Vibrational spectroscopic analysis of complex systems

Chima Robert

A thesis submitted for the degree of:

Doctor of philosophy

University of Otago

June 2021

Abstract

Vibrational spectroscopy (infrared and Raman spectroscopy) is based on the interaction of matter with electromagnetic radiation. These techniques probe a wide range of information including physical and chemical properties yielding large dimensions of data. The advancement in the art of chemometrics for accurate prediction of sample properties presented a crucial turning point in the development of the spectroscopic methods, facilitating the interpretation of the large spectral data.

This thesis investigates the application of different chemometric modelling techniques to vibrational spectroscopic data from samples like meat, pharmaceuticals and immune cells. The nature of information probed determined the choice of chemometric techniques; however, some techniques gave better approximation (performance) for one sample than the others might. The analytical applications conducted in this thesis includes; discrimination of red meat species, quality assessment of red meat using conventional Raman and infrared spectroscopy, discrimination of cancer and immune cells using Raman microscopy and isothermal dehydration of pharmaceutical crystalline hydrates using low-frequency Raman spectroscopy.

Chapter one of this thesis introduces the concept of vibrational spectroscopy, spectroscopic instrumentation, spectral preprocessing, multivariate analysis (chemometrics) and highlights the aims of this thesis.

The second chapter investigates the discrimination of red meat species (beef, lamb and venison) using Raman and infrared spectroscopic techniques. Classification models were built using partial least square discriminant analysis (PLSDA) and support vector machine classification (SVMC); whereas principal component analysis (PCA) was employed for exploratory

analysis. Visual assessment of the PCA scores revealed distinct separation of the three meat species for both spectroscopic techniques. Classification models built using the Raman data and validated against an independent test set yielded a classification accuracy ≥ 80 % and ≥ 92 % for the PLSDA and SVMC methods, respectively. Classification model created using the infrared data yielded an accuracy ≥ 94 % for both chemometric methods. This suggests that both Raman and infrared spectroscopic methods posit an effective tool for red meat discrimination.

Chapter three of this thesis investigates the implementation of Raman and infrared techniques as well as three data fusion strategies to evaluate pH and percentage intramuscular fat (% IMF) content of red meat. Quantitative models were built using partial least square regression and validated against an independent test set. Results obtained suggest a good correlation between the reference and predicted pH values using the Raman, infrared and high-level data fusion strategy whereas Raman and low-level fusion showed similar level of performance for predicting the % IMF content in red meat.

In chapter four, low-frequency Raman spectroscopy was shown to be very sensitive for monitoring the *in situ* isothermal dehydration of piroxicam and theophylline monohydrates. The dehydration was performed at four different temperatures and monitored in both the low-wavenumber (20 – 300 cm^{-1}) and mid-wavenumber (335 – 1800 cm^{-1}) Raman regions. Analysis performed using multivariate curve resolution (MCR) suggested the formation of specific anhydrous forms of piroxicam and theophylline upon dehydration of their respective monohydrates. The formation of the anhydrous forms was also detected on different timescales (approx. 2 min) between the low- and mid-wavenumber Raman regions. This finding highlights the differing nature of the vibrations being detected between these spectral regions.

In chapter five, Raman spectroscopy was demonstrated as a sufficient tool to discriminate cancer and immune cells. Phenotype of T-cells and monocytes were incubated with media conditioned by glioblastoma stem-cells (GSCs) showing different molecular background. Multivariate analysis performed using principal component - linear discriminant analysis (PCA-LDA) and SVM yielded sensitivities and specificities ≥ 70 % and ≥ 67 %, respectively. The results were in agreement with the flow cytometry analysis.

Acknowledgment

I would like to express my profound gratitude to my supervisors; Professor Keith C. Gordon and Dr. Sara J. Fraser-Miller for their constant guidance, support and encouragement throughout the course of my PhD. I have gained so much from their valuable feedbacks at various stages of my projects. In addition, thank you for the many opportunities given to me over the duration of this PhD, I am eternally grateful.

I would also like to thank my collaborators whose contributions resulted in the overall formation of this thesis. First, Dr. Cameron Craigie – led team at AgResearch, New Zealand for the providing and performing the wet chemistry analysis on the meat samples. Professor Jukka Rantanen and Professor Thomas Rades for their insight during the interpretation of some of the results in chapter four. Dr. Jaroslaw Maciaczyk – led team at Department of Surgical Science, University of Otago for their contribution on the cells project. Notable mention to Julia Tsiampali (Neurosurgery Department, University of Duesseldorf), Donata Maciaczyk and Silke Neumann (both of Department of Pathology, University of Otago) for providing and performing further analysis on the cells samples.

Many thanks to the members of the KCG group, Department of Chemistry, University of Otago for always ensuring a friendly and positive work environment. Special mention to Jeremy Rooney for getting me started on chemometrics, Joshua Sutton for helping me with instrumentation and Kārlis Bērziņš for some of your brilliant ideas and graphical abstract designs.

I would like to thank the University of Otago for providing me with the Doctoral scholarship that ensured I had financial stability over the duration of my PhD programme.

I would also like to thank Dr. Ukpai Eze (University of Coventry, UK) for his constant encouragements and support throughout my study years.

Finally, I would like to thank my family for their patience and prayers. I love you all.

Dedication

This PhD thesis is dedicated to the millions of Biafran children who were starved to death and the gallant men and women who have given their lives for a free independent Biafra.

List of Abbreviations

API	active pharmaceutical ingredients
ATP	adenosine triphosphate
ATR	attenuated total reflectance
CCD	charge coupled device
CD73	ecto-5'-nucleotidase
DFD	dark firm and dry
DFT	density functional theory
DSC	differential scanning calorimetry
EMT	epithelial to mesenchymal
FT	Fourier transform
FTIR	Fourier transform infrared
GBM	glioblastoma
GSC	glioblastoma stem cells
HLF	high-level fusion
IMF	intramuscular fat
IR	infrared
LBC	linear baseline correction
LDA	linear discriminant analysis

LFR	low frequency Raman
LLF	low-level fusion
LV	latent variable
MASM	moving average smoothing
MCR	multivariate curve resolution
MFR	midfrequency Raman
MLF	mid-level fusion
MLR	multilinear regression
MNorm	mean normalization
MSC	multiplicative scatter correction
MVA	multivariate analysis
NA	numerical aperture
ND:YAG	Neodymium-Yttrium Aluminium Garnet
NIPAL	non-iterative partial least square
NIR	near infrared
NMR	nuclear magnetic resonance
NRMSE	normalized root mean square error
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PKE	palm kernel expeller

PLS	partial least square
PLSDA	partial least square discriminant analysis
PLSR	partial least square regression
PSE	pale soft and exudative
PXA	piroxicam anhydrous
PXM	piroxicam monohydrate
RBC	rubberband correction
RFN	reddish firm and non-exudative
RMSE	root mean square error
RMSEC	root mean square error of calibration
RMSECV	root mean square error of cross-validation
RMSEP	root mean square error of prediction
RS	Raman spectroscopy
SBC	shape baseline correction
SD	standard deviation
SGSM	Savitzky Golay smoothing
SNR	signal to noise ratio
SNV	standard normal variate
SVMC	support vector machine classification
TCM	tumour condition medium
TGA	thermal gravimetric analysis

TPAH	theophylline anhydrous
TPMH	theophylline monohydrate
UV/Vis	ultraviolet / visible
XRPD	X-ray powder diffractometry
ZEB	zinc finger e-box binding homeobox

Table of Contents

List of Figures

List of Tables

Publications

1. Chima Robert, Sara J. Fraser-Miller, William T. Jessep, Wendy E. Bain, Talia M. Hicks, James F. Ward, Cameron R. Craigie, Mark Loeffen and Keith C. Gordon: Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy. *Food Chemistry* 2021, *343*, 128441. https://doi.org/10.1016/j.foodchem.2020.128441

2. Chima Robert, William Jessep, Joshua J. Sutton, Talia M. Hicks, Mark Loeffen, Mustafa Farouk, James F. Ward, Wendy E. Bain, Cameron R. Craigie, Sara J. Fraser-Miller and Keith C. Gordon: Evaluating low-, mid- and high-level fusion strategies for combining Raman and infrared spectroscopy for quality assessment of red meat. *Food Chemistry* 2021, *361*, 130154. https://doi.org/10.1016/j.foodchem.2021.130154

3. Chima Robert, Sara J. Fraser-Miller, Kārlis Bērziņš, Peter O. Okeyo, Jukka Rantanen, Thomas Rades and Keith C. Gordon: Monitoring the isothermal dehydration of crystalline hydrates using low-frequency Raman spectroscopy. *Molecular Pharmaceutics*, 2021, 18, 1264-1276. https://pubs.acs.org/doi/10.1021/acs.molpharmaceut.0c01126

4. Chima Robert, Julia Tsiampali, Sara J. Fraser-Miller, Silke Neumann, Donata Maciaczyk, Sarah L. Young, Jaroslaw Maciaczyk and Keith C. Gordon: Molecular monitoring of glioblastoma's immunogenicity using a combination of Raman spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 2021, 252, 119534. https://doi.org/10.1016/j.saa.2021.119534

Chapter 1

Introduction

In this thesis, a wide range of systems were investigated using vibrational spectroscopic techniques. This introduction briefly outlines Raman and infrared spectroscopy, experimental methodologies, spectral preprocessing techniques and multivariate analytical tools used within this thesis. Owing to the variation in the samples studied, each chapter has its own introduction to review the peculiarity of the system, analytical methods and data analysis pertaining to that chapter. With the exception of chapter one, each chapter comprises inclusion of published materials with permission from the publishing journals.

1.1 Spectroscopy

Spectroscopy encompasses the study of interaction of matter with electromagnetic radiations (or light) often resulting in transition between quantised electronic, vibrational and/or rotational states. This interaction could result in different phenomena (or processes) including absorption, emission or scattering of the incident photons (light) by the illuminated material [1, 2]. These phenomena can facilitate the understanding of the matter under illumination.

The absorption phenomenon involving transitions between the lowest energy state to a vibrationally excited state results in transitions observed in infrared (IR) and near infrared (NIR) spectroscopy. Absorption is a single photon phenomenon, where one photon provides the energy for transition from the ground state to an excited state and disappears after the transfer. In IR spectroscopy, the photon induces a transition between

vibrational energy levels. Those from $\nu = 0 \rightarrow \nu = 1$ are called fundamental transition and can be strong. The vibrational levels are associated with normal vibrational modes such as bond stretching and bending. In addition to fundamental transition ($\nu = 0 \rightarrow \nu = 1$), one may also observe overtones and combination bands of fundamental vibrations at near IR frequencies. These are much weaker than fundamental transitions and are observed on transitions from $\nu = 0 \rightarrow \nu = 2, 3, 4 \ldots$ [3].

Conversely, the scattering phenomenon is a two-photon process involving an incident and a scattered photon. On interaction of matter with monochromatic light, the incident photons are scattered by the matter in different directions. First observed by J. W. Strutt, the majority of the scattered photons will consist of the same energy as the incident photon (hv_0). This is commonly referred to as elastic or Rayleigh scattering [4]. However, approximately 0.000001 % of the scattered photons were observed to have energies above or below the incident photon, this is referred to as inelastic or Raman scattering after C.V Raman who discovered this phenomenon [5]. The energy difference between the incident and scattered photon is informative on the nature of the sample under illumination.

If the Raman scattering process involves a loss of energy to the molecule under illumination, the energy of the scattered photon (hv_s) is less than the incident photon ($hv_0 - \Delta E$). This is the Stokes scattering. Whereas if the scattering process occurred by gain of energy from the molecule, which is possible for transition from a vibrationally or rotationally excited state to a ground state, the energy of the scattered photon (hv_s) is obtained as ($hv_0 + \Delta E$). This is the anti-Stokes scattering [5, 6]. Figure 1.1 describes the energy levels leading to possible transitions that can occur on interaction of light with matter.

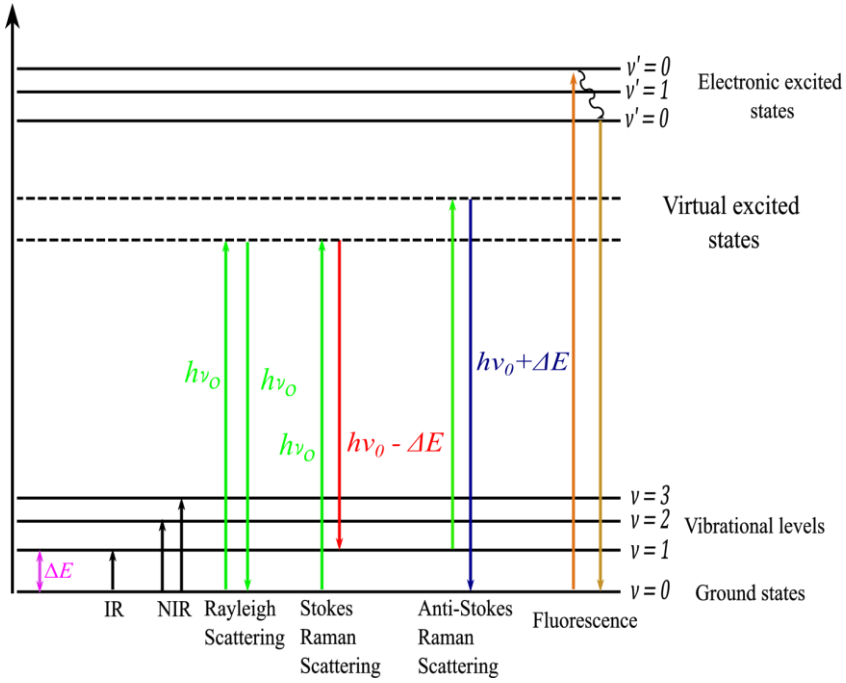

Figure 1.1: Energy diagram showing the processes on interaction of light with matter including absorption, fluorescence and scattering of photons. The energy of the incident photon is given as $h\nu_0$. The energy difference for transition between the ground state ($\nu = 0$) and first vibrationally excited state ($\nu = 1$) is labelled as ΔE. Fluorescence process involves absorption of photon into an excited state with subsequent emission from a lower level excited state to the ground state. This Figure was adapted from Fraser S.J [7].

Vibrational spectroscopy commonly refers to Raman scattering and IR absorption techniques, as well as inelastic neutron scattering [8]. As earlier highlighted, infrared spectroscopy is based on the absorption of electromagnetic radiation at fundamental frequencies, where the frequency of the electromagnetic radiation matches the vibrational frequency. For a vibrational or rotational transition to be IR active, a net change in dipole moment in the molecule is required upon vibration or rotation. For example, the charges in HBr are not evenly distributed since bromine is more electronegative than hydrogen, and has a higher electron density. The more electronegative bromine atom has a slightly negative charge, δ^-, and the hydrogen atom has a slightly positive charge, δ^+, as such vibration of this bond results in a net change in dipole moment (IR active mode). Upon vibration of the molecule, a fluctuation in the dipole moment is experienced resulting in

a field that interacts with the electric field of the electromagnetic radiation. Absorption occurs if there is a match between the frequencies of the electromagnetic radiation with the natural vibration of the molecule. In contrast, molecules such as Cl_2, N_2 and Br_2, do not possess a net dipole moment, as such the associated vibrations or rotations are IR inactive [9].

Conversely, for a molecule to be Raman active, rotational and vibrational transitions within the molecule must result in changes in molecular polarizability (α). On interaction of a molecule with an electric field (from the laser beam), an induced dipole moment (P) is produced due to distortion of the electron cloud. This distortion is caused by the attraction of the positively charged nucleus towards the negative pole in the electric field and the negatively charged electron towards the positive pole of the electric field. As such, for a Raman active molecule, $\left(\frac{d\alpha}{dq}\right)_0 \neq 0$ (i.e., slope near equilibrium position). A simple illustration of Raman activity in smaller molecules like CO_2 can be shown using the polarizability ellipsoid, which is a surface that extends $\frac{1}{\sqrt{\alpha_i}}$ from the centre of the molecule (where i is the direction from the centre). A diagrammatical representation of the polarizability ellipsoid is shown in Figure 1.2(a) [10].

From Figure 1.2(a), the size of the ellipsoid in the v_1 vibration is changing; hence, it is Raman active. However, despite the changing size and shape of the ellipsoid in the v_3 vibration, the ellipsoid at +q and -q are the same, as such the vibration is Raman inactive. Similarly, despite the shape of the ellipsoid in the v_2 vibration, the shape and size at the two extremes are the same, hence, v_2 is Raman inactive [10].

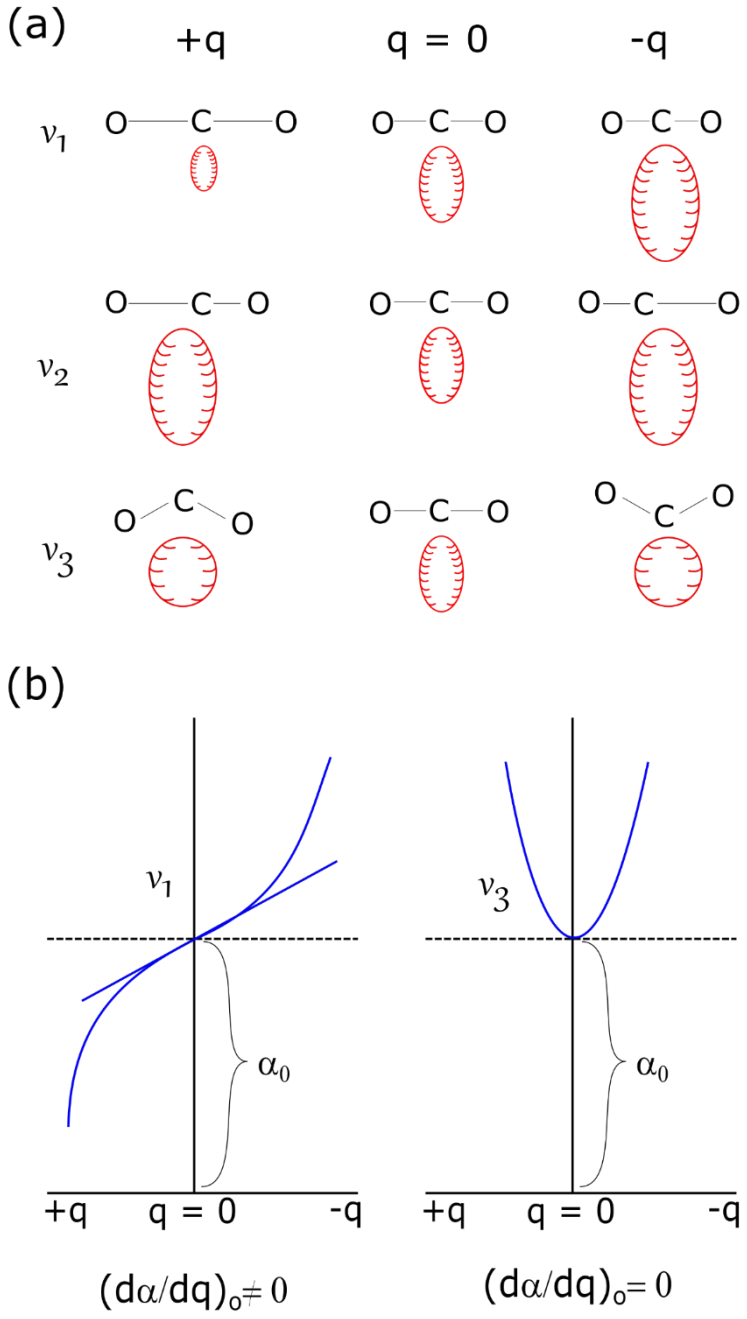

Figure 1.2: (a) Vibration of a CO_2 molecule showing changes in the polarizability ellipsoids (b) Difference between Raman active (ν_1) and Raman inactive (ν_2, ν_3) vibrations in a CO_2 molecule. These figures have been adapted from Ferraro, J.R. *et al.*[10]

For symmetric molecules with a centre of inversion, Raman and IR are mutually exclusive. Since the selection rules for IR and Raman spectroscopy are different, bonds

that are IR-active will be Raman inactive and vice versa. Hence, Raman spectroscopy is often considered a complementary technique to IR spectroscopy. Despite providing molecular level information, each of these techniques possesses distinct pros and cons for analytical applications.

