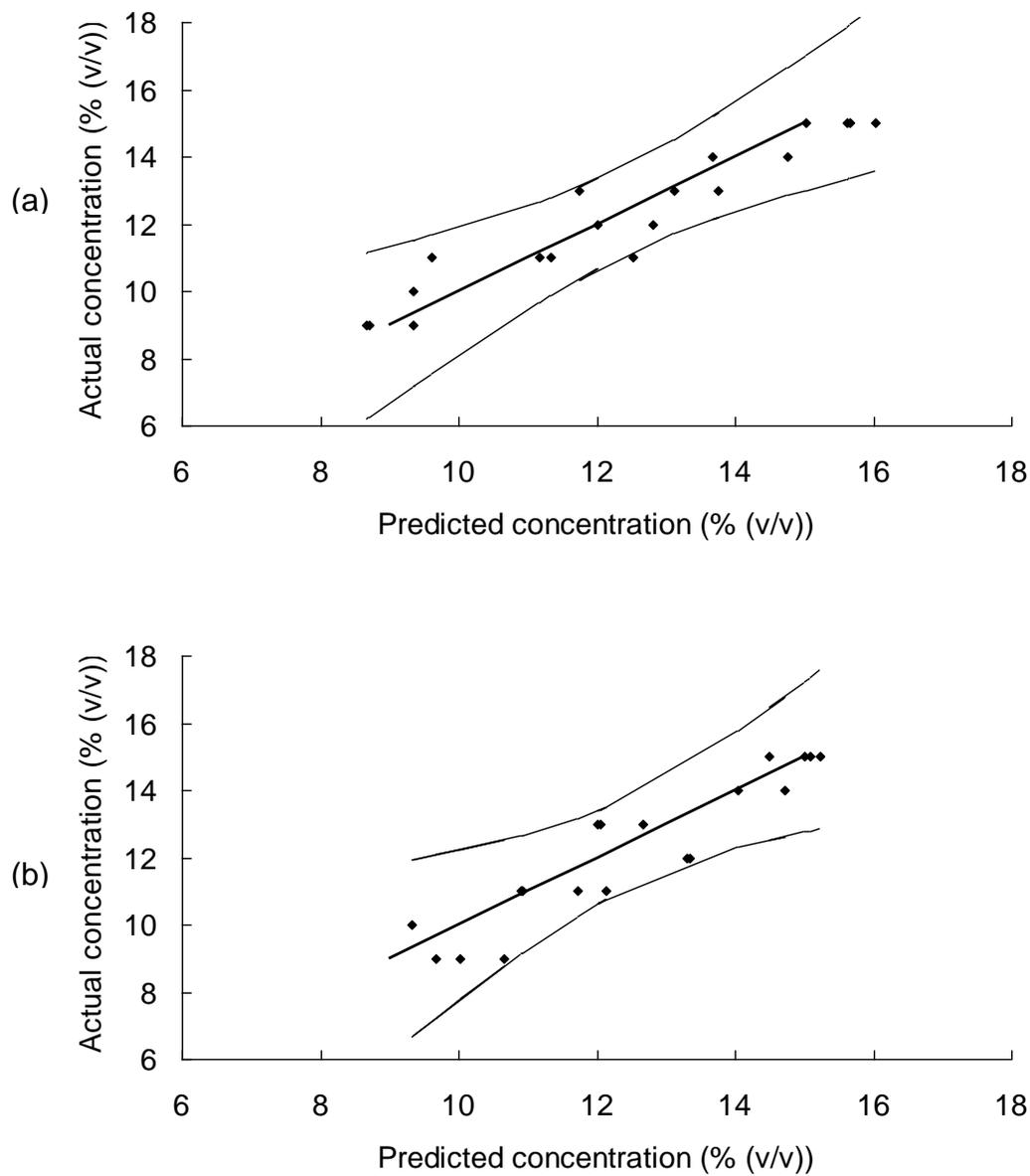


Fig. 4.24 Denatured albumin gel concentration predicted using calibration developed with (a) Raman peak height; (b) integrated peak area.

Table 4.19 Model statistics for stepwise calibrations with Raman spectra of ovalbumin aqueous solutions

	Albumin solution		Denatured albumin gel	
	Peak height	Peak area	Peak height	Peak area
R_c^2	0.79	0.77	0.76	0.76
$RMSEC(\%(w/v))$	1.28	1.44	1.30	1.34
Number of predictors/latent variables	2	2	2	3
R_t^2	0.80	0.66	0.91	0.86
$RMSEP(\%(w/v))$	1.05	1.53	0.74	0.82

4.3.2 Quantification using Near-infrared spectroscopy

The absorbance spectra of 15% (v/v) ovalbumin solution and solid ovalbumin powder were acquired using the multimodal spectrometric system (Fig. 4.25). Compared to water, which had a relatively intense absorption band around 970 nm, both solid ovalbumin powder and 15% (v/v) ovalbumin solution demonstrated featureless absorbance spectra throughout the entire spectral range. The spectrally flat spectra could be explained in two aspects. First, the lack of or rather weak absorption of the ovalbumin in this spectral region (750-980 nm) was mainly accountable for the absence of observable spectral features. Second, both the highly reflective ovalbumin powder and translucent solution did not allow sufficient optical penetration into the samples, which exacerbated the already poor absorbance situation. The second aspect could be evidenced by the absence of water absorption features in the ovalbumin solution.

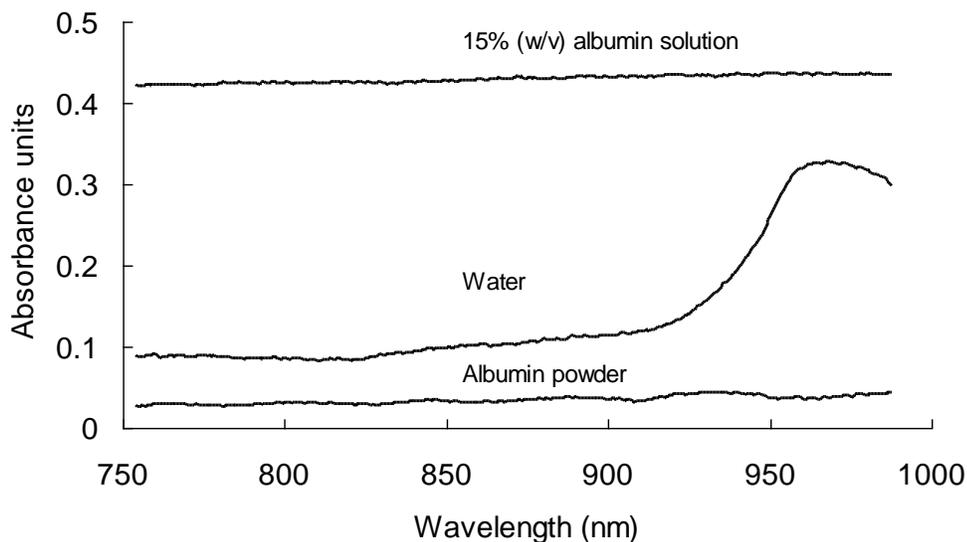


Fig. 4.25 Absorbance spectra for 15% (w/v) albumin solution, water, and solid albumin powder.

Use of PLSR could capture concentration related absorption features that were not readily observable in the unintelligible NIR spectra. Here, no features in the ovalbumin powder and ovalbumin solution were obvious. Therefore, PLSR calibrations were built to find out if quantitative analysis of ovalbumin concentration was possible. A second order 35-point Savitzky–Golay derivative filter was applied to decrease the scattering effect (Arakaki et al. 2007). Diffuse reflectance spectra of denatured ovalbumin gel were also measured and transformed into apparent absorbance, $\log(1/R)$. The apparent absorbance spectra were then preprocessed with the same Savitzky–Golay derivative filter. One outlier was found at a significance level of 0.05 and subsequently removed. The beta coefficients are shown in Fig.4.26. As shown in Fig. 4.26, the beta coefficients did not demonstrate any patterns and had sharp peaks and valleys with large numerical values. This could mean the absence of concentration correlated variables and highly unstable prediction performance. The model statistics in Table 4.20 showed that neither absorbance mode for ovalbumin solution nor diffuse reflectance mode for denatured ovalbumin gel produced any acceptable prediction model. Thus, within the 750 – 980 nm region on this instrument, NIR spectroscopy was not capable of determining the ovalbumin concentration either in aqueous solution or in denatured gel solids.

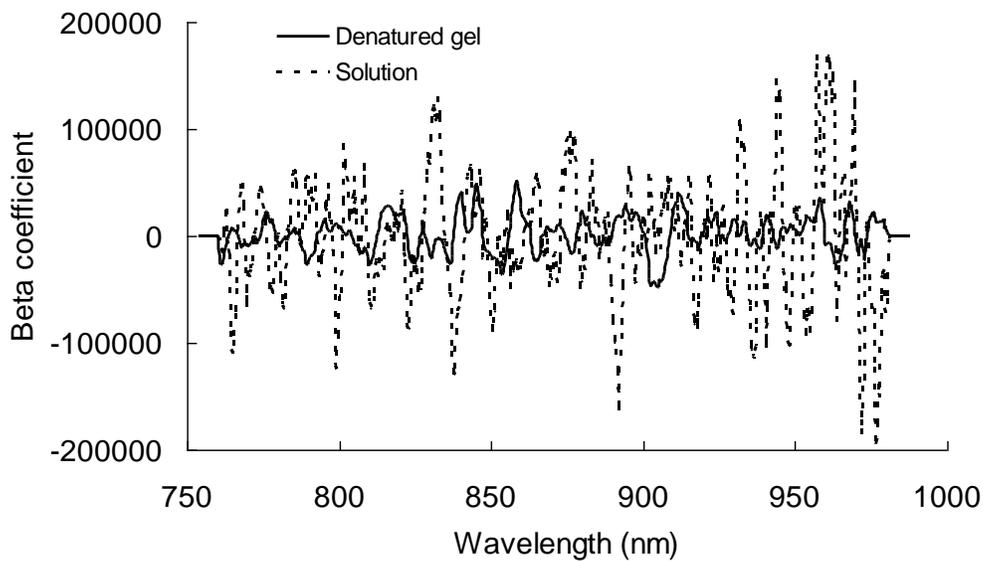


Fig. 4.26 Beta coefficients for ovalbumin solution and denatured gel.