Raman spectroscopy is ideal for analysing samples in aqueous solution or in samples with a high moisture content like meat. This is because water is a weak Raman scatterer and thus gives negligible interference on the Raman spectra. Unlike Raman, specific regions of the IR spectrum can be obscured by water signal owing to the absorption of IR radiation by water.

The inherent weak nature of Raman scattering (1 in 10^6 to 10^8) necessitates the use of powerful lasers for improved signals, which could lead to photodecomposition of samples [10]. In contrast, IR spectrometers require the use of easily accessible broad-spectrum light sources (such as Globar) owing to absorption being a common phenomenon. Limited sample preparation is required for the use of Raman, and measurements can be performed through optically transparent materials such as glass window and packaging. This is challenging in IR, owing to the absorption of the materials that might obscure the signal of the analyte of interest. The use of transparent matrix such as solid KBr for IR analysis results in sample contamination.

Conventional Raman spectroscopic instruments focus on Stokes signals (200 cm^{-1} – 4000 cm^{-1}) providing information relating to the intramolecular interaction within a sample. Access to the low frequency Raman (LFR) region (< 10 – 200 cm^{-1}) is possible on incorporation of ultra-narrow bandpass filters such as volume Bragg gratings, allowing for filtering of the dominant Rayleigh signal [11, 12]. The LFR probes the collective translation, libration and deformation of the molecular skeleton within a crystal lattice, i.e., intermolecular vibrations. The LFR region is characterized by much stronger signal as compared to the mid frequency Raman (MFR) region for many materials [12, 13]. One reason for the strong intensities at low frequency Raman shifts, $\tilde{\nu}_o$, is that the Raman cross section for a vibrational mode j, $\left(\frac{\partial \sigma_j}{\partial \Omega}\right)$ is strongly dependent on the scattering frequency, $(\tilde{\nu}_0 - \tilde{\nu}_j)^4$:

$$\left(\frac{\partial \sigma_j}{\partial \Omega}\right) = \left(\frac{2^4\pi^4}{45}\right)\left(\frac{(\tilde{v}_0 - \tilde{v}_j)^4}{1 - exp\left[\frac{-hc\tilde{v}_j}{kT}\right]}\right)\left(\frac{h}{8\pi^2 c\tilde{v}_j}\right)S_j \qquad 1.1$$

where $\tilde{v}_0$ refers to the laser frequency, *T* is the temperature in Kelvin and the remainder are physical constants [13, 14]. For a typical measurement of polarized Raman optical activity, the Raman scattering activity, S_j, which is related to the derivative terms of the mean polarizability ($\bar{\alpha}_j^2$) and anisotropy (β_j^2), can be expressed as:

$$S_j = g_j\left[45\bar{\alpha}_j^2 + 7\beta_j^2\right] \qquad 1.2$$

where g_j refers to the degeneracy of the vibrational model *j* [12, 15].

1.2 Dispersive versus Fourier transform (FT) spectrometers

For the purpose of this thesis, data collection was performed using both dispersive and Fourier transform (FT) instruments. As such, a general description of both setups is presented below.

Current dispersive Raman instrument setup consists of the excitation source (laser), optics for sample illumination and light collection, wavelength selector (filter or spectrometer) and the detector. There are varieties of excitation sources including argon (488 and 514.5 nm), krypton ion (530.9 and 647.1 nm), helium-neon (632.8 nm), diode (785 and 830 nm) and Neodymium-Yttrium Aluminium Garnet (ND:YAG,) (1064 nm). The choice of laser depends on the intended application, with shorter wavelengths prone to inducing possible photodecomposition and fluorescence (argon and krypton ion) whereas longer wavelength can reduce fluorescence and photodecomposition while

operating at higher power [2, 6]. Isolation of single laser beam is performed using the bandpass filters, with the lens guiding the laser beam to the sample. The dominant Rayleigh scattering is separated from the weak Raman scattering using a combination of notch filters and grating monochromator. The detectors such as Germanium (Ge), Indium-Gallium-Arsenic (InGaAs) or multichannel based charged couple device (CCD) then convert the incoming optical signal into charge, which is integrated and transformed into readout [2, 16].

The Michelson interferometer is the main component of the FT instruments. In FT spectrometers, the electromagnetic radiation from the source is split into two partial beams using a beamsplitter, which are reflected on a moving and fixed mirror then back to the beamsplitter, where the beams recombine. The optical path length of the beam is changed owing to the moving mirror, resulting in either constructive or destructive interference of both beams (interferogram). A Fourier transformation is performed on the interferogram to obtain the spectrum [3]. A schematic representation of the dispersive and FTIR spectrometers is presented in Figure 1.3.

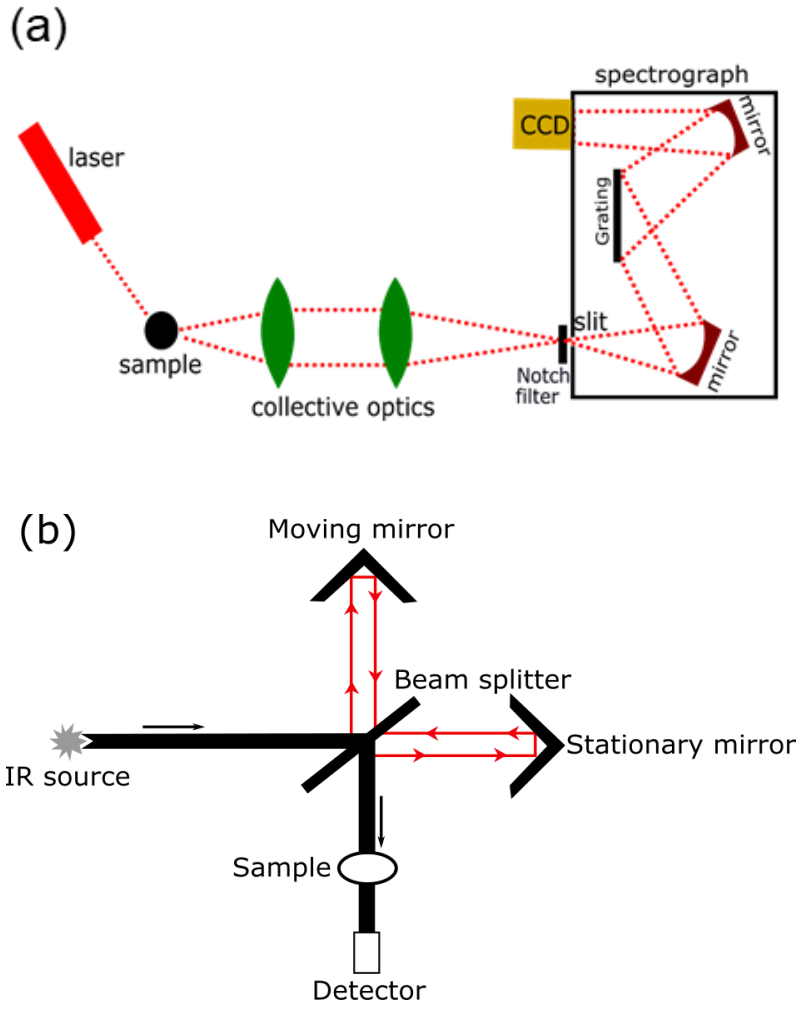

Figure 1.3: A schematic representation of spectrometers (a) dispersive Raman (b) Fourier transform instrument. The notch filter in the Raman spectrometer is used in removing the predominant Rayleigh scattering and laser line, with the Raman signal imaged onto the CCD detector by a combination of mirrors and grating system.

FT spectrometers have the multiplex advantage due to the simultaneous measurement of multiple wavelengths in the interferometer increasing acquisition speed. Each point in the interferogram contains information from each wavelength of light, since light is not separated into individual frequencies as compared to the dispersive instruments. This allows for the acquisition of many scans, which are further averaged on the FTIR in a shorter time. Also, the use of internal calibration in FTIR instrument offers the advantage of high precision as compared to dispersive IR instruments, which use

external calibration [3]. Despite these potential advantages, the application of modern dispersive spectrometers, particularly in Raman spectrometers has surged. This is due to advances in optical technologies such as notch filters, which have facilitated the removal of the dominant Raleigh scattering without the introduction of additional gratings. The incorporation of charge coupled device (CCD) detection has enabled the collection of multiple wavelength simultaneously (the multiplex advantage), as such, rapid co-addition of spectral data has become a viable method in improving the signal to noise ratio (SNR). The reduced mirrors and grating component has also increased the throughput and reduced the scan time in dispersive spectrometers [17].

1.3 Sampling

The sampling technique and volume play a vital role in the acquisition of accurate and representative spectral data using different spectroscopic techniques. Large area scan relative to sample size is required for heterogeneous samples like foodstuffs and pharmaceuticals whereas small area sampling might be appropriate for studies where the substructure is of interest. An illustration of the large area scan, used in the meat studies, along the *x*- and *y*-axis is shown in Figure 1.4.

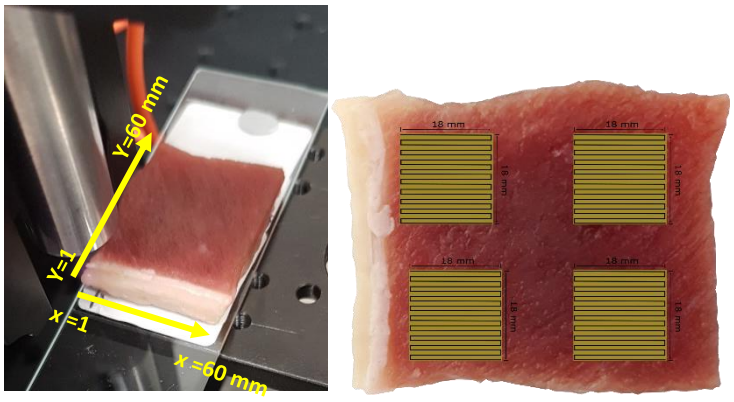

Figure 1.4: Illustration of the large area scan technique adopted for analysis of heterogeneous samples like red meat by incorporating an x, y stage. This facilitates better representation of the sample under study with the sample attached to it.

1.4 Instrumentation

The experiments described in this thesis were performed using a range of Raman and infrared spectroscopic instruments including a Vertex 70 spectrometer (Bruker Optics, Ettlingen, Germany) with a Platinum attenuated total reflectance (ATR) diamond accessory (GladiATR, Pike Technologies, Madison, USA) equipped with a wide range, room temperature, DLaTGS (Lanthanum α Alanine doped Triclycine sulphate) detector. This setup is the FT-IR instrument utilized herein.

The FT-Raman setup comprised of a MultiRam FT-Raman spectrometer (Bruker Optics, Ettlingen, Germany) equipped with a liquid N_2 cooled Ge detector and 1064 nm Nd:YAG laser. In this work the downward looking objective with a defocused aperture (~2 mm diameter spot size) in back scattering geometry was used. This is the FT-Raman instrument. Both the FT-IR and FT-Raman instruments were used for preliminary studies on samples requiring bulk analysis and had the disadvantage of subsampling when analysing heterogeneous samples like meat, and long acquisition time specifically for FT-Raman.

Large area scanning was facilitated by incorporating a raster system moving along the *x* and *y*-axis while measuring with the WP 1064 nm Raman spectrometer (Wasatch Photonics, USA) equipped with TEC cooled G9214-512S InGaAs array Hamamatsu detector. The Raman system was connected using flexible fibre optics and an adaptable probe head for flexible sampling. The raster system consisted of two UTS100CC linear motorized platforms (Newport Corp. Irvine, CA, USA) and ESP301-3G 3-axis controller (Newport Corp. Irvine, CA, USA) was used for positioning the sample. The raster system allowed for measurement across 100 mm x 70 mm dimension.

A custom-built low-frequency Raman setup allowed for simultaneous collection of the conventional mid-wavenumber and low-wavenumber Raman regions. The setup consisted of a 785 nm laser module excitation source (Ondax, Inc. Monrovia, CA, USA) which was filtered by BragGrate bandpass filters (OptiGrate Corp. Oviedo, FL, USA) to remove amplified spontaneous emission before irradiating the sample. Backscattered

light from the sample was collected and filtered through a set of volume Bragg gratings (Ondax Inc., Monrovia, CA) and focused into an LS 785 spectrograph (Princeton Instruments, Trenton, NJ, USA) where the light was dispersed onto a CCD detector (PIXIS 100 BR CCD, Princeton Instruments, Trenton, NJ, USA).

Microscale sample analysis was performed using an Alpha 300R+ confocal Raman microscope (WITec GmbH, Ulm, Germany), with 532 and 785 nm incident laser options (Coherent, California). The confocality of the system allows for depth measurements and imaging, with the spatial resolution particularly through the z-axis dependent on the sample shape and refractive index [18]. The lateral resolution or diameter (d) of the spot size is limited by the Abbe relationship as expressed in Equation 1.3

$$d = \frac{0.62 * \lambda\ (nm)}{NA} \qquad 1.3$$

where λ is the wavelength of the laser (nm) and *NA* is the numerical aperture of the microscope objective used [19]. A summary of the diffraction limited spot size with our current Raman microscope instrument is presented in Table 1.1.

Table 1.1: The calculated spot sizes for the Alpha 300R+ confocal Raman microscope at different wavelengths and objectives.

Objective	*NA*	*d (nm)* at 532 nm	*d (nm)* at 785 nm
10x	0.25	1319	1947
50x	0.8	412	608
100x	0.9	366	541

1.5 Spectral preprocessing

Despite not being a substitute for optimal data acquisition, preprocessing of spectral data is commonly applied prior to building chemometric models. Spectral preprocessing techniques were employed to remove features not associated with IR absorption and Raman scattering including random noise, intensity variations due differences in sample focus as well as baseline element from fluorescence and thermal emission sources. Selection of appropriate preprocessing technique to be applied on a spectral data can be performed via a variety of ways including visualisation of spectral data, trial and error approach or use of data quality parameters. For the purpose of this thesis, a range of spectral preprocessing techniques was adopted for different samples based on educated guess; with these techniques discussed below.

1.5.1 Noise reduction

Noise is a term that encompasses undesirable components that occur over time, which could negatively influence the overall spectral quality and interpretation. In general, electrical and thermal signals are the main contributors to the noise component in spectral data. Some noise sources can be minimized or eliminated using dark spectrum subtraction whereas others can be averaged out by increasing acquisition time, laser power or co-addition of multiple spectra [20, 21]. Spectral quality has been described in relation to the signal-to-noise ratio (SNR), which is the inverse of the relative standard deviation of the measured value [22]. For example, the SNR of the peak intensity of a Raman band can be expressed as the ratio of the mean peak height ($\bar{S}$) to the standard deviation (σ_y) as shown in Equation 1.4.

$$SNR = \frac{\bar{S}}{\sigma_y} \qquad 1.4$$

where $\bar{S}$ is the mean peak height and σ_y is the standard deviation of the peak height [22].

Optimising analytical techniques will potentially lead to reduction in noise signal and consequently lead to increase in the SNR. One approach to achieving this is in repetitive scanning and co-addition of individual spectra [20, 21]. The relationship between the *SNR* and number of co-additions (*N*) is expressed in Equation 1.5.

$$SNR \propto \sqrt{N} \qquad 1.5$$

From Equation 1.3, a 1.4-fold increase in SNR is obtained on doubling the number of scans or length of scan, whereas a 10-fold increase is obtained with a 100-fold increase in the number of co-additions. Continually increasing the number of co-additions may not be feasible owing to constraints in computer memory for storing accumulated spectral data, longer acquisition time and possibility of thermal degradation on increased exposure of samples to electromagnetic radiation [21]. However, further noise reduction can be performed post-spectral acquisition using varied smoothing schemes including Savitzky-Golay smoothing (SGSM) and moving average smoothing (MASM), but their selection and adaptation depends on the application. For the purpose of this thesis, only SGSM was applied and is described below.

1.5.2 Savitzky-Golay smoothing (SGSM)

Smoothing is regularly applied for the removal of high frequency components in spectral data. The Savitzky-Golay smoothing utilises a moving window polynomial fitting technique, replacing the original values with more regular variations after the fit. The SGSM method requires parameters like the polynomial order (*o*) and filter windows size (*w*) [23].

For a given signal measured at *n* points, and filter windows, *w*, the SGSM algorithm fits a polynomial of order, *o*, in each filter window as the filter is moved across the signal. The polynomial fit at the center gives rise to the filter estimate which replaces the centre point value for each window (for spectral data, *w* is usually an odd integer for ease of

calculation) [24, 25]. The polynomial order must be less than or equal to the number of filter windows and the number of filter windows less than the number of variables to be estimated. A representation of noise reduction on Raman spectral data using different SGSM filter windows is shown in Figure 1.5.

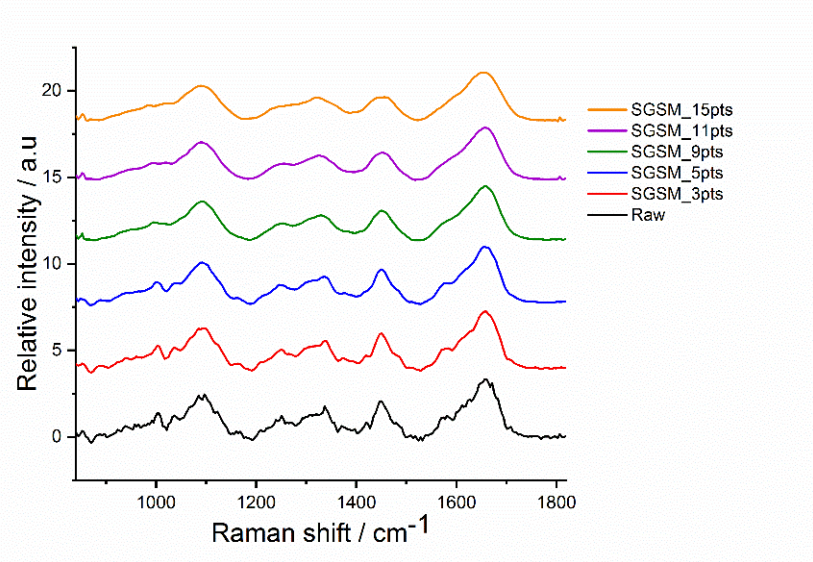

Figure 1.5: Representation of Raman spectra pre- and post-smoothing using Savitzky-Golay smoothing filter, first order polynomial with varied filter windows from 3 to 15. The figure shows improved spectral quality on smoothing; however, selecting the appropriate filter window is of critical importance to avoid removal of important spectral information.

1.5.3 Baseline correction

Baseline artefacts in spectroscopic data arise from a range of sources including fluorescence, phosphorescence, Mie scattering, detector thermal noise, sample focus and particle size variations. This induces unequal shifts in the amplitude across different wavenumbers and consequently leading to curved baseline [23, 26]. It is important that baseline artefacts be corrected to enhance subtle differences desirable for comparing spectral data, particularly if further analysis is performed using chemometrics. This is also essential for objectivity and reproducibility.

Several methods are available for baseline correction including linear baseline correction (LBC), rubber band correction (RBC), polynomial fitting, differentiation-based methods and the more software specific shape baseline correction (SBC) (WITec software).

1.5.3.1 Linear baseline correction (LBC)

Linear baseline correction identifies two variables, which defines the new baseline and subtract a linear slope from the spectral data as shown in Figure 1.6. The two selected variables or points are defined as 0, with the rest of the variables transformed with linear interpolation/extrapolation [27]. It is important that the variables selected possess no spectroscopic bands. LBC is effective in transforming a sloped baseline to a horizontal baseline, working best for smaller wavenumber ($\tilde{\nu}$) range, as such, LBC is not effective for transforming a curved baseline. LBC for a single spectrum is mathematically described by the equation

$$x = \tilde{x} + \alpha + \beta x \qquad 1.6$$

where x is the vector representing the spectrum, $\tilde{x}$ is the vector representing the spectral features of interest, α and β are the estimated slope and offset which are based on at least two points, which are thought to contain only baseline information [28].

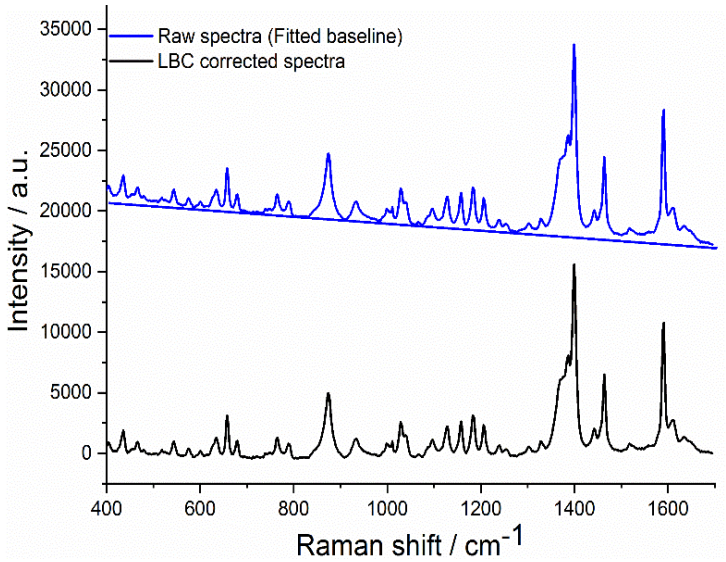

Figure 1.6: Representative Raman spectrum of piroxicam monohydrate pre- (blue) and post- (black) linear baseline correction.

1.5.3.2 Polynomial fitting and subtraction

Polynomial fitting offers a distinct advantage of retaining the input spectral intensities and contours [29]. A specific number of baseline points is fitted with an *n*-order polynomial to estimate the baseline. The estimated curve baseline is compared to original data and if the baseline matches, it is subtracted from the spectrum to remove the curved baseline [30]. This baseline correction technique may not be feasible for large spectral data sets, owing to the manual selection of the polynomial order and visualisation of the spectral data. An illustration of a polynomial subtraction for baseline correction is shown in Figure 1.7. The original spectrum was fitting with a 3rd order polynomial and the baseline component subtracted.

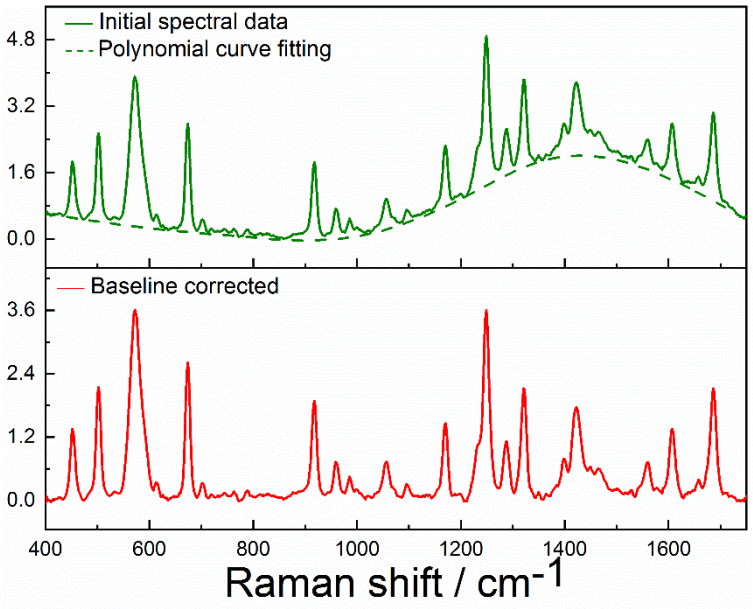

Figure 1.7: A representation of polynomial subtraction for baseline correction performed on a Raman spectrum of theophylline monohydrate.

1.5.3.3 Rubber band correction (RBC)

Rubber band baseline correction is effective in transforming curved or nonlinear baseline to a horizontal baseline in a given spectrum, particularly in emissive materials like biological samples. RBC utilizes series of polynomial function to fit a concave curve on a part or entire portion of the spectrum. The concave curve is further subtracted from the spectrum resulting in a horizontal baseline [31]. The RBC method has been incorporated in spectral acquisition software like OPUS 7.5 software (Bruker Optik, Ettlingen, Germany) and data analysis software packages like Orange Canvas (University of Ljubljana, Slovenia). Figure 1.8 shows a representation of RBC on an example IR spectrum.

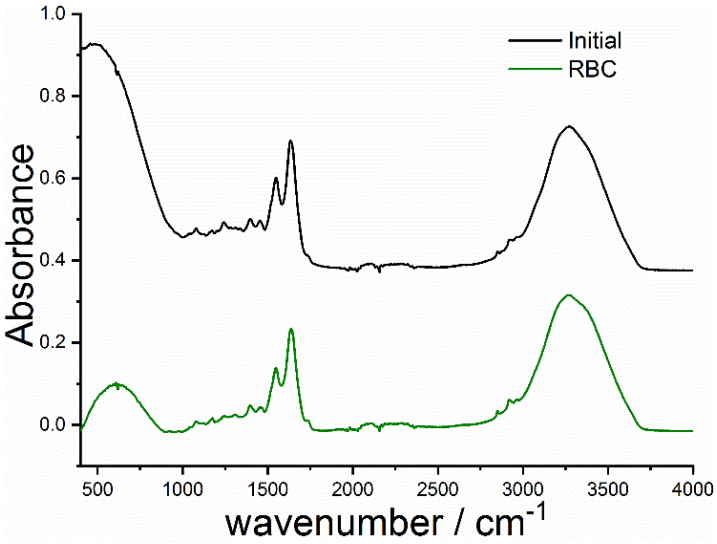

Figure 1.8: A representation of IR spectra of Lamb pre- (black) and post- (green) rubberband baseline correction.

1.5.3.4 Baseline offset correction

Baseline offset can be employed in correcting baseline differences for slight offset in intensities. It also compensates for other spectral preprocessing techniques which might generate negative intensity values. This is particularly effective when constraints such as non-negative variables are required during chemometric analysis. For baseline offset, the value of the lowest point in each spectrum is subtracted from all the variables, resulting in all the variables having a positive value and zero (0) being the minimum point [27]. An example of baseline offset correction on a spectrum is given in Figure 1.9.

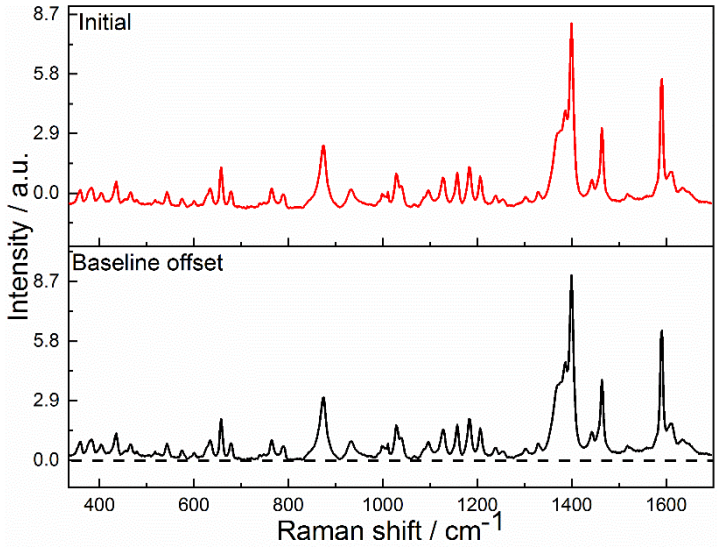

Figure 1.9: Example of Raman spectrum of piroxicam monohydrate pre- (red) and post- (black) baseline offset correction. The baseline corrected spectrum has all its variables above zero (0).

1.5.3.5 Shape baseline correction

The shape baseline function is unique to the WITec Project Five software (WITec, Ulm, Germany) and highly effective for subtraction of fluorescent signals. The baseline function utilizes a rounded shape to map out nonlinear baseline in a given spectrum on a pixel-by-pixel basis, which is then subtracted. The selection of shape size is important to avoid loss of spectral information. A smaller shape size will subtract more details from the spectrum, unlike larger shape size, which subtracts more rough shapes. An example of spectral baseline correction using the shape function is shown in Figure 1.10.

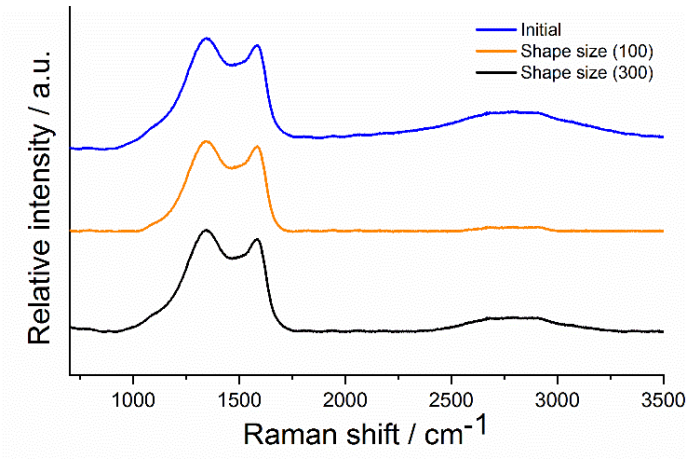

Figure 1.10: Example of Raman spectra of graphite pre- (blue) and post- (orange and black) shape baseline correction. The figure shows a substantial loss of information between 2500 – 3000 cm^{-1} while using a smaller shape size (100) as compared to a larger shape size (300).

1.5.4 Normalisation and scaling

Normalisation aims at correcting disparity in intensities across spectral data acquired using the same experimental parameters but under slightly different conditions. These differences in intensity can arise from various sources including variation in sample surfaces, change in optical path length and laser intensity [32]. Correction of spectral intensities can be performed using a range of methods including standard normal variate (SNV) transformation, mean normalisation (MNorm) and multiplicative scatter correction (MSC), which also rectifies baseline variations.

1.5.4.1 Mean normalisation

Mean normalisation computes new variables for each row (single observation) in a data matrix by diving the original spectral variables by the mean of the spectrum. For each point along a spectrum, a new variable is computed by dividing the original variable in the point x_j by the mean of the observation, $\bar{x}_i$. This technique can easily be adopted for spectral data with negative variables by computing the absolute values for each point as shown below

$$x_{new}(i,j) = \frac{x_{old}(i,j)}{|\overline{x}(i,\cdot)|} \quad 1.7$$

where $x_{new}(i,j)$ refers to the new variable for each point in *i* sample and *j* variable, $x_{old}(i,j)$ is the old variable for that point and $\overline{x}(i,\cdot)$ is the mean of the variables in an observation (the spectrum of the sample) [27].

1.5.4.2 Standard normal variate (SNV) transformation

SNV was introduced to reduce multiplicative effects of scattering and particle size, as well as the disparity in global intensities of spectral data. The SNV transformation centers and scales an individual spectrum using their standard deviation as described in the equation [33]

$$x_{i,j}^{SNV} = \frac{(x_{i,j} - \bar{x}_i)}{\sqrt{\frac{\sum_{j=1}^{p}(x_{i,j} - \bar{x}_i)^2}{p-1}}} \quad 1.8$$

where $x_{i,j}^{SNV}$ is the element of the transformed spectrum, $x_{i,j}$ is the corresponding original element *i* at variable *j*, $\bar{x}_i$ is the mean of the spectrum *i*, and *p* is the number of variables in the spectrum (wavenumbers). This equation can be simplified as

$$x_{i,j}^{SNV} = \frac{(x_{i,j} - \bar{x}_i)}{\sigma} \quad 1.9$$

where the denominator, σ is the standard deviation in intensity of the spectrum. A representation of the SNV transformation on infrared spectra of red meat sample is shown in Figure 1.11.

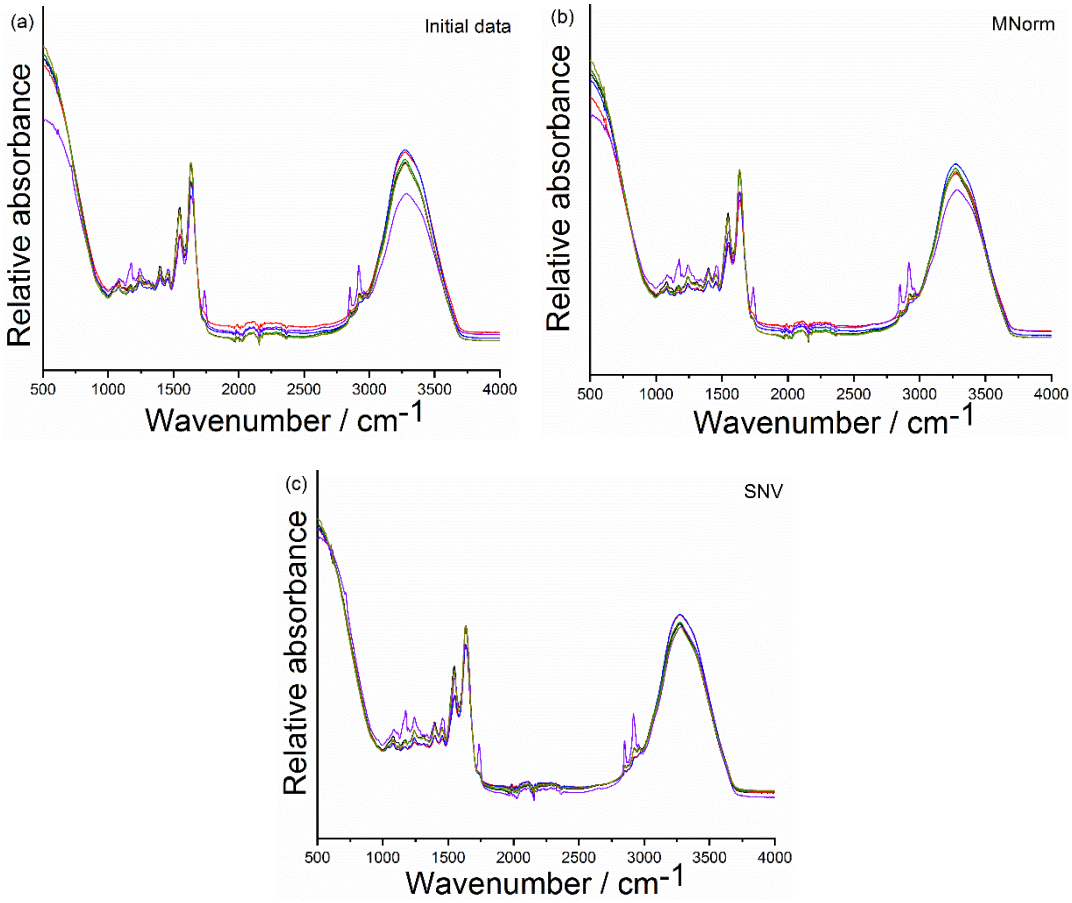

Figure 1.11: Representation of spectral data pre- and post-normalisation (a) initial spectral data (b) mean normalised data (c) standard normal variate transformed data

1.6 Multivariate analysis (chemometrics)

Spectroscopic investigation of systems studied in this thesis yielded large spectral data sets with numerous data points. The use of univariate data analysis (single or few peaks) for identification, classification or quantification of these spectra is not ideal owing to the significant loss of information across the spectral regions not selected. A more global approach utilizing wider spectral regions was facilitated by the use of methods that incorporate various mathematical and statistical techniques. This approach is referred to as multivariate analysis (MVA) or chemometrics.