Table 4.20 Model statistics for NIR calibrations of ovalbumin aqueous solutions

	Albumin solution	Denatured gel in glass vial
R_c^2	0.88	0.53
$RMSEC(\%(w/v))$	1.04	1.70
Number of predictors/latent variables	5	3
R_t^2	0	9.0×10^{-3}
$RMSEP(\%(w/v))$	3.30	3.35

4.3.3 Summary

It was shown that NIR spectroscopy was unable to predict ovalbumin concentration due to the lack of absorption features within the 750 nm to 980 nm spectral region (Table 20). Raman spectroscopy appeared to have the potential to quantitatively determine the ovalbumin concentration in aqueous solutions. For both albumin solution and denatured albumin gel, the wavenumber variable at 1448.68 cm^{-1} appeared to be a good predictor for ovalbumin concentration. This applied to prediction equations developed using either peak height or peak area. The wavenumber variable at 1662.66 cm^{-1} and 1667.78 cm^{-1} , which corresponded to the C=O stretching and N-H wagging modes of amide I region, was not useful for overall ovalbumin concentrations in both aqueous solution and denatured gel samples (Table 17 and 18). Between peak height and peak area at assigned wavenumbers, peak height consistently achieved better calibration and prediction statistics compared to those of peak areas (Table 19). The lowest RMSEP values were 1.05% for ovalbumin solution and 0.71% for denatured gel. When concentration reached 15%, it became harder for the ovalbumin powder to get completely dissolved in distilled water. The dissolved ovalbumin tended to precipitate after a suspension was formed and allowed to stand undisturbed for an extended period. The poor prediction results for ovalbumin solution with concentration at 15% might corroborate this fact. One other reason might be due to the thermal stability of ovalbumin protein during spectra measurement. The intense laser radiation at the focal point, which was powerful enough to burn plastics within a fraction of a second, might induce conformational transition of secondary structure due to thermal unfolding of protein in the aqueous solution. This also explained the lower prediction errors for denatured ovalbumin gels, which were in a much more stable form compared to the ovalbumin in aqueous solution.

4.4 Ovalbumin Secondary Structure Changes upon Thermal Denaturation Using Raman Spectroscopy

Heat-induced denaturation of ovalbumin was studied by heating ovalbumin solution at different concentrations at 70 °C and 90 °C for 1 h. The Raman spectra of 16% (w/v) ovalbumin solution and ovalbumin gel formed at 70 °C and 90 °C are shown in Fig. 4.27. As shown in Fig. 4.27, peaks around 1243.35 cm⁻¹ began to show after the ovalbumin solution samples were heated for 1 h at both temperatures. This spectral change in the amide III region corresponded to the formation of antiparallel β-sheet between ovalbumin molecules. The other change was a frequency shift from 1662.66 cm⁻¹ to 1667.78 cm⁻¹ for the amide I region. These findings corroborated results previously reported by Painter and Koenig (1976). Because different data acquisition parameters were used in acquiring the spectra of samples at different states (Appendix A), the Raman peaks showed varying intensities. Therefore, differentiating the Raman spectral changes due to ovalbumin denaturation was impossible by directly comparing peak heights or peak areas. The ratio between values at 1243.35 cm⁻¹ and 1662.66 cm⁻¹ were calculated for both peak height and peak area. A mixed model including a mixture of fixed and random factors was used (Verbeke and Molenberghs 1997). The effect of heating temperature and solution concentrations on the ratios was studied using the SAS MIXED procedure (PROC MIXED). Heating temperature was assigned as the fixed factor. Concentration and temperature multiplied by concentration (temperature*concentration) were assigned as the random factor (Green and Tukey, 1960). There were three levels in temperature factor, i.e. 25 °C (room temperature), 70 °C, and 90 °C. A total of 9 levels were assigned to concentration, i.e. 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, and 16%. Results produced by SAS are shown in Table 4.21– 4.24.

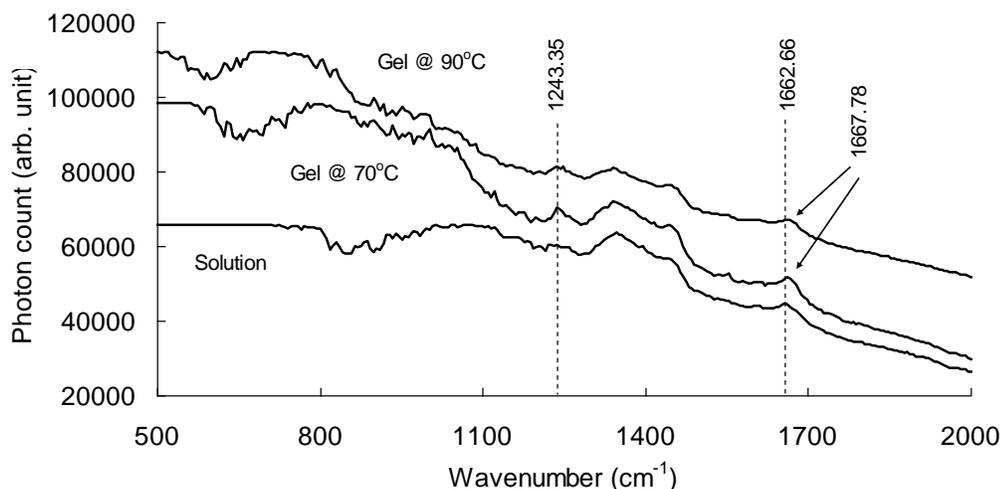


Fig. 4.27 Raman spectra of 16% (w/v) ovalbumin solution and ovalbumin gels formed after heating at 70 °C and 90 °C for 1 h.

Table 4.21 Type 3 Analysis of Variance (peak height)

Source	Error Term	Error DF	F Value	Pr > F
Temperature	MS(Temperature*Concentration)	16	21.32	<.0001
Concentration	MS(Temperature*Concentration)	16	0.16	0.9930
Temperature*Concentration	MS(Residual)	54	10.80	<.0001

Table 4.22 Type 3 Analysis of Variance (peak area)

Source	Error Term	Error DF	F Value	Pr > F
Temperature	MS(Temperature*Concentration)	16	6.73	<.0076
Concentration	MS(Temperature*Concentration)	16	0.82	0.5941
Temperature*Concentration	MS(Residual)	54	13.34	<.0001

Table 4.23 Differences of Least Squares Means (peak height)

Effect	Trial	Trial	Estimate	Standard Error	DF	T Value	Pr> t
Temperature	25	90	-0.2047	0.03646	16	-5.61	<.0001
Temperature	25	70	-0.2047	0.03646	16	-5.70	<.0001
Temperature	90	70	-0.00297	0.03646	16	-0.08	0.9360

Table 4.24 Differences of Least Squares Means (peak area)

Effect	Trial	Trial	Estimate	Standard Error	DF	T Value	Pr> t
Temperature	25	90	-0.08667	0.02538	16	-3.41	0.0035
Temperature	25	70	-0.01385	0.02538	16	-0.55	0.5929
Temperature	90	70	-0.00297	0.02538	16	2.87	0.0111

The F-tests of fixed effects factor (temperature) and random effect factor (concentration) are given in Tables 4.21 and 4.22. It is obvious that temperature effects were significantly different from one another for both peak height ratio and peak area ratio. On the other hand, concentrations were not significantly different for both peak height ratio and peak area ratio. Thus, the denaturation of ovalbumin was found to be independent of its concentration in aqueous solution. This result corroborated with the findings of Weijers et al. (2003).

The temperature effect on peak height ratio was significantly different between 25 °C and 70 °C and between 25 °C and 90 °C (Table 4.23). No significant difference

was found between 90 °C and 70 °C for peak height ratio. Table 4.24 shows that there was a significant difference between 25 °C and 90 °C as well as between 70 °C and 90 °C for peak area ratio but no significant difference was found between 25 °C and 70 °C. Therefore, we can see that both peak height ratio and peak area ratio were significantly different between 25 °C and 90 °C. The temperature effect at 70 °C was significantly different from 90 °C for peak area ratio and significantly different from 25 °C for peak height ratio. A closer examination of Raman spectra shows that spectral peaks at 1243.35 cm^{-1} are sharp at 70 °C heating compared to that at 90 °C heating. This might be explained by the fact that ovalbumin, like other types of protein, did not gel properly at 70 °C though it started to undergo conformational changes (Nonaka et al. 1993). The ovalbumin gel formed at 70 °C was very tender and contained significant amount of fluid when examined physically. On the other hand, ovalbumin gels formed at 90 °C were firm and were almost devoid of any watery content. The significant difference between 70 °C and 90 °C effect for peak area ratio might explain the fact that only a small portion of protein globules underwent heat-induced conformational changes upon heating at 70 °C for extended period (1 h). More extensive region of anti-parallel β -sheet structure was formed at heating temperature of 90 °C as was evidenced by the broadening of the spectral region around 1243.35 cm^{-1} .

4.5 Data Fusion

Data fusion was performed on Raman and NIR data of ethanol solution. Previous discussion in section 4.3.2 found that ovalbumin concentration was unable to be determined in the available spectral region in this study. Therefore, data fusion on spectra of ovalbumin was not further discussed. Raman and NIR absorption tended to have

different sources of interference, i.e., light scattering effect for NIR and fluorescence for Raman. Raman spectra and NIR absorbance spectra were acquired on different scales (2.8×10^4 – 1.6×10^5 photon counts for Raman and 0.04–0.20 absorbance units for NIR absorbance) which made rescaling necessary. Because the CCD sensor was saturated due to intense Rayleigh scattering below 724.22 cm^{-1} , Raman data for wavenumbers lower than 724.22 cm^{-1} were truncated from the original data before performing auto-scaling. Similarly, the absorbance spectra were auto-scaled. To concatenate the spectra, the higher wavenumber end of Raman spectra was linked to the lower wavelength end of corresponding NIR absorbance spectra. It was noted that there existed a numerical mismatch at the ‘joint’ resulting in an abrupt edge. The average of the difference between autoscaled Raman and NIR spectral values was subtracted from the NIR spectra to offset the mismatch. The fused spectra are shown in Fig. 4.28.