The term ‘chemometrics’ was first coined in early 1970s by Professor S. Wold in Swedish as ‘kemometri’ and was described as “the art of extracting chemically relevant information from data produced in chemical experiments”. The idea of chemometrics is analogous to other ‘metrics’ including econometrics and biometrics. It basically involves the application of mathematical and statistical treatments to chemical data [34]. Furthermore, chemometric techniques is designed to establish relationships between different measurements from a chemical system or process through the application of a range of mathematical and statistical methods [35]. MVA techniques have found widespread use in spectroscopy providing a way of finding similarities, visualisation of variations and quantification of spectral data.

Chemometric techniques can be employed for both qualitative, quantitative and classification type analysis. Qualitative analysis usually entails identification of characteristic bands in a spectral data to determine the composition and changes occurring in the sample of interest. Principal component analysis (PCA) has been widely applied for this purpose [36]. Classification methods can be either supervised or unsupervised, with unsupervised methods requiring no prior knowledge of the dataset. Classification methods include linear discriminant analysis (LDA) [37, 38], factor analysis [39], cluster analysis [40], partial least square discriminant analysis (PLSDA) [41-43], *K*-nearest neighbors [44], hierarchical cluster analysis [45], soft independent modelling of class analogy [36] and support vector machine classification (SVMC) [46]. Quantitative analysis can be performed using tools like multilinear regression (MLR), principal component regression (PCR) [47], partial least square regression (PLSR) [13, 48] and support vector machine regression [49]. Multivariate curve resolution (MCR) can be employed for both classification and regression analysis [50, 51].

In general terms, most chemometric techniques like those mentioned above aim at separating chemically relevant data from noise, where noise refers to unrelated data or data which cannot be modelled. This is presented in simple terms in the equation

$$X = M + E \qquad 1.10$$

where X is the raw noise containing chemical data, M is the chemical model comprising the chemically relevant information and E is the noise which cannot be modeled [34].

1.6.1 Principal component analysis (PCA)

Principal component analysis is a bilinear modelling technique that provides an interpretable overview of the main information contained in a multidimensional table. It aims at extracting important information in large spectral data, reducing the dimensionality of the dataset by keeping only the important information. As such, PCA facilitates the simplification and analysis of the structure of the dataset while minimizing loss of important information [52, 53]. PCA accomplishes this by estimating new uncorrelated variables which maximizes variance referred to as principal components (PCs) as described in the equation

$$X = T \cdot P^T + E \qquad 1.11$$

where X is the original dataset (matrix of i samples by k variables), T is the scores matrix, P^T is the loadings matrix and E is the residual matrix which is the part of the matrix not explained by the PCA model. The scores, T, represent the reduced dimensionality for each sample, highlighting the sample differences or similarities. The loadings matrix highlights the information being described in each PC, that is, the variable contributions and correlations.

PCA finds the direction in space along which the distance between data points is largest. These directions are computed iteratively, such that the first PC (PC1) carries the most information (or explains the most variance). The second PC (PC2) is computed under the constraints of being orthogonal to PC1, explaining variance not described in

PC1. Calculation of subsequent PCs are performed in like manner. A geometric representation of PCA model in three dimensions is shown in Figure 1.12.

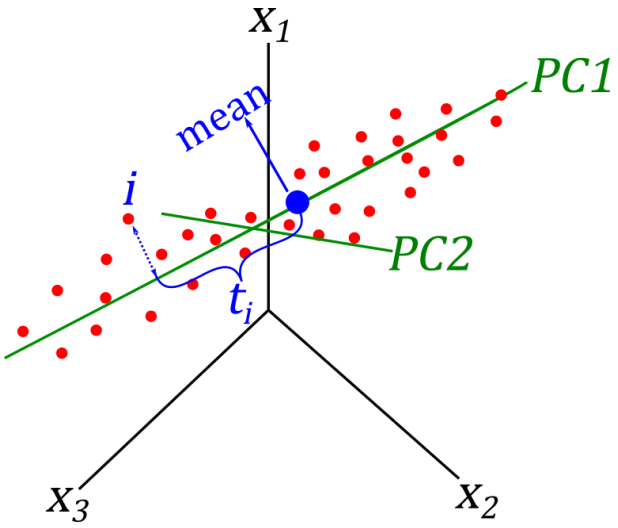

Figure 1.12: An illustration of the projection of a data matrix, *X*, with *N* points and *k*-variables using PCA. This figure shows calculation of PCs in a 3-variable dataset. The PC score of an object (t_i) is the distance from the mean along PC line.

1.6.2 Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) method is a supervised feature extraction tool used for classifying patterns between two or more classes. The LDA algorithm provides linear transformation of *n*-dimensional feature vectors (or samples) into an *m*-dimensional space called discriminant variables (where $m < n$), such that the distance between-class variance is maximized relative to the within-class variance [38]. LDA aims at finding optimal boundaries between classes, with subsequently unknown samples classified according to Euclidean distance.

LDA classifies samples by assuming normal distribution of samples as well as equal covariance matrices among the *q* number of classes. The LDA discriminant function can classify the data *x* to a class *p* using Bayes rule as shown [54]

$$\delta p(x) = x^T \Sigma^{-1} \mu_p - \frac{1}{2} \mu_p^T \boldsymbol{\Sigma}^{-1} \mu_p + log_{10}(\pi_p) \quad 1.12$$

where x is the data to be classified, μ is the mean of x variables and $\boldsymbol{\Sigma}$ is the covariance matrix for class p and π is the likelihood of data x belonging to class p.

However, with unequal covariance between classes, the shape of the curve separating the group is not linear; hence, a quadratic discriminant function is applied as shown in the expression

$$\delta p(x) = -\frac{1}{2} log_{10} |\boldsymbol{\Sigma}_p| - \frac{1}{2} (x - \mu_p)^T \boldsymbol{\Sigma}_p^{-1} (x - \mu_p) + log_{10}(\pi_p) \quad 1.13$$

1.6.3 Support vector machine classification (SVMC)

Support vector machine classification is a supervised method that utilizes linear and non-linear functions to map data, *X*, into higher dimensional space, allowing for classification of samples based on a set of predefined classes. The SVMC algorithm identifies the hyperplane with the maximum margin in the *n*-dimensional space (where *n* is the number of spectral features) and employs this to correctly classify the multivariate data into predefined groups [46, 55]. Further prediction of unknown samples can be performed using the created model and where class assignment is not possible; the sample is assigned to the nearest fit. An illustration of the SVMC support vector hyperplane is shown in Figure 1.13.

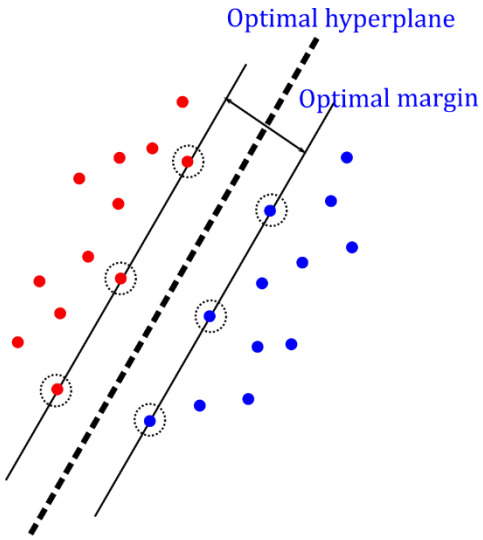

Figure 1.13 An illustration of SVMC classification in 2-dimensional space. The support vectors, market in black circles, define the optimal margin, which is the margin of largest separation between the two classes are determined.

1.6.4 Inverse regression

Inverse regression attempts to find the best relationship that correlates *X*-variables (independent variables) to *y*-variables (dependent variables), with the assumption that new samples will have the same relationship/function. Basic inverse regression methods aim at estimating the regression coefficient, β and α that solves Equation 1.14.

$$y = X \cdot \beta + \alpha + \varepsilon \qquad 1.14$$

where y is the vector of the reference values, X is a matrix of the independent variables, β is the matrix of coefficients, α is an intercept scalar and ε is the residue noise [56].

1.6.5 Multiple linear regression (MLR)

Unlike simple linear regression that deals with finding relationship between one independent variable and one dependent, multiple linear regression (MLR) attempts to model a relationship between two or more independent variables (*x*) and a response variable (*y*) by fitting a linear equation to the observed data. The *y*-variable is described as a polynomial function of the *x*-variables and the noise that cannot be modelled as shown in the equations:

$$y = f(x_1 + x_2 + x_3 \ldots + x_n) + \varepsilon \quad 1.15$$

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \varepsilon \quad 1.16$$

where, β_i $(i = 1, 2, 3, \ldots, n)$ are the regression coefficients.

The value of the coefficient can be estimated using ordinary least squares by taking the derivatives of the sum of the squared difference between the predicted $\hat{y}$-value and true *y*-value resulting in Equation 1.17.

$$\beta = (X^T X)^{-1} X^T y \quad 1.17$$

MLR is effective in cases where the number of variables is small and not significantly correlated (assumes orthogonality between the columns in the *X*-matrix). This makes MLR suboptimal for spectroscopic data, hence, the emergence of alternative regression models [57].

1.6.6 Principal component regression (PCR)

The PCR method is a combination of PCA and MLR for quantitative purpose. PCA is first performed on the *X* matrix, yielding the principal components. These PCs are further regressed against the Y matrix using MLR, as such, removing the collinearity requirement that is associated with the application of MLR. PCR decomposes the *X*-matrix using PCA as described in Equation 1.8, then regresses the PC scores against the *y*-matrix using MLR as described

$$y = T \cdot b + f \qquad 1.18$$

where y is the vector containing the response variables (concentrations) to be modelled, T is the scores matrix from the PCA, b is the vector containing coefficients and f is the y-residual from the MLR model [58].

1.6.7 Partial least square regression (PLSR)

Partial least square regression method is the variation of PLS that is used primarily for quantitative studies. The PLSR method aims at decomposing both the *X*- and *Y*-matrices simultaneously to identify the latent variables (or factors) in *X* which best predicts the latent variables in *Y* [57]. PLSR is calculated using the least square algorithm and similar to PCR. However, unlike PCR which performs a PCA first on the *X*-matrix and then regresses the scores (*T*) against the *Y*-matrix, PLSR decomposes both *X*- and *Y*-matrices simultaneously to obtain the maximized covariance between *T* (scores of independent block *X*) and *U* (scores of dependent block *Y*). This calculation can be performed as described in Equation 1.11 and 1.19 respectively [27, 59].

$$Y = T \cdot Q^T + F \qquad 1.19$$

where *Y* is the matrix containing the response vectors such as concentration, *T* are the scores vector, *Q* is the loadings matrix and *F* is the *Y*-residual errors the PLS model.

In comparison to other least square algorithms such as multilinear regression (MLR), PLSR is more robust to collinearity, noise and high dimensionality in data. One of the advantage of PLSR over PCA/PCR is in the selection of the latent variables (LVs), which is a covariance between the *X* and *Y* matrices, as such, accounting for both variability and correlation. However, it is important that appropriate LVs be selected to account for the optimal variances within the dataset, and yet not overfit the model [35].

1.6.8 Partial least square discriminant analysis (PLSDA)

The PLSDA method is a variation of PLSR employed for classification of samples. PLSDA is a linear and parametric technique that identifies LVs, which emphasizes maximal covariance between the X and Y matrices. However, unlike PLSR, the response matrix Y, comprises of vectors of zeros (0) and ones (1), which is used in assigning class membership [42]. For two-class discrimination, class members are assigned a dummy variable of one (1) and non-members assigned a variable of zero (0).

The PLSDA output gives a value for each sample with the associated standard deviation. These values can be used in determining class assignments or further employed as probability density functions in estimating the posterior probabilities of samples using Bayesian rule.

1.6.9 Multivariate curve resolution (MCR)

Multivariate curve resolution is employed in solving the mixture analysis problem. MCR is a mathematical technique that resolves multicomponent mixture systems through a bilinear modelling of the pure component contributions. It is applied in spectroscopy

for decomposition of spectral mixtures (*X*) into a product of the component spectra (*S*) and concentration profiles (*C*) [51, 60]. The MCR bilinear model decomposes mixed signal as

$$X = C \cdot S^T + E \qquad 1.20$$

where *X* is the original data matrix such as spectral mixture, *C* is the concentration of the components, S^T is the component spectra and *E* is the error matrix containing the unmodelled part of the data matrix [27, 60].

The transformation of Equation 1.20 can lead to multiple pairs of *C* and S^T due to the rotational and intensity ambiguity. Resolving these ambiguities requires the introduction of constraints during MCR analysis. These constraints ensure that the resolved profiles fulfil some preselected properties. Common applied constraints include non-negativity, unimodality and closure. The non-negative constraints ensure that the concentrations and spectra of the pure components have variables greater than or equal to zero (0). Unimodality allows for the presence of one maximum in each pure profile. This is the situation of many monotonic reaction profiles shows either increasing or decreasing component. The closure constraint is applied to closed reaction systems where the sum of the concentration of all the species (each row of the *C* matrix) involved equals a constant value for all samples [27, 60].

1.7 Model optimisation and assessing quality of calibrations

For any dataset, different pre-processing and model parameters would be tested to determine an optimum solution to the particular problem. One approach to testing this is adopting the cross-validation resampling technique.

1.7.1 Cross-validation

The cross-validation technique splits the calibration dataset into k number of groups (k-fold cross-validation). These groups can be selected at random or systematically. The groups are individually selected to form the test set while the rest of the data form the train set, which is employed in estimating the regression coefficients such as β and α (Equation 1.14). This is then utilised in estimating the ŷ-values of the test set. This process is repeated until all samples have been tested and their estimated values generated. The optimum model is commonly assessed using the minimum Root Mean Squared Error (RMSE).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y)^2}{n}} \qquad 1.21$$

where $(\hat{y} - y)$ is the difference between the predicted values and the actual reference values.

Cross-validation is useful for finding the optimum model parameters and algorithm to yield the minimum RMSE for a typical dataset, however, it is not a good measure of the model's ability to predict unknown samples. This is because samples forming the test set in cross-validation also belong to the training set. It is important to test the model against an independent test set. In such case, the RMSE would be renamed RMSECV or RMSEP to indicate cross-validation or prediction (independent test set).

1.7.2 Outlier detection

Sample outliers are points that do not follow the general trend of most (or all) other points in the model. These points might be problematic to the overall accuracy of the calibration model, as such, need to be investigated and if required removed to ensure

model stability and improved accuracy [61]. Simple outlier detection can be performed on visual inspection of the scores plot; samples that are far from the rest might suggest that they are outlier. Assessment of the RMSE on cross-validation with and without these samples could provide more insight into their status as outliers. A more sound approach will be using statistical tools such as Q-residual, F-residual, leverage, Hotelling T^2 statistics and custom local outlier factor [62].

For the purpose of this thesis, outlier detection was performed using the influence plot. This is a plot of the leverage (Hotelling T^2 statistics) against the residuals. Sample residuals (Q- and F-residuals) are derived from the error matrix and relates to the unexplained variance in a sample. It is common for samples to have some degree of unexplained variance; however, large unexplained variance suggests that the model poorly describes the sample. If this sample is not influential, leaving the sample might not be problematic to the overall model. The Hotelling T^2 relates to the explained variance in the model, with higher value suggesting that the sample is well described by the model. This sample is usually problematic in the calibration, as it has a higher than average leverage or influence in the model. The most problematic are samples with high residuals and high influence. These samples are poorly described by the model and have higher than average influence in the model. Both residuals and Hotelling T^2 are calculated within a given confidence interval, samples outside the critical limit are probably outlier and should be removed prior to recalculation of the model. A representation of the influence plot is shown in Figure 1.14.

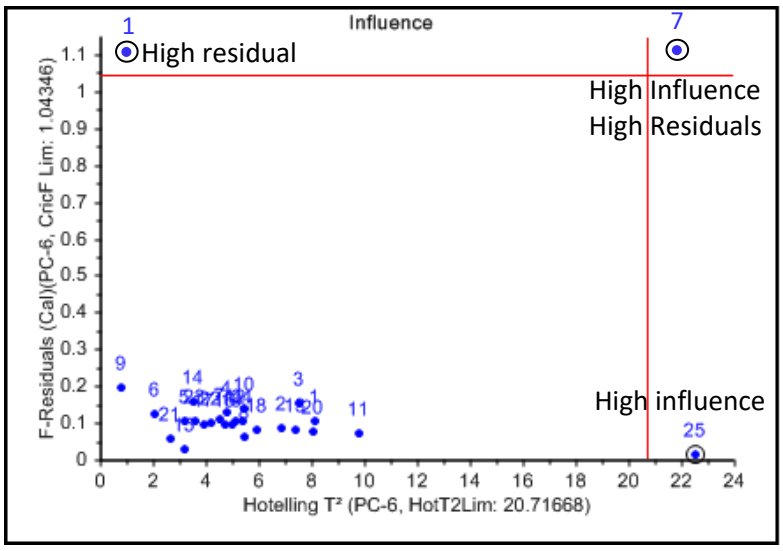

Figure 1.14: Influence plot highlight possible outlying samples in the model. Removal of these samples might improve the model stability and accuracy.

1.8 Summary

This chapter described the concepts of Raman and infrared spectroscopy alongside the various experimental considerations.

Spectral preprocessing techniques including noise reduction, removal of baseline, scaling and scatter correction were also discussed. Specifically, noise reduction techniques such as moving average and Savitzky-Golay were highlighted. Baseline correction tools included LBC, polynomial subtraction, RBC, shape function and baseline offset. Scaling and scatter correction method included MNorm and SNV.

The concept of chemometric techniques was introduced and a range of chemometric techniques used in thesis were also described. This included classification techniques such as PCA, LDA SVMC, PLSDA and regression methods such as PLSR, PCR and MCR.

Vibrational spectroscopy in combination with chemometric tools posit a powerful tool in analytical chemistry. Infrared spectroscopy in combination of chemometrics have been applied in food analysis [63-67], medicine [68-70] and pharmaceuticals [71, 72]. Raman spectroscopy in combination with a range of chemometric tools have also been applied for food analysis [73-78], pharmaceuticals [13, 79-82] and medical applications [83-85].

1.9 Thesis aims

The main aim of the thesis is the application of Raman and infrared spectroscopy in combination with chemometrics to analyse a range of samples. Each chapter in this thesis discusses specific projects and comprises published papers, with the necessary copyright permissions obtained. Owing to the variety of topics discussed in this thesis, an introduction into each project with the related important literature review is provided at the beginning of each chapter.

The following projects were discussed in this thesis:

- The use of Raman spectroscopy in combination with chemometrics to non-destructively discriminate red meat species, in particular beef, venison and lamb. The use of infrared spectroscopy, howbeit destructive owing to the use of the ATR system was also employed in combination with chemometrics for the discrimination of the red meat types.
- The quantification of meat quality parameters including pH, % intramuscular fat (IMF) and tenderness was assessed using Raman and infrared spectroscopy in combination with chemometrics. The correlation of spectral data to chemical measurements was believed to be possible using vibrational spectroscopic techniques. Fusion of Raman and infrared data at low, mid and high levels were proposed to highlight any improvement in model performance on correlation of spectral data to chemical measurements of pH, % IMF and tenderness.

- The use of Raman microscopy in combination with chemometrics to discriminate the phenotypes of T-cells and monocyte after incubation with media conditioned glioblastoma stem-cells (GSCs) showing different molecular background.
- The use of low frequency Raman spectroscopy in combination with chemometrics to investigate the in situ isothermal dehydration of crystalline hydrates, particularly piroxicam and theophylline monohydrates at different temperatures. This examined the simultaneous application of the low and mid frequency Raman region to identify the changes occurring as well as the timescales in which these changes were first observed.

Chapter 2

Analysis of intact meat

2.1 Acknowledgments

The samples discussed in this chapter were sourced and processed by AgResearch, New Zealand. The team at AgResearch, New Zealand also carried out the reference chemical measurements.

2.2 Results publication

Some of the results discussed in this chapter have been previously published:

Chima Robert, Sara J. Fraser-Miller, William T. Jessep, Wendy E. Bain, Talia M. Hicks, James F. Ward, Cameron R. Craigie, Mark Loeffen and Keith C. Gordon: Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy. *Food Chemistry* 2021, *343*, 128441.

My contribution to this paper included conceptualization, data collection, data curation, formal analysis, drafting of the original manuscript and correction of the reviewed manuscript.

2.3 Introduction

Meat continues to play a key role in human dietary requirements, with an approximate composition of 75 % water, 19 % protein, 3.5 % soluble non-protein, 2.5% fat and other micronutrients [86]. The Food and Agriculture Organisation of the United

Nations estimated that 330 million tonnes of meat was produced globally in 2017, with the average meat price increasing by nearly 9 % from 2016 [87]. This is indicative of the growth potential of meat and meat products industries. This growth potential has also increased the need for proper identification of meat and meat products, as their consumption impact on the overall health, safety, religious and ethical beliefs of consumers.

The growing consumer interest in high quality meat, and assurances on the integrity of meat products, have increased the need for identification and traceability of meat and meat products across the production chain [88]. Following the horsemeat scandal in Europe in 2013 [41], research relating to meat adulteration and discrimination has significantly increased [89-93]. These research studies are predominantly aimed at proffering alternative techniques for faster, cheaper, accurate and robust meat assessment.

Several techniques have been applied for the identification of meat origin and meat discrimination. These techniques include, but are not limited to, DNA-based molecular techniques such as polymerase chain reaction [94-96], DNA barcoding [97], DNA fingerprinting [98] and DNA polymorphism [99] or protein and fat-based techniques including electrophoresis [89], liquid chromatography – mass spectrometry [100], gas – chromatography [101, 102] and enzyme-linked immunosorbent assay [103]. Despite the advantages associated with the use of these techniques, there are some limitations to their applicability. These limitations include sample destruction, requirement for skilled personnel and expensive instruments as well as long analysis time [104]. Specific limitation associated with DNA and protein-based technique include the denaturation of key analyte markers [105] whereas low reproducibility is common with chromatographic fat-based techniques [106]. Owing to these limitations, there is a need to develop an improved, easy to use and reliable technique for meat discrimination.

Over the last decade, vibrational spectroscopy has garnered significant attention as an applicable alternative for meat assessment. Vibrational spectroscopy (Infrared and Raman spectroscopy) posit great advantages over traditional methods for meat discrimination and quality assessment [107-112]. These advantages include minimal sample preparation, rapid measurement / short analysis time, non-destructive, non-

invasive, cost-effective, increased sensitivity as well as the potential for inline/online integration [64]. In addition, water molecules have negligible Raman signature making Raman spectroscopy less susceptible to interferences by water molecules in either meat sample or the environment in which the measurement is occurring.

Raman spectroscopy is based on the principle of inelastic scattering reported by C. V. Raman [6]. On interaction of a sample with light, a small portion of the incident light is scattered, with the resulting scattered light having an energy, which is different from the incident light. The measurable energy change between the incident and scattered light is referred to as Raman shift. This Raman shift generates molecular specific spectra from the sample under analysis; which can be used for qualitative identification [113].

Infrared spectroscopy has also been investigated for meat assessment [114]. In contrast to Raman scattering, infrared spectroscopy is based on the resulting transition between the ground and excited states upon absorption of electromagnetic radiation by matter. Detailed description of Raman and infrared techniques have been presented in Chapter 1.

A number of studies have reported on the potential of Raman [41, 64, 86, 88, 106, 112], near infrared [115, 116] and mid infrared spectroscopy [117, 118] in combination with chemometrics to discriminate meat species. Despite the successful discrimination presented in these reports, some of the studies utilized extracted fat and not the intact meat . This tends to increase the analysis time, margin for error and unsuitability for inline integration [41, 106]. The chemometric techniques employed in this study included principal component analysis (PCA) for exploratory analysis; partial least square discriminant analysis (PLSDA) and support vector machine classification (SVMC) for developing classification models.

For the purpose this study, a one versus all PLSDA was adopted, with dummy variables used in assigning predefined classes. Three PLSDA models were built for the discrimination of each of the meat specie. In building the PLSDA models, class members were coded with the dummy variable 1 ($class\ w_1$) and 0 for non-members ($class\ w_0$). The PLSDA model was calculated using a k-fold (k = 3) systematic cross-validation

technique. The optimum PLSDA model was determined against the minimum root mean square error of cross-validation (RMSECV) of the training set. The PLSDA model prediction values formed the probability density function, which were further utilized in estimating the posterior probabilities of samples belonging to each meat class. The a priori probabilities $P(w_0)$ and $P(w_1)$ were calculated from the training set assuming that meat classes had equal number of participating samples. The Bayesian decision rule for class assignment was based on maximum probability value as shown in the equations

$$P(w_c|\hat{y}_u) = \frac{P(\hat{y}_u|w_c) * P(w_c)}{P(\hat{y}_u)} \quad 2.1$$

The denominator can be obtained using

$$P(\hat{y}_u) = P(\hat{y}_u|w_0) * P(w_0) + P(\hat{y}_u|w_1) * P(w_1) \quad 2.2$$

where, *c* refers to the classes, $P(w_0)$ and $P(w_1)$ are the a priori probabilities and can be calculated from the training dataset assuming the number of samples of each class is representative of the entire population, $P(\hat{y}_u|w_0)$ and $P(\hat{y}_u|w_1)$ are the probability density functions for the classes w_0 and w_1 which can be obtained for the calculated model [119].

The Bayesian decision rule attempts to minimize error in class assignment, $P(error|\hat{y}_u)$. The probability that a sample belonging to class w_0 will be wrongly assigned is given by the probability that the sample with prediction $\hat{y}_u$ actually belongs to class w_1, *i.e.*, $P(w_1|\hat{y}_u)$. Likewise, the probability of a wrong decision for class w_1 is given by $P(w_0|\hat{y}_u)$ [119]. As such in assigning unknown samples to a class using Bayesian decision rule:

$$decide\ class\ w_0\ if\ P(w_0|\hat{y}_u) > P(w_1|\hat{y}_u);\ otherwise\ decide\ class\ w_1 \quad 2.3$$

The SVMC model was developed using both a linear and non-linear kernel function. This was to highlight any possible improvement by the non-linear kernel to capture complex spectral relationships. Samples were assigned to predefined classes prior to the model development.

Model performances were assessed with respect to the prediction accuracy, sensitivity and specificity. Sensitivity is the ability of a model to correctly identify true positives. In this case, samples belonging to a specific meat class. Specificity relates to the ability of a model to identify true negatives, *i.e.*, samples not belonging to a meat class. Accuracy is the ability of the model to correctly classify the samples into their assigned classes. These parameters can be calculated using Equations 2.4 to 2.6, with a diagrammatical representation presented in Figure 2.1.

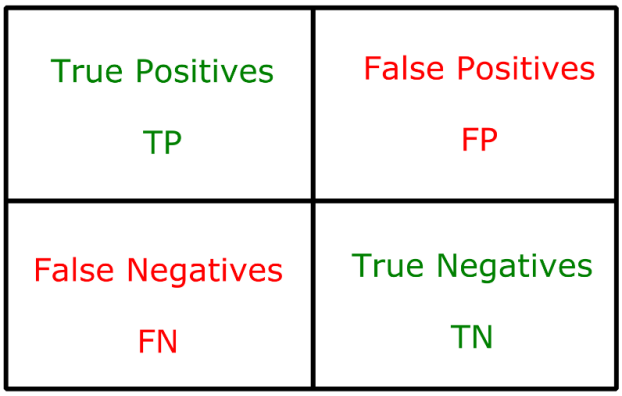

Figure 2.1: A 2x2 table illustrating the basis of deriving the sensitivity, specificity and accuracy of predicted classes.

$$Sensitivity = 100\left[\frac{TP}{TP + FN}\right] \qquad 2.4$$

$$Specificity = 100\left[\frac{TN}{FP + TN}\right] \qquad 2.5$$

$$Accuracy = 100\left[\frac{TP + TN}{TP + FP + FN + TN}\right] \quad 2.6$$

where, *TP* is the true positives, *i.e.,* correct classification of samples to their assigned class, *FP* is the false positives, *i.e*., wrong classification of a sample into a class, *FN* is the false negatives, which wrongly indicates that a sample is not a member of a class and *TN* is the true negatives which correctly indicates that a sample is not a member of a class [120].

In this work, an FTIR and fibre optic - coupled Raman spectrometer was employed for fast discrimination of intact beef, venison and lamb samples. To minimize the challenge of subsampling, the meat samples were scanned on a raster system allowing for representative sampling during measurement using Raman spectrometer. This was not possible with the FTIR setup, as such multiple spot measurements was implemented. The acquired spectral data were pre-processed using a combination of techniques and analysed with three different chemometric tools including PCA, PLSDA and SVMC.

2.4 Materials and methods

2.4.1 Sample collection and preparation

Samples from three species representative of the New Zealand red meat sector were obtained from AgResearch Invermay, New Zealand. A total of ninety beef (*Bos Taurus*), lamb (*Ovis aries*) and venison (*Cervus elaphus scoticus, hippelaphus and pannonensis*) (30 samples per meat specie) were used in this study.

The venison samples were obtained from farmed red deer. The deer included a mixture of European sub-species (*Cervus elaphus scoticus, hippelaphus and pannonensis*). Four different stags sired the males and seven different stags sired the female. The animals were farmed with 50 % of their diet on an energy basis provided using palm kernel expeller (PKE) and the other 50 % from pasture (rye grass dominant,

with some white clover). The males and non-pregnant females were 11-months and 23-months old, respectively, at the time of slaughter. The ram lambs (*Ovis aries*) were sourced from the AgResearch Progeny Test flock which comprised of a mixture of pure maternal and maternal cross terminal genetics. They were approximately six months old at the time of slaughter. The beef samples were obtained from *Bos Taurus*, with the left and right striploin and inside round (*M. Longissimus dorsi and M. semimembranosus*) collected for various reference and spectroscopic measurements. The muscles were grouped based on their pH and subjected to either ageing or non-ageing treatment. The meat samples were collected from the left and right-hand side loins of the animal carcasses to account for differences originating from the animal sides. Meat cuts of 100 mm x 70 mm were collected from each muscle and wet-aged by sealing in a vacuum bag. The samples were stored below -17 $^{\circ}$C for 1-2 weeks prior to Raman measurement.

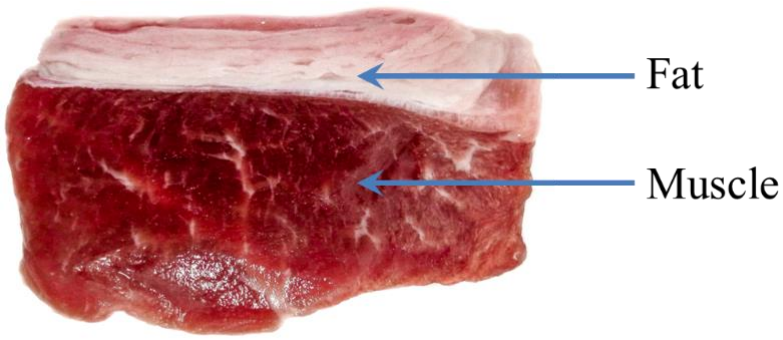

Figure 2.2: A representation of the meat sample.

2.4.2 Instrument and sample measurement

2.4.2.1 Reference measurement

pH measurements were obtained using a portable meat pH meter - HI99163 (123 Hanna Instruments, Mediray, New Zealand). pH point measurements were taken after the subcutaneous fat was removed from the M. Longissimus dorsi and M. semimembranosus.

This was further cut into sections and the median pH (n = 5) was used as the pH in the subsequent modelling. The meat samples were freeze-dried prior to percentage intramuscular fat content (% IMF) analysis.

Reference % IMF analysis was performed using gas chromatography, as outlined by Craigie et al. [121]. The % IMF analysis was performed using a direct trans-methylation technique involving the extraction and esterification of lipids in a single step. The team at AgResearch, New Zealand carried out the reference chemical measurement.

2.4.2.2 Raman Spectroscopy

Raman spectral data were collected using a Wasatch Photonics (WP) 1064 nm Raman spectrometer (Wasatch Photonics, USA) equipped with TEC cooled G9214-512S InGaAs array Hamamatsu detector (at -15 $^{\circ}$C) and fibre optics probe for routing the excitation light from the laser and collected light to the spectrometer. The Raman probe tip can be adjusted to the samples and has a spot size of 1 mm diameter. The probe consists of hollow optical fibre of 600 µm outer diameter and matched numerical aperture to the spectrometer input (NA = 0.36), thus reducing loss of light within the fibre optics. The probe is connected to the laser and spectrometer using an SMA 950 adaptor. Measurements were acquired with a Dash spectrometer control software v2.1 over a spectral window of 313-1895 cm^{-1} and a spectral resolution of 6 cm^{-1}. The instrument was calibrated using 1,4-bis(2-methylstyryl) benzene.

The meat samples were measured at 450 mW laser power, 15 s integration time x 12 co-added scans. Each 12 s scan had a background (dark scan) subtracted. For representative sampling, the 100 mm x 70 mm meat cuts were scanned on a raster setup with each scan extending over an area of 50 mm x 30 mm per portion. The raster system comprised of two UTS100CC linear motorized platforms (Newport Corp. Irvine, CA, USA) and ESP301-3G 3-axis controller (Newport Corp. Irvine, CA, USA) was used for positioning the sample. All measurements were performed in triplicate to obtain a representative sample and avoid the limitation of sub-sampling associated with spot

measurements. The acquired spectral data were subsequently averaged to three spectra per sample prior to analysis.

2.4.2.3 Infrared spectroscopy

Sample measurement was performed using a Vertex 70 (Bruker Optics, Ettlingen, Germany) spectrometer with a Platinum attenuated total reflectance (ATR) diamond accessory (GladiATR, Pike Technologies, Madison, USA) equipped with a wide range of room temperature DLaTGS Lanthanum α Alanine doped TriGlycine sulphate) detector. Spectra information were acquired with OPUS software-version 7.5 (Brucker Optics, Ettlingen, Germany) over 400 - 4000 cm^{-1} spectra region with sample scan-time of 32 s and spectral resolution of 4 cm^{-1}. Background spectra were collected from the cleaned blank ATR crystal prior to the presentation of each sample replicate. Each meat sample was divided into three portions; and for each portion, five measurements were taken.

2.4.2.4 Spectral Processing

To account for instrument variation, sample presentation, environment changes and background interferences over the duration of measurement, spectral data were preprocessed using a range of techniques. Baseline correction was applied to reduce the effect of fluorescence and sample focus; smoothing was performed on the Raman data to reduce noise, and standard normal variate transformation was performed to account for intensity variation across measurements. The background (dark scan) subtracted Raman spectra were preprocessed in Orange Canvas (University of Ljubljana, Slovenia) and The Unscrambler X 10.3 (CAMO, Norway). All Raman spectra were baseline corrected using rubber band baseline correction (RBC) in Orange Canvas [122]. The baseline corrected spectra were normalized using standard normal variate (SNV) transformation and smoothing performed using the Savitzky-Golay smoothing filter (first order polynomial, 7-point filter window) for noise reduction without distorting the acquired signal.

The MIR spectral regions 3025 – 2760 cm^{-1}, 1780 – 1710 cm^{-1} and 1485 – 1055 cm^{-1} were independently baseline corrected using linear baseline correction (LBC) in The Unscrambler X 10.3 (CAMO, Norway). The baseline corrected spectral region were collectively normalized using standard normal variate (SNV) transformation to account for intensity variations across replicate measurements.

2.4.2.5 Multivariate data analysis

In this study, PCA was utilized as an unsupervised MVA for exploratory analysis whereas SVMC and PLSDA were both employed in building a supervised classification model for meat discrimination. PCA model development was performed on the entire sample set using a k-fold (k = 3) systematic cross validation procedure. Model optimization was carried out by the removing outlying samples and the PCA was then recalculated. The PLSDA and SVMC classifications were performed on the training set and the model validated against an independent test set. The training set comprised two-thirds $\left(\frac{2}{3}\right)$ of the sample set and the test set consisting of one-third $\left(\frac{1}{3}\right)$ of the sample set. The training and test sets selections were performed such that the samples are representative of the entire population. Selection was carried out such that each set had samples covering the % IMF range used in this study (samples were first grouped into categories depending on their % IMF value (low, intermediate and high fat). For example, in the training set, 20 samples each were selected from the sample category “low fat” (% IMF < 2.0), “intermediate fat” (% IMF = 2.0 - 3.0) and “high fat” (% IMF > 3.0) [123]. Similarly, 10 samples were selected from each of the category forming the test set.

2.5 Results and Discussion

The results and discussion are divided into two main components: initial exploratory analysis with PCA and classification of meat samples.

2.5.1 Exploratory analysis

The reference chemical measurements shown in Table 2.1 highlight the variability across the meat samples. The meat samples had a percentage intramuscular fat (% IMF) content of 1.0 – 2.0 (venison), 1.0 – 4.2 (lamb) and 0.7 – 3.2 (beef). The venison meat samples had the least intramuscular fat for the three meat species used in this study.