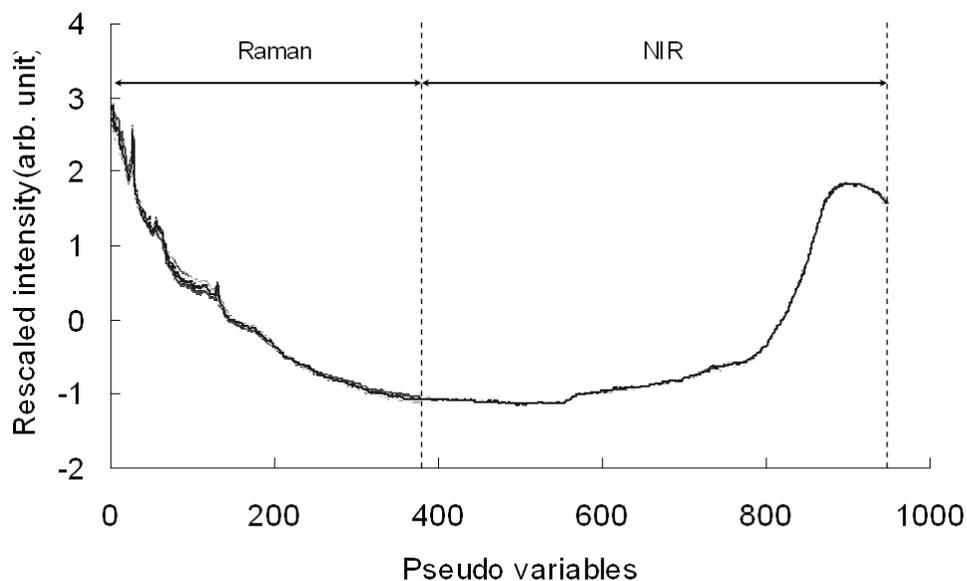


Fig.4.28 Fused spectra of Raman and NIR absorbance data for ethanol aqueous solution

A PLSR model was calibrated on the fused spectra. Leave-one-out cross validation was employed to assess the model generalization performance and find the optimal model complexity. A second order 19-points Savitzky-Golay derivative filter was applied to the fused data to remove sloped baseline. The predicted residual sum of squares values versus latent variable numbers is shown in Fig. 4.29. After including the first six latent variables in the model, the prediction performance did not improve significantly by adding more latent variables. Therefore, six factors were recommended by the program as the optimal model complexity. The beta coefficients are shown in Fig. 4.30.

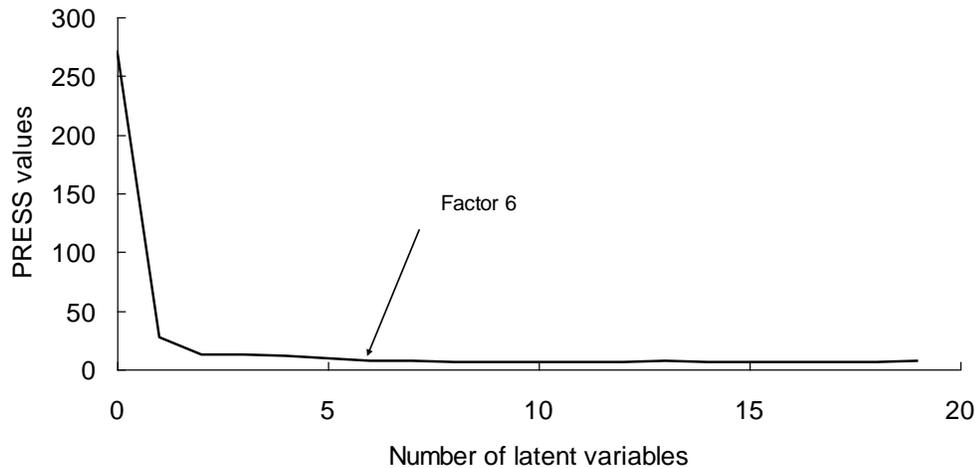


Fig. 4.29 Predicted residual sum of squares values versus factor number.

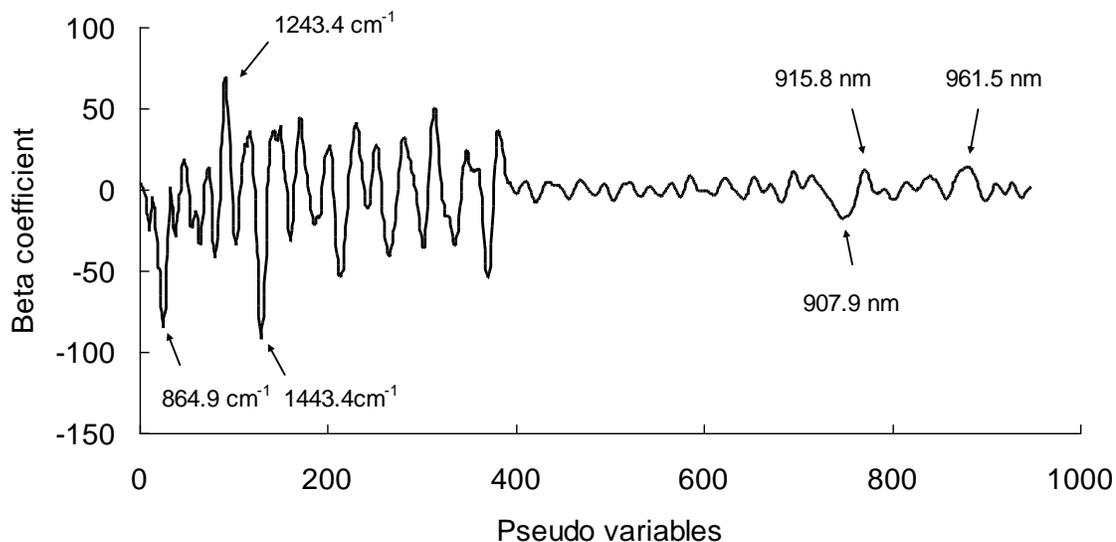


Fig. 4.30 Beta coefficients for fused Raman and NIR spectra

The first 376 pseudo variables corresponded to wavenumber variables in the Raman spectra and the rest of the pseudo variables belonged to the wavelength variables in NIR absorption. As seen in the figure, there were several wavenumber variables that had influence on the prediction model due to the noisy Raman data. But the wavenumber variables at 864.9 cm^{-1} , 1243.4 cm^{-1} , and 1443.4 cm^{-1} demonstrated the three largest weights. It was reasonable because they related to the Raman peaks of ethanol at 880 cm^{-1} , 1280 cm^{-1} , and 1460 cm^{-1} , though there were frequency shifts due to derivative preprocessing of the spectra. The beta coefficient of the NIR absorbance data also demonstrated three significant peaks at 907.9 nm , 915.8 nm , and 961.5 nm , which also corresponded well with ethanol and water absorption bands in the short wavelength NIR region. Other than these prominent peaks, there were some other variables with higher beta coefficient values in the Raman data. These variables could be attributed to the high level of noise in the spectra region, which might lead to chance correlation with the concentration variation. This type of chance correlation could cause degradation in model

performance. This was evidenced by the prediction performance of fused data model, which did not compare favorably with those of models derived from NIR absorbance or Raman peaks respectively. The prediction result is shown in Fig. 4.31. Model statistics are summarized in Table 4.25.

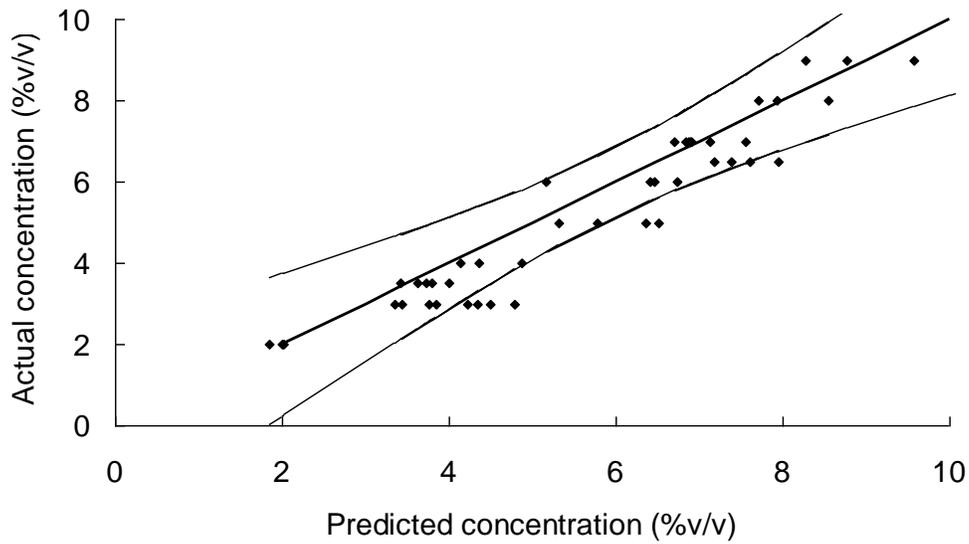


Fig. 4.31 Prediction results for fused Raman and NIR data.

Table 4.25 Model statistics for fused NIR/Raman data

R_c^2	$RMSEC$ (% (v/v))	Number of predictors/latent variables	R_t^2	$RMSEP$ (% (v/v))
0.97	0.49	6	0.93	0.73

4.6 Two-dimensional Correlation Spectroscopy

Two-dimensional correlation spectroscopy can reveal spectral information that is not readily observable in one-dimensional spectra. In this section, 2D correlation spectroscopy was applied to spectral data of ethanol and ovalbumin aqueous solutions. Raman/NIR heterospectral correlation spectroscopy was also attempted to correlate the spectral changes in NIR absorption to Raman peak intensity variations.

4.6.1 Analysis of ovalbumin and ethanol solution using 2D correlation spectra

The basic concept of 2D correlation spectroscopy was to observe the spectral intensity variation patterns induced by external perturbation. The variation patterns were expressed as 2D correlation spectra along two separate spectral variables within specific intervals. Synchronous and asynchronous spectra were generated after 2D correlation spectroscopy was applied to the test data. In this experiment, concentration was introduced as the external perturbation to the samples. In Fig. 4.32, the synchronous, asynchronous, and autopower spectrum of ovalbumin solution and denatured gels are shown.