Table 2.1: Summary of pH and % IMF reference measurements for red meat species

Meat species	pH	% IMF
Venison	5.26 ± 0.07	1.26 ± 0.24
Lamb	5.96 ± 0.42	2.17 ± 0.95
Beef	5.61 ± 0.10	1.48 ± 0.65

2.5.1.1 Raman Spectral Data Analysis

Raman spectra were acquired over a spectral range of 313-1895 cm^{-1}, however only the selected fingerprint region (750-1750 cm^{-1}) was used for analysis. Figure 2.3 shows the mean ± standard deviation of Raman spectra of beef, venison and lamb samples. The composition of meat cut is influenced by numerous factors including animal genetics, breed, age at slaughter and rearing condition. The averaged spectra of the meat species show relatively little difference in their Raman signatures and are consistent with published spectra. The observed bands may be assigned based on the literature – this is summarized in Table 2.2. Broadly speaking the bands are either attributed to protein, amide bands (1660, 1316, 1267 cm^{-1}), amino acid residue features (1362, 1556 cm^{-1}) or are due to lipids (1451 cm^{-1}). The positons of the amide bands depend on the conformation of the polypetide backbone and intra- and intermolecular hydrogen bonds; as such, these bands can be associated to the secondary structures of protein [124]. The band at 1660 cm^{-1} corresponds to amide I band and does not overlap with the vibrational bands of other functional groups, as such can be employed in characterising the secondary structure of protein [125]. For example, insulin fibrils and globular proteins have been

characterised the position and intensity of amide I band [126, 127]. The bands in the region 850 – 933 cm^{-1} are due to both amino acid residues, the polypeptide backbone and fats [74, 86, 128-139]. Amide II has been reported to be observed around 1520 – 1570 cm^{-1} and is primarily related to C-N stretching, N-H bending and C-C stretching vibration modes. Amide II has a weak Raman band and nearly undetectable in non-resonant conditions, hence it was not observed in the meat spectra. Amide III band, observed around 1316 cm^{-1} overlap with vibrational modes of other functional groups including CH_2 and C-C, which complicates their use in secondary structure interpretation [140]. The band at 1362 and 1556 cm^{-1} have been associated to tryptophan whereas CH_2 and CH_3 bending vibrations are observed at 1451 cm^{-1} [132, 133].

Table 2.2: Summary of Raman band assignments for red meat species

Band (cm^{-1})	Origin	Assignment	References
911	Short chain lipid	C-C stretching	[136, 141]
850 - 933	Amino acid (glutamic acid, lysine) and polypeptide	C-CN stretching	[131, 136]
977	Protein	=C-H wag	[134]
1075	Amino acid		[142]
1175	Peptide backbone (secondary amine)	C-N and C-C stretching	[134]
1267	Protein	Amide III globular α-helix	[129, 143]
1316	Protein	Amide III fibrous α-helix	[86, 129, 144]
1362	Tryptophan	Indole ring	[131, 133, 137]
1451	Lipids	CH_2 and CH_3 bending	[74, 132, 137]
1502	Phenylalanine	-	[134]
1556	Tryptophan	-	[74, 129, 134]
1660	Peptide backbone	Amide I	[86, 131, 133]

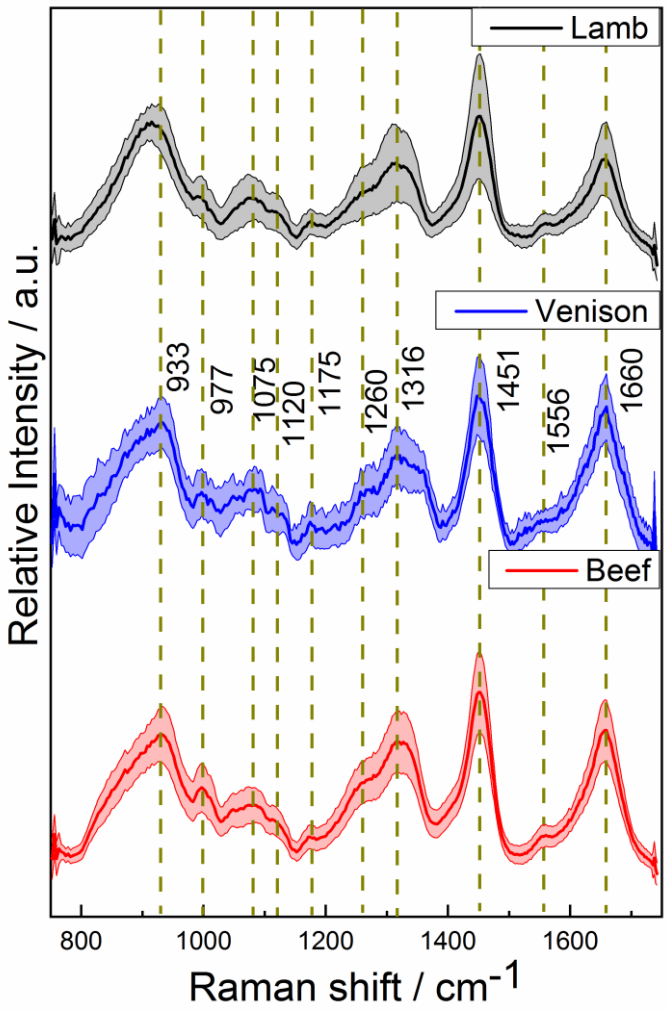

Figure 2.3: Mean ± standard deviation of Raman spectral data acquired from beef (red), venison (blue) and lamb (black) meat (n = 60 measurements per meat specie).

2.5.1.2 Infrared spectral data analysis

The MIR spectral data were collected over a spectral range of 400 – 4000 cm^{-1} and the selected regions (3025-2760 cm^{-1}, 1780-1710 cm^{-1} and 1485-1055 cm^{-1}) were used for further analysis. Figure 2.4 shows the mean IR spectral data alongside the standard deviation. The strongest bands were observed at 1546, 1634 and 3275 cm^{-1} whereas weaker bands appeared at 2924 cm^{-1}, 2850 cm^{-1}, 1736 cm^{-1}, 1453 cm^{-1}, 1396 cm^{-1}, 1242 cm^{-1}, 1170 and 1161 cm^{-1}. The band assignments suggest fat, protein and water

to be the major components of the red meat samples detected. A summary of the band assignments is presented in Table 2.3. The 1485-1710 cm^{-1} region was excluded due to interfering water signals. Also, the broad absorption features around 3100 – 3600 cm^{-1} corresponding to the O-H stretching vibrations of water molecules and NH stretching of amide in proteins, were not considered in predicting the pH and % IMF due to the large variability in the region as compared to others [145]. The mean ± standard of the FTIR spectral data is shown in Figure 2.4.

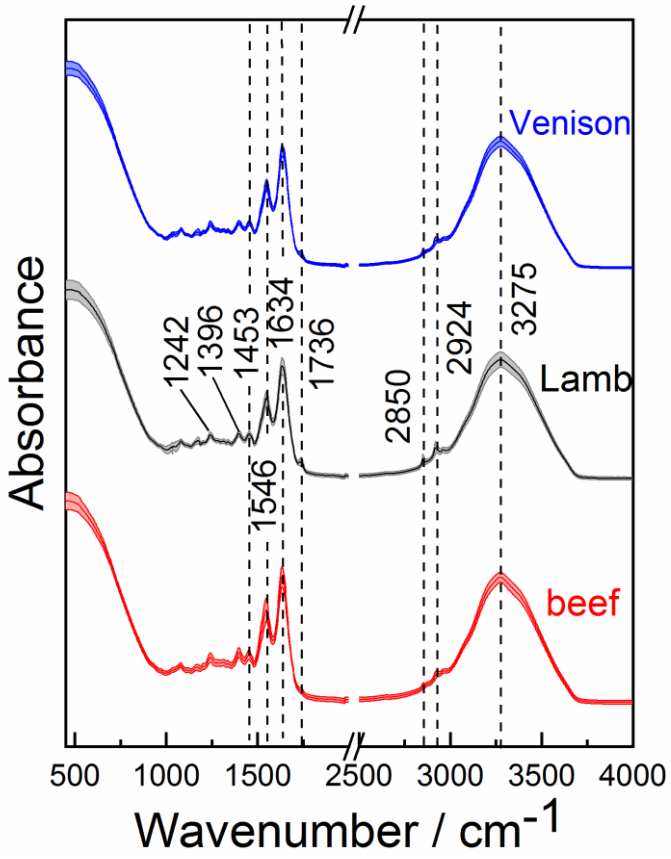

Figure 2.4: Mean ± standard deviation of spectral data acquired using the FTIR instrument (n = 60 measurements per meat specie). The FTIR show little deviation across meat measurements. The diamond crystal signal from 1900 – 2400 cm^{-1} has been excluded in this figure.

Table 2.3: IR Spectral bands assignment for the chemical functional groups in venison, lamb and beef according to literature

IR (cm^{-1})	Origin	Assignments	References
950 - 1200	Glycogen	C-O-C stretching	[145, 146]
1161	Triglycerides (lipids)	C-O stretching and C-H bending	[145, 146]
1242	Protein, Phospholipids	CN stretching and NH bending (amide III), PO_2^- stretching	[147, 148]
1396 / 1400	Lipids / Protein	COO^- stretching and amino side chains	[146, 149]
1453	Lipids	CH_2 and CH_3 scissoring	[36, 150]
1466	Lipids / Protein	C-H scissoring	[114]
1546	Protein	NH bending and CN stretching (amide II)	[151]
1634	Protein / water	CN stretching, C=C stretching (olefins), C=O stretching (amide I), O-H bending	[146, 148, 152, 153]
1736	Lipids	C=O stretching of ester, aldehyde, ketone: free fatty acids	[146, 148]
2850	Lipids	CH_2 symmetric stretching	[145, 153]
2924	Lipids	CH_2 asymmetric stretching	[145, 153]
3275	Protein	NH asymmetric and symmetric stretching	[145, 151]

2.5.1.3 Principal component analysis (PCA) – Raman data

Ninety meat samples (30 samples per meat specie) were used in creating the PCA model. The first two PCs explained the most variance across the meat samples (59 % of the variance). The PCA scores and loadings plot are shown in Figure 2.5 and 2.6, respectively. Despite minor overlaps, the PCA scores plot shows that the beef, venison and lamb meat are distinctly separate from one another. The venison samples clustered in the positive PC2 space while the beef and lamb samples clustered in the negative PC2 space. The beef and lamb samples despite being in the negative PC2 space are distinguishable in the scores plot, with the beef cluster seen in the negative PC1 space and the lamb cluster in the positive PC1 space.

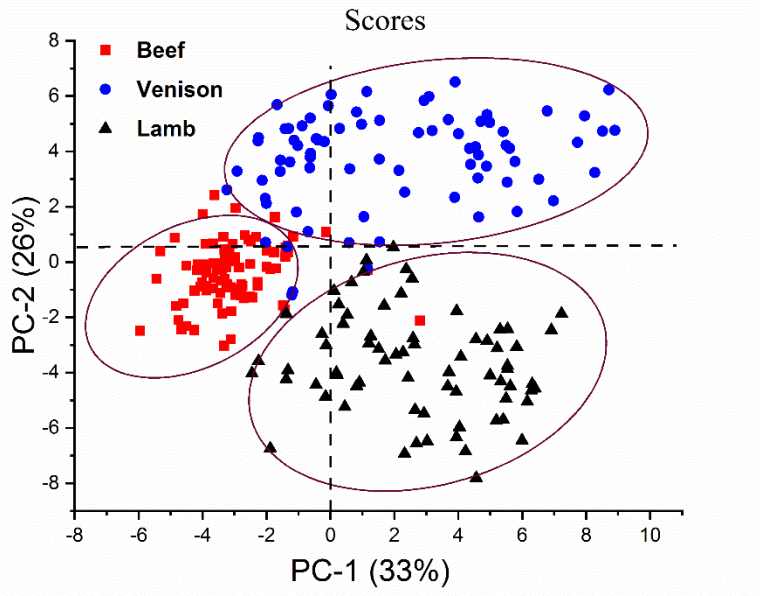

Figure 2.5: PCA scores plot showing separation of beef, venison and lamb meat samples.

The loadings plot in Figure 2.5 shows the Raman features responsible for the separation. Raman spectral features at 911, 977, 1267, 1316, 1362, 1451, 1556 and 1660 cm^{-1} were associated with the greatest variance. Of note is the distinct feature at 911 cm^{-}

[1] that is obscured in the normal Raman spectra by the broad number of features from 850 – 933 cm^{-1}. This feature has been reported previously as being associated with CH_3 rocking vibrational mode in lipids [128, 129] as well as α-helix structure in protein [126]. However, this peak was assigned to lipid because it was somewhat unique to the lamb samples, and might be indicative of the shorter chain lipids found mainly in lamb and goat meat. The separation along the PC2 axis is associated with the lipids (911 cm^{-1}) which is a stronger band in the lamb samples versus 1362 and 1660 cm^{-1} associated with tryptophan and amide I, respectively [74, 131, 133, 136, 137]. These features are more intense in the venison samples. The beef samples which cluster in negative PC1 space show stronger amide III bands at 1267 – 1316 cm^{-1} [86, 131, 134]. A summary of the peak assignments is presented in Table 2.2.

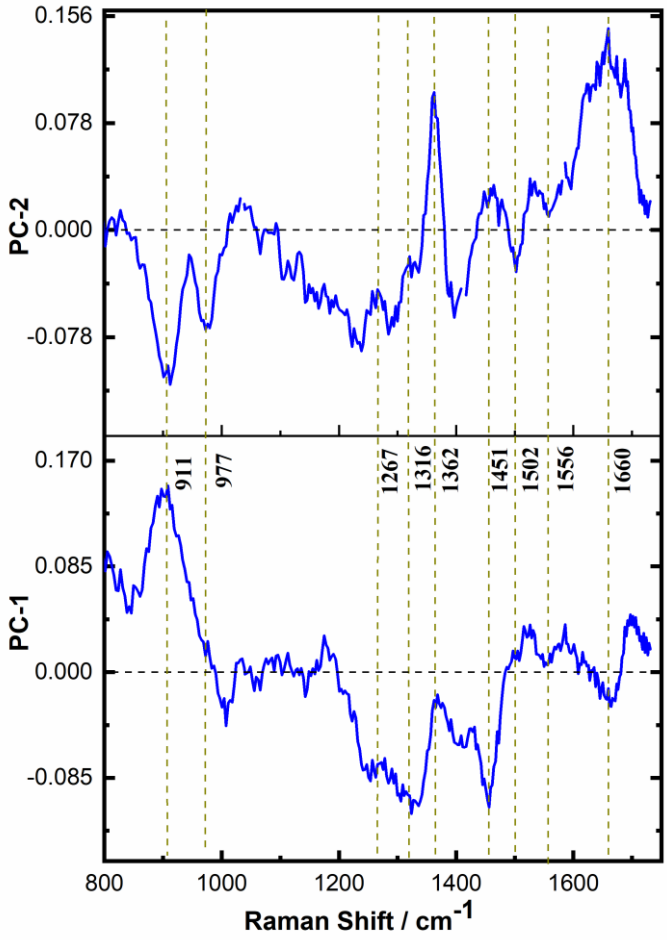

Figure 2.6: PCA loadings plot with Raman spectral features associated with variation in beef, venison and lamb meat.

2.5.1.4 Principal component analysis (PCA) – IR data

The PCA performed on the spectral acquired using the FTIR highlight the variability of spectral data across the meat samples. Despite the higher SNR of the IR spectra compared to the Raman, the scores plot shows greater overlap of the meat types. This is possibly due to subsampling from spot measurements reiterating the heterogeneity of meat samples. Clusters of the meat species can be somewhat distinguished in the first two PCs accounting for 85 % variance across the dataset. The PCA scores and loadings plots is shown in Figure 2.7 and 2.8, respectively. From the scores plot, majority of the beef samples clustered in the positive PC1 space whereas lamb clustered in the negative PC1 space. The venison samples are distinguishable in the PC2 space, clustering in the positive PC2 space whereas majority of the beef samples clustered in the negative PC2 space. The overlap observed in some of the samples across the PC spaces could be attributed to variability within samples of the same meat group owing to the heterogeneity of meat samples.

The loadings plot highlight the spectral features responsible for the delineation of the meat species. IR spectral features at 1176, 1396, 1467, 1736, 2850 and 2917 cm^{-1} were associated with the greatest variance. These features where dominated by lipid signals evident in the intense peaks at 1176, 1396, 1736, 2850 and 2917 cm^{-1} across both PCs, with the overlapping samples making it difficult to highlight contribution of these peaks across the three sample groups. As such, further classification analysis was performed using PLSDA and SVMC.

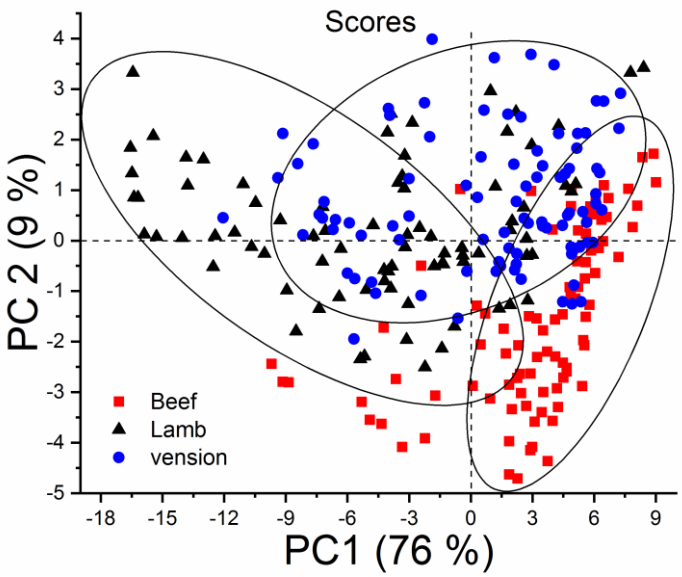

Figure 2.7: PCA scores plot performed using FTIR spectral data showing separation of beef, venison and lamb meat samples.