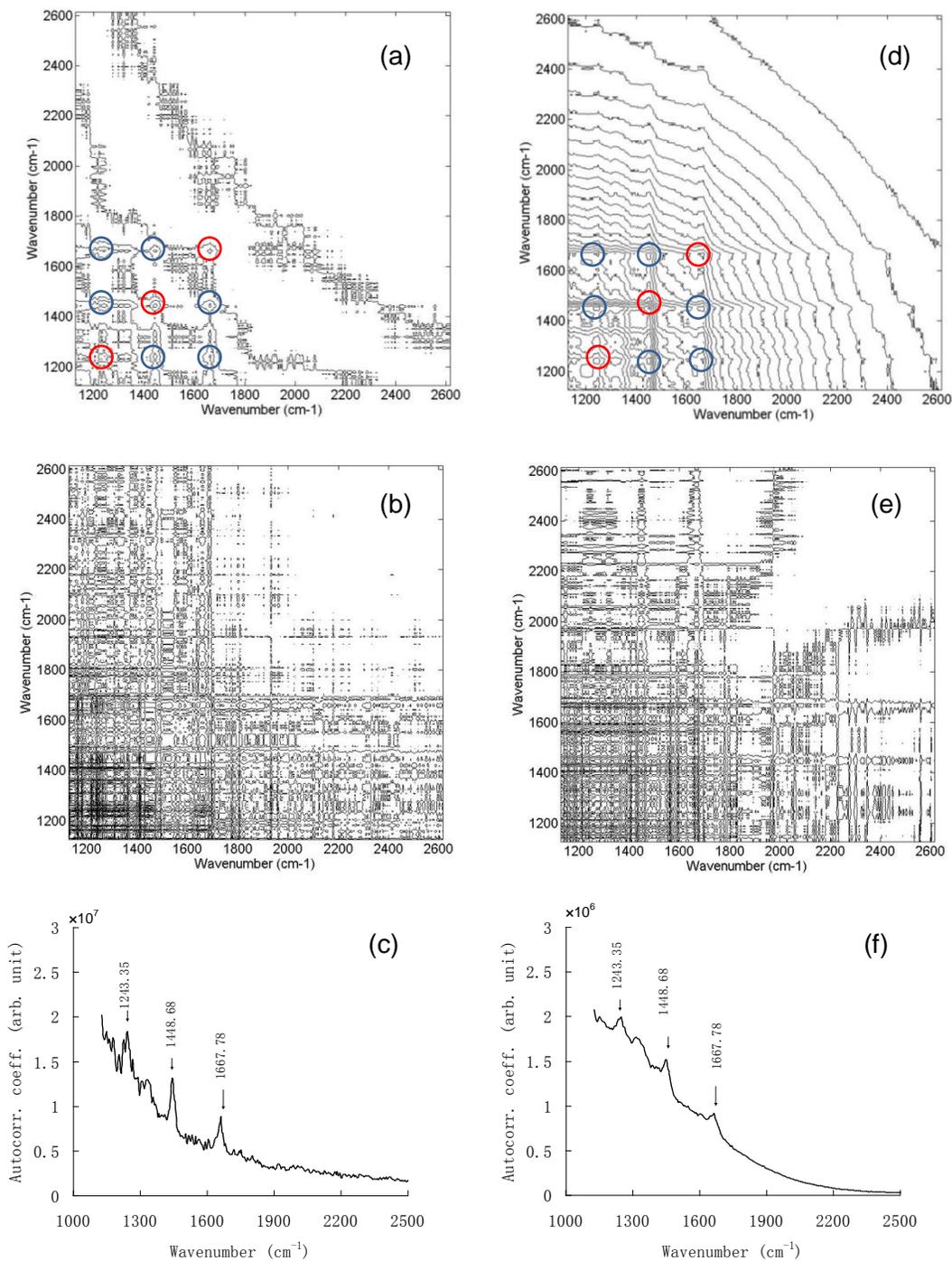


Fig. 4.32 2D correlation spectra of ovalbumin. (a) synchronous, (b) asynchronous, and (c) autopower spectra for ovalbumin aqueous solution; (d) synchronous, (e) asynchronous, and (c) autopower spectra of denatured ovalbumin gel.

The contour map shown in Fig. 4.32 depicts the 2D correlation spectra of ovalbumin in the spectral region between 1126.99 cm^{-1} and 2614.75 cm^{-1} due to the varying ovalbumin concentration. The streaking and ridges in the 2D correlation spectra were mainly due to a high level of baseline variation in Raman spectra (Ashton et al. 2006). In the synchronous spectra (Fig. 4.32(a) & (d)), two types of correlation peaks, viz., autopeaks and crosspeaks, were revealed. The autopeaks located along the diagonal positions (marked with red circle) corresponded to the autocorrelation function of spectral intensities when concentration was varied. The autopower spectrum, which was extracted along the diagonal of the synchronous spectra, revealed the autopeaks at 1243.35 , 1448.68 , and 1662.66 cm^{-1} in Fig. 4.32 (c). These peaks corresponded well to the Amide III, C–H bending, and Amide I modes found in Raman spectra of ovalbumin. Figure 4.32(f) showed autopeaks at the same positions except 1662.66 cm^{-1} , which appeared at 1667.78 cm^{-1} , indicating the frequency shift due to thermal conformational changes in ovalbumin. All crosspeaks, which appeared at off-diagonal positions, possessed positive signs, indicated that all concentration changes induced spectral intensity changes increased or decreased together.

Similarly, the 2D correlation spectra for intensity changes in the Raman and NIR spectra of ethanol in aqueous solution induced by concentration changes are shown in Fig. 4.33.

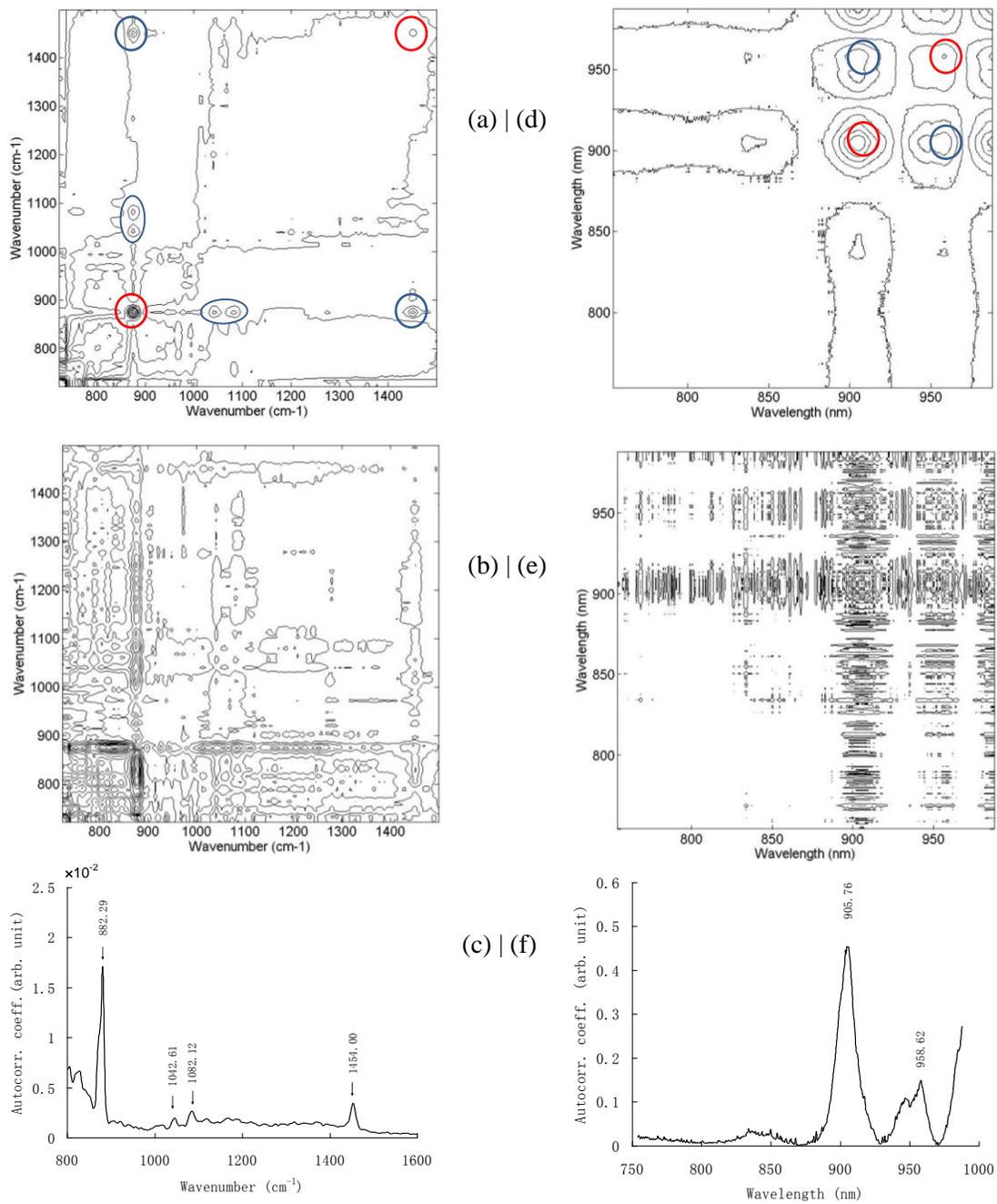


Fig. 4.33 2D correlation spectra of ethanol in aqueous solution. (a) synchronous, (b) asynchronous, and (c) autopower spectra for Raman measurement; (d) synchronous, (e) asynchronous, and (f) autopower spectra NIR absorption.

In Fig. 4.33(c), four autopeaks at 882.29, 1042.61, 1082.16, and 1454.00 cm^{-1} are observable. They corresponded well to the Raman peaks of ethanol in aqueous solution (Fig. 4.13). The two autopeaks at 1042.61 and 1082.16 cm^{-1} are not visible in Fig. 4.33(a) due to its relatively low amplitude. However, the two autopeaks (marked with red circles) along the diagonal and three crosspeaks (marked with blue circles) along shown in Fig. 4.33(a), corresponds well to the autopeaks identified in the autopower spectrum. The crosspeaks all had positive signs, indicating all spectral intensities decreased or increased simultaneously. The autospectra in Fig. 4.33(f) shows that there are two prominent autopeaks corresponding to spectral intensity changes due to concentration variation. Two autopeaks appeared at 905.76 and 958.62 nm, which corresponded well to the third overtone of the C–H stretch on the methyl group and the combination transition of O–H stretch. The synchronous 2D correlation spectrum for the NIR data of ethanol shown in Fig. 4.33(d) shows crosspeaks at positions where 905.76 and 958.62 nm intersects. These crosspeaks had negative signs, indicating the spectral intensities at 905.76 and 958.62 nm varied in different directions. This was reasonable because each peak represented different constituents in the binary solution.

4.6.2 Raman/NIR 2D heterospectral correlation spectral analysis

For the two closely spectroscopic techniques, viz., Raman and NIR, attempt was made to assign spectral bands in the highly overlapping NIR spectra with the assistance of well-resolved Raman features. In the NIR spectrometric experiment, respective absorbance features of ethanol and water in their binary mixtures were hard to interpret in the one-dimensional spectra. By correlating the NIR absorbance spectra with the better understood Raman spectra under the same external perturbation, assignments of NIR

spectral bands to its molecular vibrational origin could be made. The truncated spectral region for Raman spectra in the heterospectral correlation analysis was between 724.22 cm^{-1} and 1600.85 cm^{-1} . The wavelength range for NIR spectra was between 753.91 nm and 987.74 nm. The 2D correlation spectra and slice spectra are shown in Fig. 4.34.

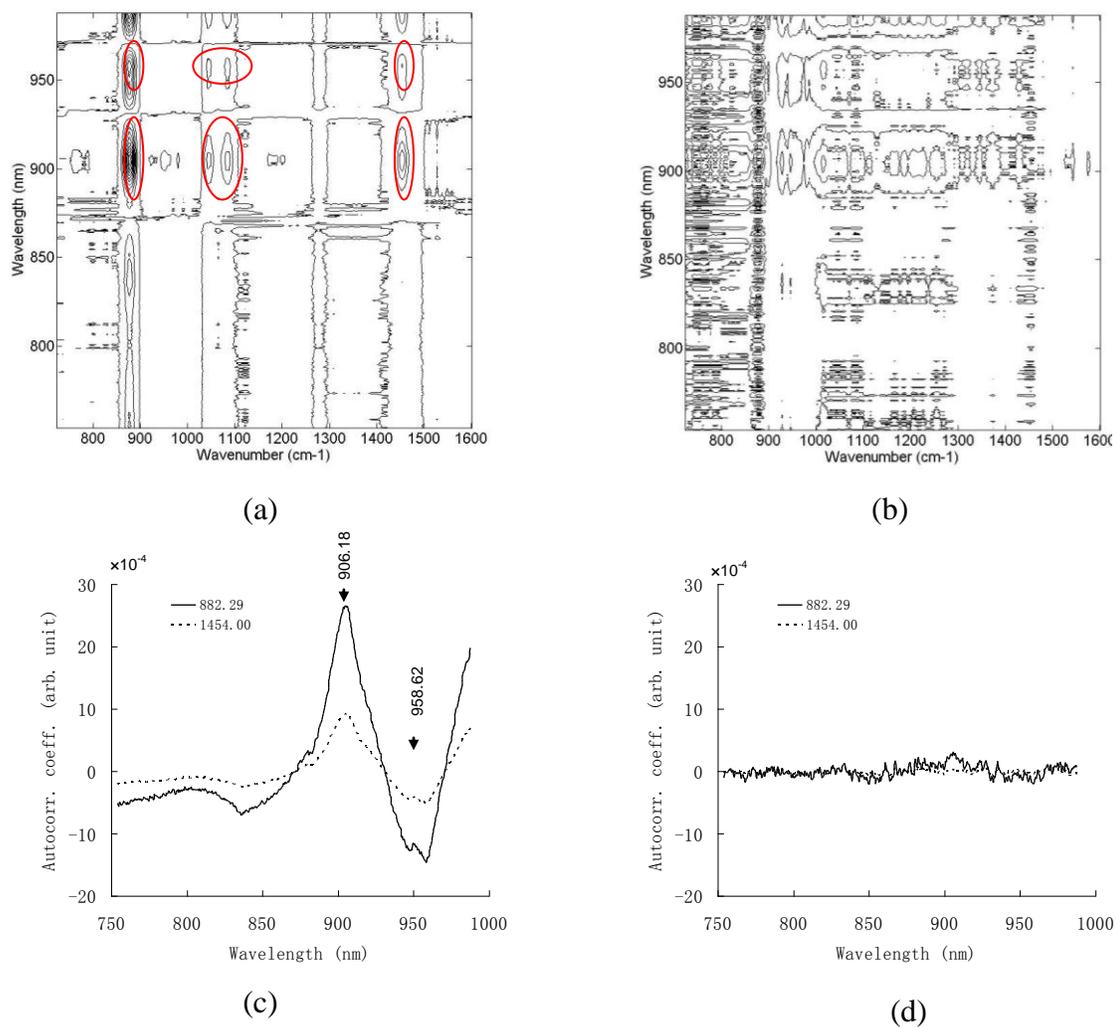


Fig. 4.34 2D Raman/NIR heterospectral correlation spectra generated from Raman and NIR spectra of ethanol aqueous solution. (a) synchronous spectra (b) asynchronous spectra (c) slice spectra from synchronous spectra at 882.29 cm^{-1} and 1454.00 cm^{-1} (d) slice spectra from asynchronous spectra at 882.29 cm^{-1} and 1454.00 cm^{-1} .

In Fig. 4.34(c), the slice spectra of Raman spectra at 882.29 cm^{-1} and 1454.00 cm^{-1} revealed that there were two NIR absorbance peaks. A positive crosspeaks in the synchronous spectra of Raman/NIR 2D correlation spectra meant the two spectral bands in NIR and Raman shared the same chemical origin. In this scenario, the spectral intensities of Raman peaks at 882.29 cm^{-1} and 1454.00 cm^{-1} and the NIR absorbance feature around 906.18 nm changed in phase with the varying concentration. This corroborated with prior knowledge of the Raman peaks and absorbance wavelength for ethanol. On the other hand, NIR feature around 958.62 nm shared negative crosspeaks with the spectral intensities of Raman peaks at 882.29 cm^{-1} and 1454.00 cm^{-1} . Absorbance at 958.62 nm corresponded to the absorbance of water and thus was of different origin with Raman intensity variations at 882.29 cm^{-1} and 1454.00 cm^{-1} . The lack of structures in the slice spectra of asynchronous spectra indicated the absence of out-of-phase intensity changes of Raman/NIR with concentration.

4.6.3 Summary

Using concentration as the external perturbation, 2D correlation spectroscopy identified spectral signatures for both ovalbumin and ethanol. The synchronous spectra for ovalbumin revealed autopeaks at 1243.35 , 1448.68 , 1662.66 , and 1667.78 cm^{-1} . These spectral bands corresponded well with the Raman peaks of ovalbumin in aqueous solution and denatured gels respectively. Autopeaks at 905.76 and 958.62 nm from synchronous spectra of NIR spectra of ethanol aqueous solution matched the absorption bands for ethanol and water respectively. Heterospectral correlation analysis found that Raman peaks at 882.29 cm^{-1} and 1454.00 cm^{-1} and NIR absorbance at 958.62 nm shared negative crosspeaks, which indicated different origins of molecular vibrations.

5. CONCLUSIONS

A multimodal spectrometric system was designed and applied to the analysis of ethanol and ovalbumin. The unique spectrometric system combined the capacities to acquire Raman spectra as well as NIR absorption data in one compact device. This invention presented scientists and engineers with a powerful tool to analyze samples in various states and of varying constitutional composition. The device consists of fixed optics for a rugged design suitable for industrial applications. Fiber optical probes were utilized which allow remote sensing of samples. The unique system also has the advantages of multiplexing, i.e, simultaneous acquisition of data in multiple tracks. It also has the capacity to carry out multimodal spectroscopy and two-dimensional correlation spectroscopy, which are two advanced techniques that are becoming increasingly popular in the spectroscopy community. After preliminary assembly, calibration was carried out to optimize optical alignment and spectral resolution. Relative Raman shift standards, i.e., benzonitrile and 4-acetamidophenol, were used to calibrate the wavenumbers versus pixel relationship. Second order polynomial fit was found to provide the optimal model complexity judging by the variance. Variance values of wavenumber calibration developed using 4-acetamidophenol were lower than that of benzonitrile. Assignment of wavelength variables to pixels was carried out using the developed wavenumber polynomial equation and validated with a standard argon calibration lamp. A maximum of 0.57 nm wavelength error was found. The available spectral region was calculated to be from 752.36 nm to 987.74 nm. Spectral resolution was calculated by fitting Gaussian profiles to individual spectral peaks and to calculate the full width at half maximum

(FWHM). Spectral resolution was found to be lower than 1 nm in wavelength and better than 14.39 cm^{-1} in wavenumbers. Spectral response linearity was tested using reflectance standards at 2%, 25%, 75%, and 99%. Linearity expressed as coefficient of determination (R^2) was 0.998 across all tested wavelengths. Instrumental stability was tested and found to have an average of 0.72% for the Raman channel and 0.39% for the NIR channel during several hours of testing.