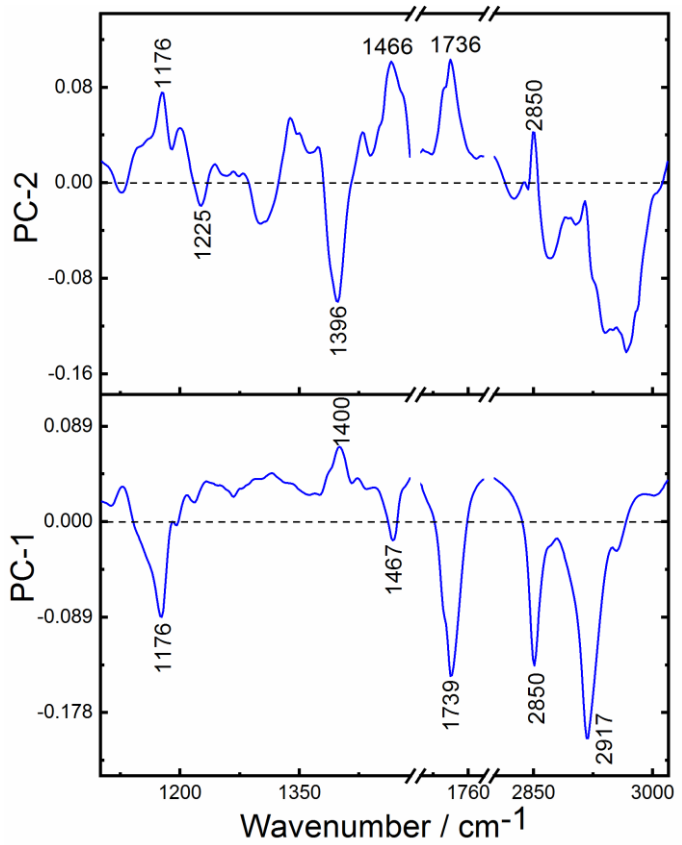

Figure 2.8: PCA loadings plot with IR spectral features associated with variation in beef, venison and lamb meat.

2.5.2 Classification of meat samples – Raman data

2.5.2.1 Partial Least Square - Discriminant Analysis

A one versus all PLSDA approach was adopted with one model created for classifying samples belonging to each meat class (beef or venison or lamb). Optimum PLSDA model was determined using the minimum root mean square error of cross-validation (RMSECV) of the training set. This procedure resulted in the selection of four LV for the venison model and three LV for both beef and lamb models. A summary of each classification model performance and parameters used is shown in Table 2.2.

The PLSDA model discriminating venison from beef and lamb was constructed using first four latent variables, describing 85 % of the variance within the dataset. A plot of the class predictions for the venison sample is shown in Figure 2.9. The model yielded a prediction accuracy of 97 and 100 % was obtained for each of the training and test sets. Validation of the model on the test set yielded maximum sensitivity and specificity (100 %). The PLSDA model was discriminating lamb meat species were built using the first three factors, which account for 77 % of the total variance within the dataset. The model yielded an accuracy of 92 and 81 % for each of the training and test sets, with sensitivity and specificity of 87 and 60 % achieved upon validation against the test set. Similarly, in the beef PLSDA model, three latent variables were adopted which accounted for 62% of the variability in the dataset. The model yielded an accuracy of 88 and 79 % for the training and test sets, respectively, with sensitivity and specificity of 87 % and 75 % achieved on validation against the test set, respectively. The venison model had the best performance in predicting unknown samples with lamb model having the lowest performance.

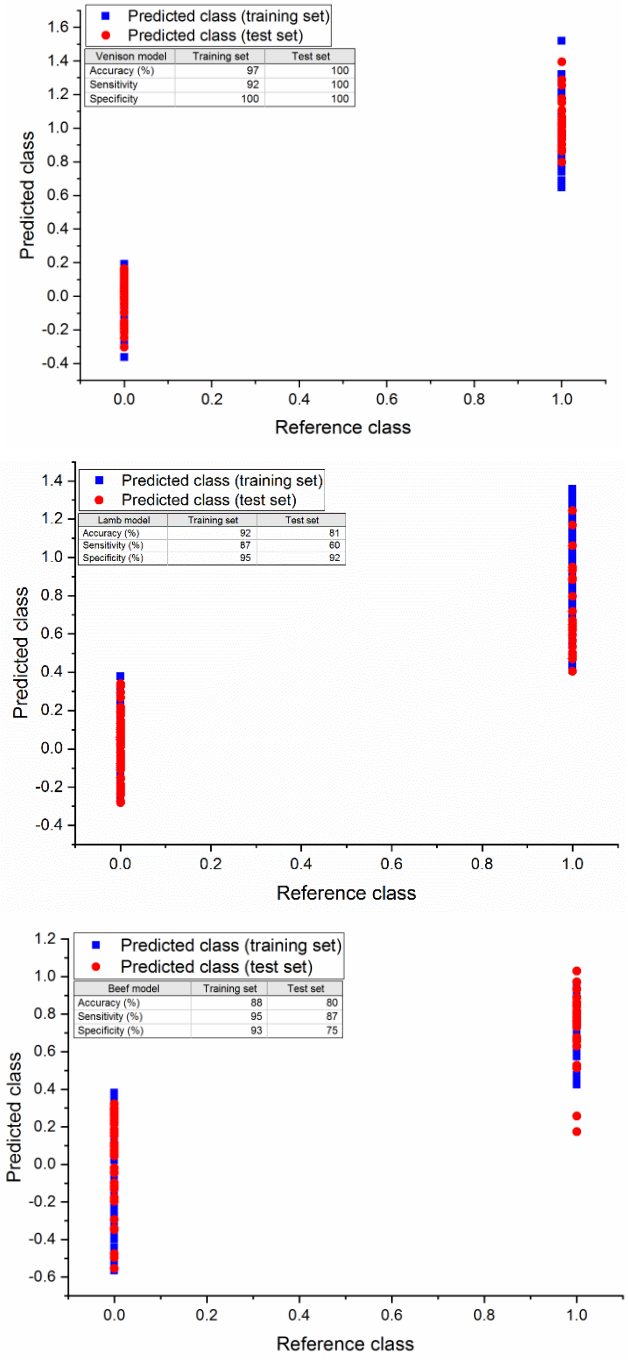

Figure 2.9: Prediction of venison (top), lamb (middle) and beef (bottom) meat classes using PLSDA and Raman data, where 1 represents class members and 0 represents non-members. Samples with value close to 1 in the predicted class suggest that they belong to the meat class, whereas samples with value close to 0 suggest they are non-members.

The regression coefficient, which highlights Raman features responsible for the discrimination of the meat types, is shown in Figure 2.10. In the venison PLSDA model, Raman peaks at 933, 1362 and 1451 cm^{-1} were responsible for the discrimination. The peaks at 933 cm^{-1} and 1362 cm^{-1} are consistent with amino acid residue such as lysine and tryptophan (Table 2.2) whereas the peak at 1451 cm^{-1} coincided with CH_2 and CH_3 bending vibration (aliphatic hydrophobic residue) [74, 133, 137]. In the beef model, the regression coefficient showed major contribution from vibrations at 911, 977, 1267, 1362 and 1660 cm^{-1}. The band at 1362 cm^{-1} has been associated with C-C stretching vibration in tryptophan [74, 133, 134]. The features at 911 and 1267 cm^{-1} are associated with lipid [74, 133, 143] and the feature at 1660 cm^{-1} is consistent with an amide I band [86, 130]. Vibrations associated with discrimination in the PLSDA lamb model were observed at 911, 977, 1362, 1451 and 1660 cm^{-1}. These are associated with protein and lipid as described above.

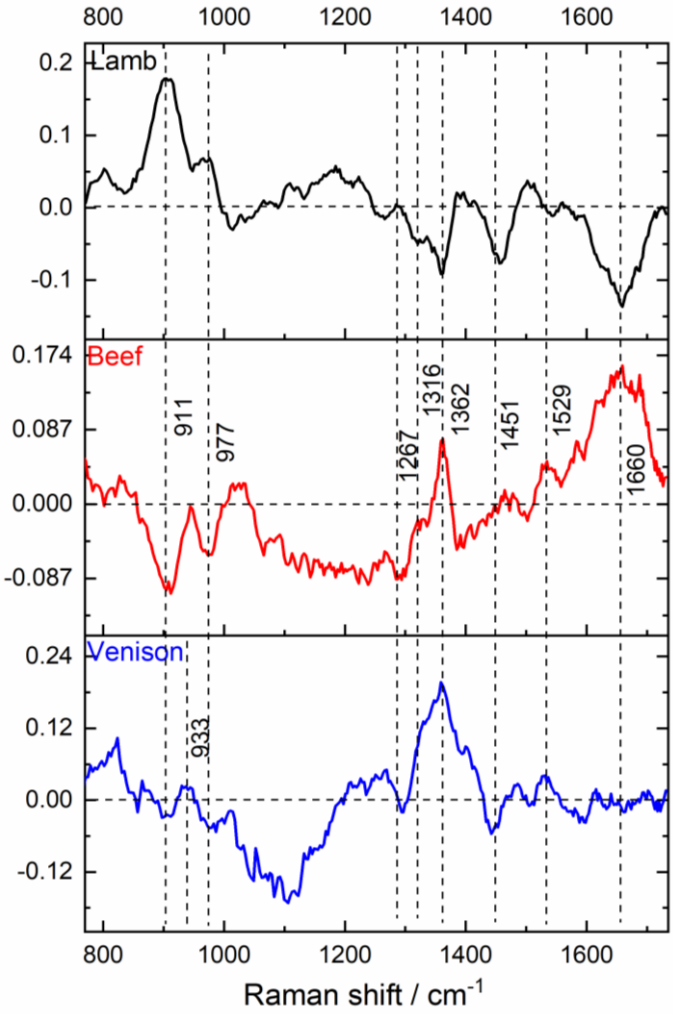

Figure 2.10: PLSDA regression coefficient showing Raman spectral features associated with classification of beef, venison and lamb samples

2.5.2.2 Support Vector Machine Classification (SVMC)

The SVMC classification was performed using a c-SVMC type; with a linear and non-linear (radial basis function, $\gamma = 0.01$) kernel function. The linear c-SVMC model yielded a training accuracy of 100 %, with validation on an independent test set yielding an accuracy of 92 %. Sensitivity and specificity of over 87 % were achieved for each meat specie using the linear kernel function. The non-linear c-SVMC gave an improved sensitivity and specificity on validation against the test set. The model yielded sensitivity

and specificity of over 90 and 95 %, respectively. The model had the best performance for classifying venison samples with a sensitivity of 100 % for both linear and non-linear SVMC models. A summary of model performances is shown in Table 2.4.

Table 2.4: Model performance for the discrimination of beef, venison and lamb meat Raman spectroscopy

PLS-DA models	Accuracy (%)	Sensitivity (%)	Specificity (%)
Beef vs others	88 (80)	95 (87)	93 (75)
Venison vs others	97 (100)	92 (100)	100 (100)
Lamb vs others	92 (81)	87 (60)	95 (92)

SVMC (training and test set)

Beef vs venison vs lamb	Accuracy (%)	Sensitivity (%)			Specificity (%)		
		Beef	Venison	Lamb	Beef	Venison	Lamb
SVMC (linear kernel)	100 (92)	100 (90)	100 (100)	100 (87)	100 (93)	100 (88)	100 (95)
SVMC (RBF)	100 (93)	100 (90)	100 (100)	100 (90)	100 (95)	100 (90)	100 (95)

2.5.3 Classification of meat samples – IR data

2.5.3.1 Partial Least Square - Discriminant Analysis

The PLSDA model development was performed using the same parameters as described for the Raman spectral analysis. Optimum PLSDA model was selected with respect to the minimum RMSECV for each of the PLSDA models. These models were validated against an independent test set.

The PLSDA model discriminating venison meat samples was constructed using the first seven (7) LVs accounting for 86 % of the variance within the datasetThe model

yielded a prediction accuracy of 99 and 98 % for the training and test sets, respectively. A sensitivity and specificity of 100 and 97 %, respectively was achieved on validation against an external test set. The model discriminating lamb meat species accounted for 63 % of the variances within the dataset. A model accuracy, sensitivity and specificity of 94, 93 and 95 %, respectively upon validation against an external test set. Similarly, the beef model yielded a maximum sensitivity and specificity of 100% upon validation against an external using 7 LVs, describing 81% of the total variance within the dataset. A larger number of LV was selected for the IR analysis as compared to Raman owing to the minimum RMSECV obtained across each data set; this was utilised in determining the optimum model.

The regression coefficient plot highlighting spectral features responsible for the discrimination of the meat species is shown in Figure 2.12. In the venison model, IR peaks at 1100, 1139, 1453, 2850 and 2920 cm^{-1} had the most contribution to the loadings. The peaks are consistent with lipids (1453, 2850 and 2920 cm^{-1}) and glycogen (1100 and 1137 cm^{-1}) [114, 146]. In the beef model, the major contribution were from bands at 1161, 1213, 1400 and 1469 cm^{-1} which have been associated with lipids (1161, 1400, 1469 and 2920 cm^{-1}) and asymmetric PO_2 stretching bands (1213 cm^{-1}) [154]. Major contribution for discrimination of lamb samples were observed at 1157, 1213 and 2920 cm^{-1}, corresponding to nuclei acid and lipid vibrations. The beef model had the best performance in predicting unknown samples with the lamb sample having the least performance. This is not surprising as the PCA highlighted higher variability within samples of the lamb and venison group.

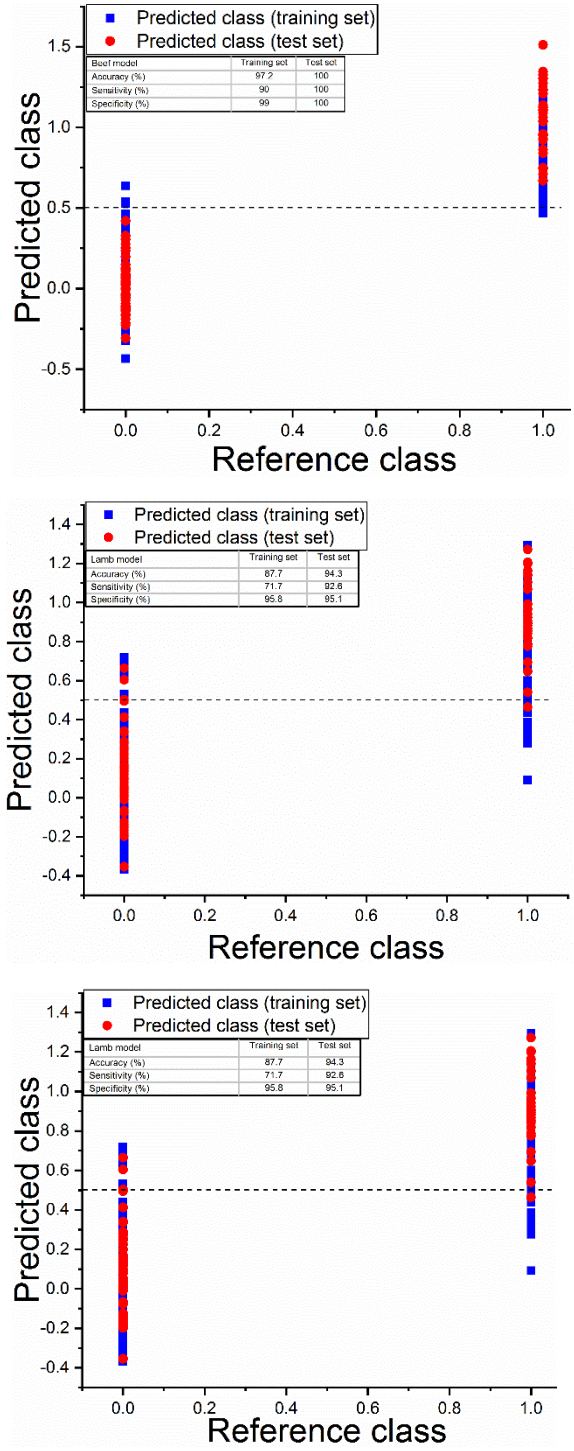

Figure 2.11: Prediction of venison (top), lamb (middle) and beef (bottom) meat classes using PLSDA and IR data, where 1 represents class members and 0 represents non-members. Samples with value close to 1 in the predicted class suggest that they belong to the meat class, whereas samples with value close to 0 suggest they are non-members.

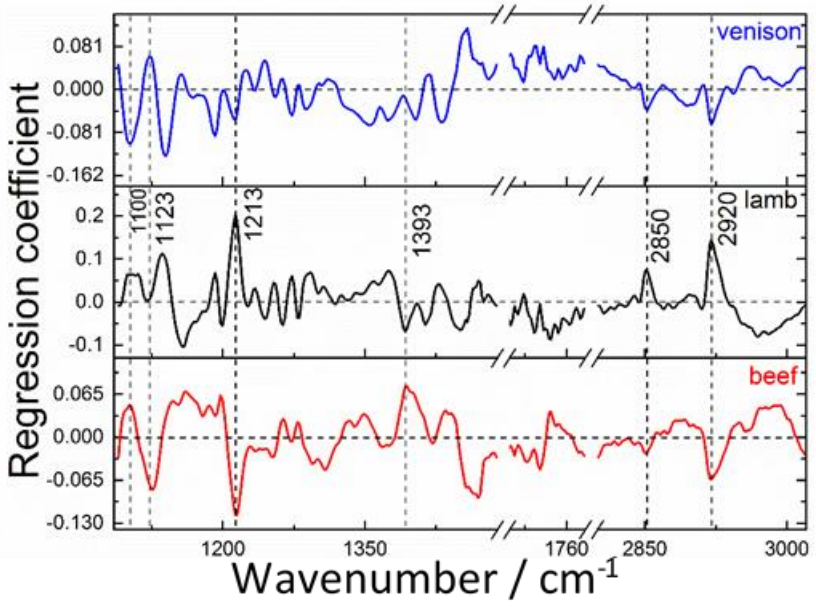

Figure 2.12: PLSDA regression coefficient showing IR spectral features associated with discrimination of beef, venison and lamb samples.

2.5.3.2 Support vector machine classification

The SVMC classification was performed in similar manner as the Raman analysis. The linear c-SVMC model yielded a training accuracy of 100 %, with validation on an independent test set yielding an accuracy of 98.9 %. Sensitivity and specificity of over 96 % were achieved for each meat specie using the linear kernel function. The non-linear c-SVMC yielded similar result without any significant model improvement. Both the linear and non-linear kernel had the best performance for classifying the lamb and venison sample. A summary of model performances is shown in Table 2.5.