The multimodal spectrometric system was applied to study the hydrogen bonding effect of ethanol in water and quantitative analysis of ethanol in aqueous solution. A frequency shift for the C–O stretch at 1053.92 cm^{-1} of anhydrous ethanol sample to 1042.61 cm^{-1} in the 10% (v/v) ethanol aqueous solution was observed. The change of rocking band of CH_3 from 1093.36 cm^{-1} in anhydrous ethanol to 1082.16 cm^{-1} was also reported. Quantitative determination of ethanol concentration in aqueous solution was performed using univariate regression, stepwise multiple regression, and partial least squares regression. Baseline corrected spectral peak height at 882.16, 1042.61, 1082.16, and 1454.00 cm^{-1} and peak areas around these peaks were used as predictors. It was found that univariate calibration using peak height at 882.16 cm^{-1} produced the most succinct model and best prediction results with a root mean squared error of prediction (RMSEP) of 0.45% (v/v) and a R^2 of 0.96. The NIR absorption spectra were acquired on the same ethanol aqueous solution and PLSR models were developed. Beta coefficients of the developed models agreed with the absorption peaks for pure ethanol and water. Preprocessing using mean-centering followed by 5-point first-order Savitzky–Golay derivative produced the best prediction model with a RMSEP of 0.49% (v/v) and a R^2 of 0.96. Quantification of ovalbumin concentration in aqueous solution as well as in

denatured gels heated at 90 °C was explored using Raman spectroscopy. Stepwise multiple regression was performed using variables at 1243.35, 1448.68, 1662.66, and 1667.78 cm^{-1} . Peak height was a good predictor compared to peak area and 1448.68 cm^{-1} was chosen to be included in all calibrated models by stepwise multiple regression. Prediction errors were 1.05 % (w/v) for ovalbumin solution and 0.74 % (w/v) for ovalbumin gel. Near-infrared spectroscopy failed to quantify ovalbumin concentration due to the lack of absorption features in the available spectral region. Secondary structural changes in ovalbumin upon thermal denaturation was studied using the peak height and peak area ratios between 1243.35 cm^{-1} and 1662.66 cm^{-1} . Temperature effects on the conformational changes of ovalbumin were analyzed using mixed procedure in SAS 9.2. It was reported that temperature effects were significant for both peak height and peak area ratio. Concentration effect was not significant on the other hand for both predictors. Temperature effect on peak height ratio explained the onset of thermal denaturation of ovalbumin at 70 °C and temperature effect on peak area ratio revealed more extensive formation of anti-parallel β -sheet structure at 90 °C.

Data fusion of Raman and NIR absorption spectra was attempted. The important variables corresponded well with the spectral signatures of ethanol in both Raman and NIR spectra. However, fusion of spectral data resulted in a more complicated PLSR model and higher prediction error. Two-dimensional (2D) correlation spectra were calculated on the ovalbumin and ethanol aqueous solution with concentration as the external perturbation. The autopower spectra revealed Raman peaks and NIR absorption features that corresponded well to the analytes. This was especially useful in interpreting NIR absorbance spectra, in which absorbance bands for ethanol and water were otherwise

indiscernible in one-dimensional spectra. Heterospectral correlation analysis was carried out on Raman/NIR spectra of aqueous solution of ethanol with concentration. Raman peaks at 882.29 cm^{-1} and 1454.00 cm^{-1} corresponded well with the NIR absorption feature around 906.18 nm . The negative sign of crosspeak shared by NIR absorbance at 958.62 nm and Raman intensity variations at 882.29 cm^{-1} and 1454.00 cm^{-1} indicated the spectral intensity variations were of different origin, i.e., water for NIR absorption and ethanol for Raman scattering.

In summary, a multimodal spectrometer capable of measuring Raman scattering and NIR absorption spectra was successfully tested and calibrated. The prototype device was applied to study ethanol and ovalbumin in aqueous solutions. Research findings on ethanol and ovalbumin aqueous solutions proved its potential as a rapid analyzing tool for both quantitative and qualitative analysis. Unique capability of the multimodal spectrometer to carry out advanced spectroscopic techniques, viz., hyphenated Raman/NIR technique and 2D correlation spectroscopy, was also successfully demonstrated.

6. RECOMMENDATIONS

The prototype device demonstrated unique features. Particularly, it allowed for multimodal spectroscopic measurement on one single system. It opened up possibilities such as heterospectral two-dimensional correlation spectroscopy. To enhance its performance, there were several aspects in the hardware system that needed further improvement.

First, a camera with more pixels would be essential to improve the wavenumber accuracy and spectral resolution. The current system had 576 pixels along the spectral dimension. Spectral intervals ranging from 4.22 cm^{-1} to 6.85 cm^{-1} were achieved between adjacent pixel columns. Spectra resolution was better than 14.39 cm^{-1} . For Raman spectroscopy, this kind of accuracy and resolution might offer reasonable estimations of Raman peak positions in identifying chemical compounds and functional groups. It became the subject of concerns when frequency shift or quantitative analysis was subject of study. In these situations, minute changes in frequency might go undetected and it was also expected to lose some accuracy in determining the spectral intensity.

Second, an available spectral range in the longer wavelength range was preferred. As is shown in the experiments in this study, the spectral range between 750 nm and 988 nm offered by CCD camera was not rich in absorption spectral features. This greatly hindered the applicability of the NIR absorption channel, which offered prediction accuracy comparable to that of Raman within a fraction of second. Commercially available Indium gallium arsenide (InGaAs) array detectors had spectral response range of $0.9\text{ }\mu\text{m}$ to $1.67\text{ }\mu\text{m}$ and $0.9\text{ }\mu\text{m}$ to $2.55\text{ }\mu\text{m}$. These spectral ranges practically covered

the first and second overtone and combination bands of the C-H, O-H, and N-H bonds. In addition, it offered 10 to 100 times increase in the relative intensity of signal compared to the third overtone bands within the 750 nm and 988 nm range. Moving the detection range into the longer wavelength region also meant significant reduction in the fluorescence background for the Raman channel. This reduction in fluorescence was especially helpful when dealing with biological or real life samples, which tended to fluoresce strongly. To enable the Raman channel, an NIR laser source emitting at longer wavelength, e.g., 1064 nm, was required.

Third, a better integrated sampling probe and automated data acquisition software would cut down measurement time and move a step further towards simultaneous measurements for both NIR and Raman channel. If the optical paths for NIR and Raman could be placed orthogonal to each other at the sampling point, exactly the same sample would be measured instead of taking measurements of duplicates for each channel. Nevertheless, programmable shutters were needed to control the sequence of NIR and Raman excitation to avoid cross talks between the two channels. A dedicated software platform could control the data acquisition sequence to enable multimodal measurement. At this point, NIR and Raman channels worked in sequence. The NIR channel only took a fraction of second to acquire the spectra and thus the dual channel measurement could be considered simultaneous. This could possibly make it a highly useful tool to study the kinetics of chemical reactions using Raman and NIR jointly.

REFERENCES

- AACC. 2009. *AACC International Approved Methods of Analysis*. St. Paul, MN: AACC.
- Anderssen, R.S., E. Carter, B.G., Osborne, and I.J. Wesley. 2003. Joint inversion of multi-modal spectroscopic data of wheat flours. *Appl. Spectrosc.* 59(7): 920-925.
- Arakaki, L.S.L., D.H. Burns and M.J. Kushmerick. 2007. Accurate myoglobin oxygen saturation by optical spectroscopy measured in blood-perfused rat muscle *Appl. Spectrosc.* 61(9): 978–85.
- Asfeth, N.K., V.H. Segtnan, B.J. Marquardt and J.P. Wold. 2005. Raman and near-infrared spectroscopy for quantification of fat composition in a complex food model system. *Appl. Spectrosc.* 59: 1324-1332.
- Ashton, L., B. Czarnik-Matuszewicz and E.W. Blanch. 2006. Application of two-dimensional correlation analysis to Raman optical activity. *J. Mol. Struct.* 799: 61–71.
- Baranska, M., W. Schütze and H. Schulz. 2006. Determination of Lycopene and β -Carotene Content in Tomato Fruits and Related Products: Comparison of FT-Raman, ATR-IR, and NIR Spectroscopy. *Anal. Chem.* 78: 8459-8461.
- Barnes, R.J., M.S. Dhanoa and S.J. Lister. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43(5): 772-777.
- Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

- Berg, R.W. and T. Nørbygaard. 2006. Wavenumber calibration of CCD detector Raman spectrometers controlled by a sinus arm drive. *Appl. Spectrosc. Rev.* 41: 165–183.
- Blanco, M. and I. Villarroya. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trends Anal. Chem.* 21(4): 240-250.
- Bruun, S.W. I. Søndergaard and S. Jacobsen. 2007a. Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. gluten powder. *J. Agric. Food Chem.* 55: 7234-7243.
- Bruun, S.W. I. Søndergaard and S. Jacobsen. 2007b. Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. hydrated gluten. *J. Agric. Food Chem.* 55: 7244-7251.
- Burikov, S. T. Dolenko, S. Patsaeva, Y. Starokurov and V. Yuzhakov. 2010. Raman and IR spectroscopy research on hydrogen bonding in water–ethanol systems. *Mol. Phys.* 108(18): 2427-2436.
- Carey, D.M. and G.M. Korenowski. 1998. Measurement of the Raman spectrum of liquid water. *J. Chem. Phys.* 108: 2669-2675.
- Cavinato, A.G., D.M. Mayes, Z. Ge and J.B. Callis. 1990. Noninvasive method for Monitoring ethanol in fermentation processes using fiber-optic near- infrared spectroscopy. *Anal. Chem.* 62: 1977-1982.
- Chang, S.K.C. Chang, K. C. 2010. Chapter 9. Protein Analysis. In *Food Analysis*, 4th edition, ed. S. S. Nielsen, 133-146. New York, NY: Springer.

- Chung, H. and M. Ku. 2000. Comparison of Near-Infrared, Infrared, and Raman Spectroscopy for the Analysis of Heavy Petroleum Products. *Appl. Spectrosc.* 54(2): 239-245.
- Chalmers, J.M. and P. Griffiths. 2002. *Handbook of Vibrational Spectroscopy*. New York, NY: John Wiley and Sons.
- Cîntă-Pînzaru, S., S. Cavalu, N. Leopold, R. Petry and W. Kiefer. Raman and surface-enhanced Raman spectroscopy of tempyo spin labelled ovalbumin. *J. Mol. Struct.* 565-566: 225-229.
- Clarke, F.C., M.J. Jamieson, D.A. Clark, S.V. Hammond, R.D. Jee, and A.C. Moffat. 2001. Chemical Image Fusion. The Synergy of FT-NIR and Raman mapping microscopy to enable a more complete visualization of pharmaceutical formulations. *Anal. Chem.* 73(10): 2213-2220.
- Colthup, N.B., L.H. daly and S.E. Wiberley. 1975. Vibrational and rotational spectra. In *Introduction to Infrared and Raman Spectroscopy*, ed. N.B. Colthup, L.H. daly and S.E. Wiberley, 1-68. New York, NY: Academic Press, Inc.
- Culbertson, J.D. 2006. Chapter 7 Food Protein Functionality. In *Handbook of Food Science, Technology, And Engineering, Volume 1*, ed. Y.H. Hui. Boca Raton, FL: Taylor & Francis LLC.
- De Beer, T., A. Burggraeve, M. Fonteyne, L. Saerens, J.P. Remon and C. Vervaet. 2011. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.* 417: (2011) 32– 47.

- De Groot, P.J., H. Swierenga, G.J. Postma, W.J. Melssen, and L.M.C. Buydens. 2003. Effect on the partial least-squares prediction of yarn properties combining Raman and infrared measurements and applying wavelength selection. *Appl. Spectrosc.* 57(6): 642-648.
- D áz-Carrillo, E., A. Muñoz-Serrano, A. Alonso-Moraga and J.M. Serradilla-Manrique. 1993. Near infrared calibrations for goat's milk components: protein, total casein, α_s -, β - and k-caseins, fat and lactose. *J. Near Infrared Spectrosc.* 1:141-146.
- Doi, E., T. Koseki and N. Kitabatake. 1987. Effect of limited proteolysis on functional properties of ovalbumin. *J. Am. Oil Chem. Soc.* 64: 1697-1703.
- Ellepola, S.W., S.-M. Choi, D.L. Phillips and C.-Y. Ma. 2006. Raman spectroscopic study of rice globulin. *J. Cereal Sci.* 43: 85-93.
- Faupel, M.L., S.B. Bambot, T. Harrell, and A. Agrawal. 2005. Multi-modal optical tissue diagnostic system. U.S. Patent 6975899 B2.
- Fisher, J. M. Baumback, J. Bowles, J. Grossmann and J. Antoniadis. 1998. Comparison of low-cost hyperspectral sensors. *SPIE Proc.* 3438: 0277-234.
- Fountain, A.W., T.J.Vickers and C.K. Mann. 1998. Factors that affect the accuracy of Raman shift measurements on multichannel spectrometers. *Appl. Spectros.* 52 (3): 462-468.
- Geladi, P. and B. Kowalski. 1986. Partial least-squares regression: A tutorial. *Anal. Chim.Acta* 185: 1-17.

- Geladi, P., D. Macdougall and H. Martens. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39(3): 491-500.
- Green, B.F., Jr. and J.W. Tukey. 1960. Complex analyses of variance: General problems. *Psychometrika* 25: 127–152.
- Gresham, C.A., D.A. Gilmore, and M.B. Denton. 1999. Direct Comparison of Near-infrared Absorbance Spectroscopy with Raman Scattering Spectroscopy for the Quantitative Analysis of Xylene Isomer Mixtures. *Appl. Spectrosc.* 53(10): 1177-1182.
- Griffiths, P.R. 2010. Introduction to the Theory and Instrumentation for Vibrational Spectroscopy. In *Applications of Vibrational Spectroscopy in Food Science*, ed. E.C.Y. Li-Chan, P.R. Griffiths and J.M. Chalmers, 31-46. Chichester, West Sussex, UK: John Wiley and Sons.
- Griffiths, P.R. and J.A. de Haseth. 2007. Introduction to vibrational spectroscopy. In *Fourier Transform Infrared Spectrometry*, ed. P.R. Griffiths and J.A. de Haseth, 1-18. New Jersey: John Wiley and Sons.
- Haan C. T. 2002. *Statistical methods in Hydrology*, 2nd edition. Ames, Iowa: Iowa State University Press.
- Hall, J.W. and A. Pollard. 1993. Near-infrared spectroscopic determination of serum total proteins, albumin, globulins, and urea. *Clin. Biochem.* 26: 483-490.
- Herrala, E. and J.Okkonen. 1996. Imaging spectrograph and camera solutions for industrial applications. *Int. J. Pattern Recogn.* 10:43-54.

- Hirschfeld, T. and B. Chase. 1986. FT-Raman Spectroscopy: Development and Justification. *Appl. Spectrosc.* 40(2): 133-137.
- Jalkanen, K.J., V. Würtz Jürgensen, A. Claussen, A. Rahim, G.M. Jensen, R.C. Wade, F. Nardi, C. Jung, I.M. Degtyarenko, R.M. Nieminen, F. Herrmann, M. Knapp-Mohammady, T.A. Niehaus, K. Frimand and S. Suhai. 2006. Use of vibrational spectroscopy to study protein and DNA structure, hydration, and binding of biomolecules: A combined theoretical and experimental approach. *Int. J. Quantum Chem.* 106 (5): 1160-1198.
- Jérez Rozo, J.I., A. Zarow, B. Zhou, R. Pinal, Z. Iqbal and R.J. Romáñach. 2011. Complementary near-infrared and Raman chemical imaging of pharmaceutical thin films. *J. Pharma. Sci.* 100(11): 4888-4895.
- Jerram, P., P.J. Pool, R. Bell, D.J. Burt, S. Bowring, S. Spencer, M. Hazelwood, I. Moody, N. Catlett and P.S. Heyes. 2001. The LLCCD: low-light imaging without the need for an intensifier. *Proc. SPIE* 4306: 178-186.
- Jung, C. 2000. Insight into protein structure and protein–ligand recognition by Fourier transform infrared spectroscopy. *J. Mol. Recognit.* 13: 325-351.
- Jung, Y.M., H.S. Shin, S.B. Kim and I. Noda. 2002. New Approach to Generalized Two-Dimensional Correlation Spectroscopy. 1: Combination of Principal Component Analysis and Two-Dimensional Correlation Spectroscopy. *Appl. Spectrosc.* 56(12): 1562-1567.

- Kaye, W. 1954. Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochim. Acta* 6(4): 257-287.
- Keller, S., T. Löchte, B. Dippel and B. Schrader. 1993. Quality control of food with near-infrared-excited Raman spectroscopy. *Fresenius J. Anal. Chem.* 346:863-867.
- Khijwania, S.K., V.S. Tiwari, F-Y. Yueh and J.P. Singh. 2007. A fiber optic Raman sensor for hydrocarbon detection. *Sens. Actuators, B.* 125: 563–568.
- Krishnan, R.S. 1971. Historical introduction. In *The Raman Effect*, ed. A. Anderson, 1-47. New York, NY: Marcel Dekker, Inc.
- Ku, M. and H. Chung. 1999. Comparison of Near-Infrared and Raman Spectroscopy for the Determination of Chemical and Physical Properties of Naphtha. *Appl. Spectrosc.* 53(5): 557-564.
- Li-Chan, E.C.Y. 1996. The applications of Raman spectroscopy in food science. *Trends Food Sci. Technol.* 7: 361-370.
- Li-Chan, E.C.Y. 2010. Introduction to Vibrational Spectroscopy in Food Science. In *Applications of Vibrational Spectroscopy in Food Science. Vol.1.* ed. E.C.Y. Li-Chan, P.R. Griffiths and J.M. Chambers, 3-29. Chichester, West Sussex, UK: John Wiley & Sons Ltd.
- Li-Chan, E.C.Y. and S. Nakai. 1989. Biochemical basis for the properties of egg white. *Crit. Rev. Poult. Boil.* 2(1): 21-59.
- Li-Chan, E.C.Y. and L. Qin. 1998. The Application of Raman Spectroscopy to the Structural Analysis of Food Protein Networks. In *Paradigm for Successful*

- Utilization of Renewable Resources*. ed. D.J. Sessa and J.L. Willett, 123-139.
Champaign, Illinois: AOCS Press.
- Liu, Y., Y-D Chen and Y. Ozaki. 2000. Two-Dimensional Visible/Near-Infrared Correlation Spectroscopy Study of Thermal Treatment of Chicken Meats. *J. Agric. Food Chem.* 48: 901-908.
- Long, D.A. 2002. *The Raman effect : a unified treatment of the theory of Raman scattering by molecules*. West Sussex, England: John Wiley & Sons.
- Madan, S.K. B. Bhaumik and J.M. Vasi. 1983. Experimental observation of avalanche multiplication in charge coupled devices. *IEEE Trans. Electron Devices* ED-30: 694-699.
- Marteau, P. N. Zanier-Szydlowski, A. Aoufi, G. Hotier and F. Cansell. 1995. Remote Raman spectroscopy for process control. *Vib. Spectrosc.* 9: 101-109.
- Matsushita, A., Y. Ren, K. Matsukawa, H. Inoue, Y. Minami, I. Noda and Y. Ozaki. 2000. Two-dimensional Fourier-transform Raman and near-infrared correlation spectroscopy studies of poly(methyl methacrylate) blends 1. Immiscible blends of poly(methyl methacrylate) and atactic polystyrene. *Vib.Spectrosc.*24: 171–180.
- McCreery, R.L. 2000. *Raman Spectroscopy for Chemical Analysis*. New York, NY: John Wiley and Sons.
- McShane, M.J., B.D. Cameron, G.L. Cot ê M. Motamedi, and C.H. Spiegelman. 1999. A novel peak-hopping stepwise feature selection method with application to Raman spectroscopy. *Anal. Chim. Acta* 388: 251-264.

- McWhirter, I. 2008. Electron Multiplying CCDs - New Technology for Low Light Level Imaging. *IRF Sci. Rep.* 292: 61–66.
- Mendes, L.S., F.C.C. Oliveira, P.A.Z. Suarez and J.C. Rubim. 2003. Determination of ethanol in fuel ethanol and beverages by Fourier transform (FT)-near infrared and FT-Raman spectrometries. *Anal. Chim. Acta* 493: 219–231.
- Mizuno, K., Y. Miyashita and Y. Shindo. 1995. NMR and FT-IR studies of hydrogen bonds in ethanol-water mixtures. *J. Phys. Chem.* 99: 3225-3228.
- Mortimer, R.G. 2008. *Physical chemistry*, 3rd edition. Burlington, MA: Elsevier Academic Press.
- Niven, R.K. 2005. Ethanol in gasoline: environmental impacts and sustainability review article. *Renew. Sust. Energy Rev.* 9(6): 535–555.
- Noble, S.D., R.B. Brown and T.G. Crowe. 2012. Design and evaluation of an imaging spectrophotometer incorporating a uniform light source. *Rev. Sci. Instrum.* 83: 0331121-0331129
- Noda, I. 1986. Two-dimensional infrared spectroscopy of synthetic and biopolymers. *Bull. Am. Phys. Soc.* 31: 520.
- Noda, I. 1993. Generalized two-dimensional correlation method applied to infrared, Raman, and other types of spectroscopy. *Appl. Spectrosc.* 47(9): 1329-1336.
- Noda, I. and Y. Ozaki. 2004. *Two-dimensional Correlation Spectroscopy – Applications in Vibrational and Optical Spectroscopy*. Chichester, West Sussex, England: John Wiley & Sons Ltd.

- Nonaka, M., E.C.Y. Li-Chan and S. Nakai. 1993. Raman Spectroscopic Study of Thermally Induced Gelation of Whey Proteins. *J. Agri. Food Chem.* 41: 1176-1181.
- Nordon, A., A. Mills, R.T. Burn, F.M. Cusick and D. Littlejohn. 2005. Comparison of non-invasive NIR and Raman spectrometries for determination of alcohol content of spirits. *Anal. Chim. Acta* 548 : 148–158.
- O'Brien, R.M. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors *Qual. Quant.* 41:673–690.
- Osborne, B.G. 1986. The use of near infrared reflectance spectroscopy in the control of flour protein content. *J. Chem. Technol. & Biotechnol.* 36(8): 364-366.
- Osborne, B.G., T. Fearn and P.H. Hindle. 1993. *Practical NIR Spectroscopy: with Applications in Food and Beverage Analysis*, 2nd edition. Essex, England: Longman Scientific & Technical; New York: John Wiley and Sons.
- Painter, P.C. and J.L. Koenig. 1976. Raman spectroscopic study of the proteins of egg white. *Biopolymers* 15: 2155-2166.
- Paradkar, M.M., S. Sakhamuri and J. Irudayaraj. 2002. Comparison of FTIR, FT-Raman, and NIR spectroscopy in a maple syrup adulteration study. *J. Food Sci.* 67(6): 2009-2015.
- Pérez-Marín, D., T. Fearn, J.E. Guerreroa and A. Garrido-Varo. 2009. A methodology based on NIR-microscopy for the detection of animal protein by-products. *Talanta* 80(1): 48-53.

- Qiao, Y. and T. A. T. G. van Kempen. 2008. Technical note: Comparison of Raman, mid, and near infrared spectroscopy for predicting the amino acid content in animal meals. *J Anim. Sci.* 82:2596-2600.
- Qiu, J.Y. and L. Shao. 2001. Near-infrared Raman instrument for rapid and quantitative measurements of clinically important analytes. *Rev. Sci. Instrum.* 72(6): 2717-2723.
- Raghavachari, R. 2001. *Near Infrared Applications in Biotechnology*. New York: Marcel Dekker, Inc.
- Samuels, M.A., S.W. Patterson, J.A. Eppstein, N.T. Yu and S. Bursell. 1996. Apparatus and methods for measuring characteristics of biological tissues and similar materials. U.S. Patent 5582168.
- Săsić, S. and Y. Ozaki. 2001. Statistical Two-Dimensional Correlation Spectroscopy: Its Theory and Applications to Sets of Vibrational Spectra. *Anal. Chem.* 73: 2294-2301.
- Savitsky, A. and M.J.E. Golay. 1964. Smoothing and differentiation of Data by simplified least squares procedures. *Anal. Chem.* 36(8): 1627-1639.
- Scholz, É., M.L. Prieto-Linde, S. Gergely, A. Salgó and E. Johansson. 2007. Possibilities of using near infrared reflectance/transmittance spectroscopy for determination of polymeric protein in wheat. *J. Sci. Food Agric.* 87(8): 1523-1532.

- Schulz, H., B. Schrader, R. Quilitzsch, S. Pfeffer and H. Krüger. 2003. Rapid classification of basil chemotypes by various vibrational spectroscopy methods. *J. Agric. Food Chem.* 51: 2475-2481.
- Shaw, R.A., S. Kotowich, H.H. Mantsch and M. Leroux. 1998. Quantitation of protein, creatinine, and urea in urine by near-Infrared spectroscopy. *Clin. Biochem.* 29(1): 11-19.
- Siesler, H.W. 2008. Basic Principles of Near-Infrared Spectroscopy. In *Handbook of Near-Infrared Analysis, 3rd edition*, ed. D.A. Burns and E.W. Ciurczak, 7-19. Boca Raton, FL: CRC Press.
- Slater, J.B., J.M. Tedesco, R.C. Fairchild and I.R. Lewis. 2001. Raman Spectrometry and its Adaptation to Industrial Environment. In *Handbook of Raman Spectroscopy*. ed. I.R. Lewis and H.G.M. Edwards. New York, NY: Marcel Dekker. Inc.
- Sorak, D., L. Herberholz, S. Iwascek, S. Altinpinar, F. Pfeifer and H.W. Siesler. 2012. New Developments and Applications of Handheld Raman, Mid-Infrared, and Near-Infrared Spectrometers. *Appli. Spectrosc. Rev.* 47(2): 83-115.
- Sostek, R.Y. and N.A. Wright. 1998. Optical Instrument providing combined infrared and Raman analysis of samples. U.S. Patent 5841139.
- Sullivan, R. 2012. *Introduction to Data Mining for the Life Sciences*. New York: Springer.
- Tate, J.D. P. Chauvel, R.D. Guenard and R. Harner. 2002. Process Monitoring by Mid- and Near-infrared Fourier Transform Spectroscopy. In *Handbook of Vibrational*

- Spectroscopy*. ed. J.M. Chalmers and P. Griffiths. New York, NY: John Wiley and Sons.
- Tøgersen, G. J.F. Arnesen, B.N. Nilsen and K.I. Hildrum. 2003. On-line prediction of chemical composition of semi-frozen ground beef by non-invasive NIR spectroscopy. *Meat Sci.* 63: 515–523.
- Treado, P.J. and M.P. Nelson. 2002. Raman Imaging. In *Handbook of Vibrational Spectroscopy*. ed. J.M. Chalmers and P. Griffiths. New York, NY: John Wiley and Sons.
- Verbeke, G. and G. Molenberghs. 1997. *Linear Mixed Models in Practice: A SAS-Oriented Approach*. New York: Springer.
- Vickers, T.J., C.A. Rosen and C.K. Mann. 1996. Compact Raman spectrometers: data handling methods. *Appl. Spectrosc.* 50(8): 1074-1081.
- Wang, M-L., Y-M. Choong, N-W. Su and M-H. Lee. 2003. A rapid method for determination of ethanol in alcoholic beverages using capillary gas chromatography. *J. Food Drug Anal.* 11(2): 133-140.
- Wang, W. and J. Paliwal. 2006. Design and evaluation of a visible-to-near-infrared electronic slitless spectrograph. *Meas. Sci. Technol.* 17(10): 2698-2704.
- Wang, W., A. Major and J. Paliwal. 2012. Grating-Stabilized External Cavity Diode Laser for Raman Spectroscopy – A Review. *Appl. Spectrosc. Rev.* 47(2): 116-143.
- Wartewig, S. 2003. *IR and Raman Spectroscopy Fundamental Processing*. New York, NY: John Wiley & Sons.

- Wen, Z. 2007. Raman Spectroscopy of Protein Pharmaceuticals. *J. Pharm. Sci.* 96(11): 2861-2878.
- Weijers, M., P.A. Barneveld, M.A.C. Stuart and R.W. Visschers. 2003. Heat-induced denaturation and aggregation of ovalbumin at neutral pH described by irreversible first-order kinetics. *Protein Sci.* 12:2693–2703.
- Weyer, L.G. and S.-C. Lo. 2002. Spectra–Structure Correlations in the Near-infrared. In *Handbook of Vibrational Spectroscopy*. ed. J.M. Chalmers and P. Griffiths. New York, NY: John Wiley and Sons.
- Wilks, S.S. 1963. Multivariate Statistical Outliers. *Sankhya A* 25: 407-426.
- Williams, P.C. 2007. Grains and Seeds. In *Near-Infrared Spectroscopy in Food Science and Technology*. ed. Y. Ozaki, W.F. McClure and A.A. Christy, 165-217. Hoboken, NJ: John Wiley & Sons Inc.
- Workman, J.J. 1999. Review of Process and Non-invasive Near-Infrared and Infrared Spectroscopy: 1993–1999. *Appl. Spectrosc. Rev.:* 34(1&2), 1–89.
- Workman, J. and L. Weyer. 2008. *Practical Guide to Interpretive Near-Infrared Spectroscopy*. Boca Raton, FL: Taylor & Francis LLC.
- Yang, H., J. Irudayaraj and M. M. Paradkar. 2005. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chem.* 93: 25–32.
- Ysacc Sato-Berrú, R., J. Medina-Valtierra, C. Medina-Gutiérrez and C. Frausto-Reyes. 2004. Quantitative NIR Raman analysis in liquid mixtures. *Spectrochim Acta A* 60: 2225–2229.

Zhang, Z., S. Chen, Y. Liang, Z. Liu, Q. Zhang, L. Ding, F. Ye and H. Zhou. 2009. An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. *J. Raman Spectrosc.* 41(6): 659-669.

Zhao, J., H. Lui, D.I. McLean and H. Zeng. 2007. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Appl. Spectrosc.* 61(11):1225-1231.

APPENDIX A. Data Acquisition Settings for Different Samples

Samples	Raman parameters		NIR acquisition parameters	
	Integration time (s)	Accumulation	Integration time (s)	Accumulation
Ovalbumin solution	60	4	0.01	60
Ovalbumin gel (70 °C)	30	6	N/A	N/A
Ovalbumin gel (90 °C)	20	6	0.01	50
Ovalbumin solid powder	20	1	0.015	60
water	30	1	0.015	60
Ethanol solution	7	6	0.01	50
Pure ethanol	3	1	0.01	50
Acetaminophen	3	1	N/A	N/A