Table 2.5: Model performance for the discrimination of beef, venison and lamb meat using infrared spectroscopy

PLSDA (training and test set)								
PLS-DA models	Accuracy (%)	Sensitivity (%)			Specificity (%)			
Beef vs others	97.2 (100)	90 (100)			99.2 (100)			
Venison vs others	98.9 (97.7)	100 (96.8)			98.3 (98.2)			
Lamb vs others	87.7 (94.3)	71.7 (92.6)			95.8 (95.1)			
SVMC (training and test set)								
Beef vs venison vs lamb	Accuracy (%)	Sensitivity (%)			Specificity (%)			
		Beef	Venison	Lamb	Beef	Venison	Lamb	
SVMC (linear Kernel)	100 (98.9)	100 (96.7)	100 (100)	100 (100)	100 (100)	100 (98.3)	100 (98.3)	
SVMC (RBF)	100 (98.9)	100 (96.7)	100 (100)	100 (100	100 (100)	100 (98.3)	100 (98.3)	

The focal point of this study was to establish the potency of vibrational spectroscopic technique in combination with combination with more than one chemometric technique for intact red meat discrimination. Different preprocessing techniques were trialled based on the quality of spectral data acquired, however, only the technique that yielded optimum solution was reported. Smoothing of the Raman spectra was necessary owing to the presence thermal and electrical signals reducing the quality of the spectral data. PCA against the IR data had considerably overlap across the meat types, owing to the subsampling error associated with the heterogeneity of meat. This error was less of a challenge in the Raman system by integrating the raster sampling system, as such clearer separation was observed in the PCA. Optimum classification model was selected based on the minimum RMSE on cross-validation. Classification models computed using data for both techniques were successful in predicted external test samples not involved in the calibration, with accuracies of over 80 % (PLSDA) and 92 % (SVMC) on the Raman data and accuracies of over 98 % (SVMC) and 94 % (PLSDA) using IR spectra.

Potential inline/at-line inspections using these techniques for food will require longer measurement time and large area sampling to enable acquisition of high quality and representative spectra data. However, this will still be a very short time compared to analysis using conventional techniques. Inclusion of processed and fresh meat samples in the calibration model might facilitate the stability, robustness and applicability of the models to real life situation. The availability of handheld Raman techniques also improve quick analysis in supermarkets and abattoirs but might reduce but might yield poor spectral data in comparison to benchtop or optical fibre coupled devices.

2.6 Conclusion

Raman and infrared spectral data of 90 red meat samples were obtained and analysed using a range of spectral preprocessing and chemometric techniques, to highlight the potential of vibrational spectroscopic techniques for discrimination of red meat samples. Optimum models were selected based on the RMSE and Raman spectroscopy in combination with PLSDA and SVMC were found to be most effective for discrimination of all meat samples. However, the SVMC tended to yield higher sensitivities for both Raman and infrared spectral data. The potential of Raman and infrared spectroscopy as an alternative technique for intact meat discrimination has been shown, with successful discrimination of the three meat types achieved despite their similar chemical composition.

Chapter 3

Quality assessment of red meat

3.1 Acknowledgments

The samples discussed in this chapter were sourced and processed by AgResearch, New Zealand. The reference chemical measurements were also performed by AgResearch.

3.2 Results publication

The results discussed in this chapter have been published:

Chima Robert, William Jessep, Joshua J. Sutton, Talia M. Hicks, Mark Loeffen, Mustafa Farouk, James F. Ward, Wendy E. Bain, Cameron R. Craigie, Sara J. Fraser-Miller and Keith C. Gordon: Evaluating low-, mid- and high-level fusion strategies for combining Raman and infrared spectroscopy for quality assessment of red meat. *Food Chemistry* 2021, *361*, 130154.

My contribution to this paper included conceptualization, data collection, data curation, formal analysis, drafting of the original manuscript and correction of the reviewed manuscript.

3.3 Introduction

The quality of meat plays a significant role in the overall purchasing attitude and eating experience of consumers [131], hence, the continual pursuit for techniques capable

of objective and accurate quality assessment by the meat industry. Meat quality envelops both the sensory and nutritional values (composition) of the meat. These attributes are influenced by several factors including animal genetics, production systems, growth rate, slaughter conditions, as well as the biochemical processes taking place during the conversion of animal muscle to meat [155]. Several indicators have been established for the assessment of meat quality including pH, percentage intramuscular fat content (% IMF), fatty acid profile, protein content, shear force, water holding capacity, drip loss, colour and marbling [64, 156, 157].

The pH value in red meat is critical as it influences the overall appearance of meat cut. The ultimate pH (pH_u) value of meat is widely measured at 24 hours post-slaughter (pH_{24}), and is a function of the post-mortem biochemical changes which occur during the conversion of stored muscle glycogen to replenish cellular adenosine triphosphate (ATP) via glycogenolysis and glycolysis, resulting in an accumulation of lactate. Furthermore, the replenished ATP is subsequently hydrolysed to generate energy (conserving cellular processes), liberating H^+ ions responsible for reducing the pH of meat after slaughter [158]. Broadly speaking, the desired pH values of red meat range from 5.4 - 5.8, with the resulting red meat often appearing reddish-firm-non exudative (RFN) [157, 159]. Nevertheless, possible deviation in glycogenolysis and glycolysis could lead to variation in the ultimate pH value. This deviation is commonly a function of the impact of pre-slaughter stress on the animal. Accelerated and delayed glycolytic processes could lead to the formation of pale-soft-exudative (PSE) meat ($pH < 5.4$) and dark-firm-dry (DFD) meat ($pH > 5.8$), respectively, resulting in meat with poor physical appearance and eating quality [160]. Hence, there is a need to assess the pH value and lactate concentration of red meat.

The % IMF impacts on the nutritional value, sensory characteristics (flavour) and shelf life of meat [156]. Among other factors, the animal genetics and rearing conditions play significant roles in the fat composition of red meat [161, 162]. The dietary benefits of fat and fatty acid composition in meat including palmitic acid, stearic acid, linoleic acid, linolenic acid, eicosapentaenoic acid and docosapentaenoic acid have been investigated [163, 164].

Conventional methods of meat quality assessment include chromatographic methods for the determination of fatty acid composition, solvent extraction or iodometric titration for determining fat content [165] and electrochemical methods for determining pH [166]. However, these methods are destructive and can require time-consuming sample preparation. Spectroscopic methods such as near-infrared, mid-infrared, hyperspectral imaging, visible, fluorescence and Raman spectroscopy have been explored for meat quality assessment [167]. Research involving the use of spectroscopic techniques as alternatives for conventional methods have soared in the last two decades. This has been mainly due to advancements in optical technologies and the potential advantages of minimal sample preparation, short analysis time, ease of use, cost effectiveness, non-destructiveness, the possibility of automation, miniaturization, on-line/in-line integration, as well as the ability to build quantitative and qualitative models [64].

Vibrational spectroscopic techniques like Raman and infrared spectroscopy provide complementary information on the structure and molecular interactions within a system. A combination of data from these techniques could provide deeper understanding into the nature, structure and interactions within such system; this is commonly referred to as data (or sensor) fusion. Data fusion techniques involve the combination of data from multiple instruments such as ultraviolet/visible (UV/Vis), near infrared, mid infrared, Raman, nuclear magnetic resonance (NMR), mass spectrometry and other physio-chemical measurements to achieve optimal information for the sample of interest [168, 169]. Fusion aims to exploit the potential synergistic effect on the combination of more than one technique; culminating in the enhancement of information quality [170, 171]. Data fusion is a recent trend in analytical chemistry and has been applied in the discrimination of alcoholic beverages [172], meat authentication [168], honey discrimination and prediction of meat spoilage [173].

Data fusion can be performed at three levels; low-level (LLF), mid-level (MLF) and high-level (HLF) [172]. LLF, also referred to as measurement level fusion, involves the combination of pre-processed spectral data (measurements) from various instruments into a single matrix. This matrix is further used in building qualitative or quantitative models [174]. Unlike LLF, MLF (feature level fusion) strategy integrates a feature extraction

step, which can hold adequate original information, with the extracted features combined to build quantitative or qualitative models. The MLF are easily adaptable with established feature extraction steps such as principal component analysis (PCA) or multivariate curve resolution (MCR) [169, 174]. In HLF (decision level fusion), classification or regression models are built using data from individual techniques and the model predictions further fused to obtain a final outcome. HLF is the least explored of all three data fusion strategies and susceptible to significant information loss, owing to the use of model outputs for final predictions [168, 171, 172]. Data from both instruments are not utilised in developing a single model rather their outputs are combined based on some assumptions. For example, majority voting is adopted for HLF in classification studies. This entails the assignment of a sample to a class based on the majority of the individual instrument assignments. This might not be accurate as individual techniques can also yield wrong classification.

For the purpose of this work, the model performance was assessed with respect to the determination coefficient (R^2) and root mean square error for the training ($RMSEC$), cross-validation ($RMSECV$) and test sets ($RMSEP$). A normalized root mean square error of prediction ($NRMSEP$) was calculated using $RMSEP$ and the observed range of the reference measurement. The $NRMSEP$ can be obtained by the expression in equation [175].

$$NRMSEP = \frac{RMSEP}{y_{max} - y_{min}} \qquad 3.1$$

where, *RMSEP* is the root mean square error of prediction and $y_{max} - y_{min}$ is the range of the values in reference measurements (*y matrix*). In general terms, a low $RMSEP$ as well as a high R^2 value suggest a good predictive ability for the model [48].

In the present study, I set out to show the performance of Raman and infrared spectroscopy to predict pH and % IMF meat quality parameters, and also to evaluate the performance of three data fusion strategies in predicting these meat quality indicators.

3.4 Materials and methods

Detailed description of the sample collection, reference chemical measurements, spectroscopic analysis and spectral preprocessing have been presented in the materials and methods section in chapter 2.

3.4.1 Data fusion strategies

Data fusion can be performed at three different levels: low, mid and high-level fusion. In the present study, the potential of all three fusion strategies to predict pH and % IMF quality parameters was investigated. An illustration of data fusion strategies is presented in Figure 3.1.

3.4.1.1 Low-level fusion (LLF)

LLF is achieved by concatenating preprocessed data from individual sources to form a single matrix, *X*. The matrix comprises of *m*-rows (individual samples) and *n*-columns (measurement variables from each source). In this study, fusing the data at the low-level yielded 713 *n*-variables, with the Raman and infrared instruments contributing 314 and 399 variables, respectively.

3.4.1.2 Mid-level fusion (MLF)

MLF strategy entails concatenating extracted features from different analytical sources. The spectral data from each technique are independently preprocessed and data reduced using techniques such as PCA and MCR. The obtained features like scores (PCA) or component concentrations (MCR) from both instruments are concatenated sample-wise to form a single matrix. The feature extraction step allows for data reduction without substantial loss of information as well as the adequate representation of information from each data block unlike the low-level fusion [174]. For example, in the LLF, Raman and infrared instruments contributed 314 and 399 variables, respectively, giving the infrared

data block a greater variable contribution in the model. However, using PCA and MCR data reduction techniques, representative features were extracted from each block and utilised in forming a single matrix. The Raman and infrared instruments contributed 5 and 6 principal components (PC), respectively (for PCA data reduction), whereas each of the instrument contributed 7 components using MCR data reduction technique. This yielded a total of 11 PCs (PCA) and 14 components (MCR) as the extracted features. Lesser number of PCs was selected for Raman to avoid inclusion of noise component.

3.4.1.3 High level fusion (HLF)

Data fusion at a high level or decision level entails fusing outcomes of classification or regression models from individual techniques to identify the best outcome. In HLF, each individual technique is treated independently; as such, poor performance from one technique does not worsen the overall performance unlike other fusion strategies. For the purpose of this study, independent model predictions were first calculated using the Raman and MIR spectral data, and the outputs for each sample averaged (n = 3) prior to calculating the HLF. The final outcome was then calculated using central limit theorem (CLT), taking the standard deviation as the weightings as described in the equations

$$Prediction\ for\ CLT,\quad X_{RaIR} = \sigma^2_{RaIR}[\,\bar{X}_{Ra}\cdot\sigma^{-2}_{Ra} + \bar{X}_{IR}\cdot\sigma^{-2}_{IR}\,] \qquad 3.2$$

$$\sigma^2_{RaIR} = [\sigma^{-2}_{Ra} + \sigma^{-2}_{IR}]^{-1} \qquad 3.3$$

where X_{RaIR} is the prediction from the fused outcomes, $\bar{X}_{Ra}$ is the average estimated pH or % IMF for an observation (sample) using Raman, $\bar{X}_{IR}$ is the average estimated pH or % IMF for the same observation (sample) using MIR, σ^2_{Ra} *and* σ^2_{IR} are the variances from the Raman and FTIR instruments, respectively and σ^2_{RaIR} is the combined variances for the observation from both sensor [176].

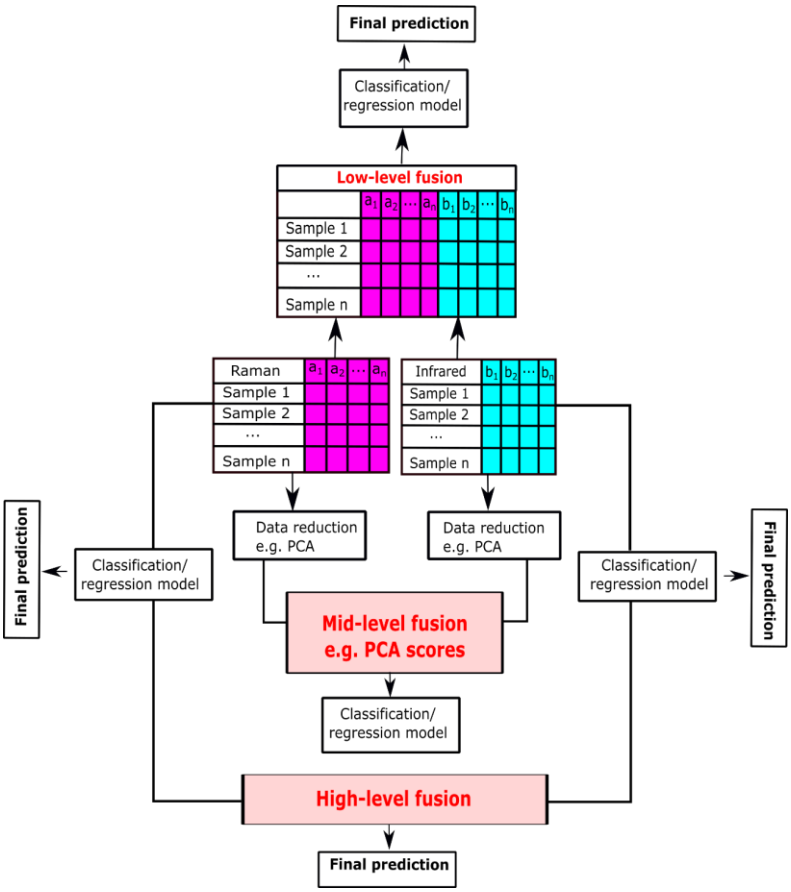

Figure 3.1: A simplified representation of low-level (top), mid-level (middle) and high-level (bottom) data fusion strategies.

3.4.2 Multivariate analysis

Reference pH and % IMF (*y-matrix)* were used in building quantitative models against acquired spectroscopic data (*X-matrix)*. Preprocessed spectral data were

correlated against the reference using partial least squares (PLS) in The Unscrambler X 10.3 (CAMO, Oslo, Norway).

The building of predictive models using PLS has been widely applied in analytical chemistry. For the purpose of this study, the PLS models were calculated using a segmented k-fold cross-validation approach, where k replicates (k = 3) from the training set formed a segment and on cross-validation each excluded segment is predicted using the training model. The cross-validation technique despite giving an estimate of the model's predictive ability is limited, in that prediction is performed on samples belonging to the training set. The use of an independent test set provides better insight into the model's robustness in predicting unknown samples.

3.5 Results and Discussion

The results and discussion are divided into two main components: initial exploratory analysis, regression analysis using individual instruments and data fusion strategies.

3.5.1 Exploratory analysis

Table 3.1 presents the values of the mean and standard deviation (SD) including minimum and maximum values for pH and % IMF content of the red meat samples. The mean pH varied from a minimum of 5.27 ± 0.07 for venison to a maximum 5.91 ± 0.46 for beef. The pH range across the beef samples is 5.1 - 6.8, with the 1.7 pH unit difference being the largest between samples of same meat specie. The % IMF content showed greater variability across the samples evident in the wider range and coefficient of variation (CV) values. The variation in the % IMF was expected as several factors including animal genetics, breed, sex, age at slaughter and rearing conditions play a role in the overall fat content in meat [161, 162]. However, the individual meat species did not provide sufficient variability in their quality traits (pH and % IMF inclusive) to build a robust predictive model, hence the use of three red meat species to cover a wider reference range.

Table 3.1: Mean and standard deviation values of pH and percentage intramuscular fat content (% IMF) in venison, beef and lamb

Parameter	Meat specie	Minimum	Maximum	Mean	SD[a]	CV (%)[b]
pH	Venison	5.10	5.44	5.27	0.07	1.33
	Beef	5.10	6.8	5.91	0.46	7.78
	Lamb	5.49	5.92	5.61	0.11	1.96
% IMF	Venison	1.12	2.05	1.48	0.24	16.2
	Beef	0.64	3.23	1.30	0.67	51.5
	Lamb	1.00	4.18	2.17	0.95	43.8

[a]SD = Standard deviation

[b]CV = Coefficient of variance $\left(\frac{100SD}{mean}\right)$

3.5.2 Spectral data analysis

The Raman spectral data were acquired over a spectral range of 313-1895 cm^{-1}, and further analysis performed using the selected region (750-1750 cm^{-1}). The observed spectral data showed little difference in their Raman signatures across the three red meat species, which suggests similarity in their chemical composition, although minor differences are detectable [177]. The mean spectrum of the Raman data with the corresponding standard deviation is shown in Figure 3.1a. The dominant bands appear at 933, 1075, 1316, 1451 and 1660 cm^{-1}. The observed bands have been assigned based on literature as summarised in Table 3.1. The most intense band at 1451 cm^{-1} can be attributed to CH_2 and CH_3 bending vibration in lipids and protein, whereas the band at 1660 cm^{-1} is consistent with the amide I band. The broad features in the region 850-933 cm^{-1} can be attributed to polypeptide backbone, amino acid residues and fats [128, 131, 133, 135].

The MIR spectral data were collected over a spectral range of 400 – 4000 cm^{-1} and the selected regions (3025-2760 cm^{-1}, 1780-1710 cm^{-1} and 1485-1055 cm^{-1}) were

used for further analysis. The strongest bands were observed at 1546, 1634 and 3275 cm^{-1} whereas weaker bands appeared at 2924 cm^{-1}, 2850 cm^{-1}, 1736 cm^{-1}, 1453 cm^{-1}, 1396 cm^{-1}, 1242 cm^{-1}, 1170 and 1161 cm^{-1}. The band assignments suggest fat, protein and water to be the major components of the red meat samples detected. A summary of the band assignments is presented in Table 3.2. The 1485-1710 cm^{-1} region was excluded due to interfering water signals. Also, the broad absorption features around 3100 – 3600 cm^{-1} corresponding to the O-H stretching vibrations of water molecules and NH stretching of amide in proteins, were not considered in predicting the pH and % IMF due to the interfering water signal as well as the larger variability in this region compared to others [145]. The mean spectrum of the IR data with the corresponding standard deviation is shown in Figure 3.1b.

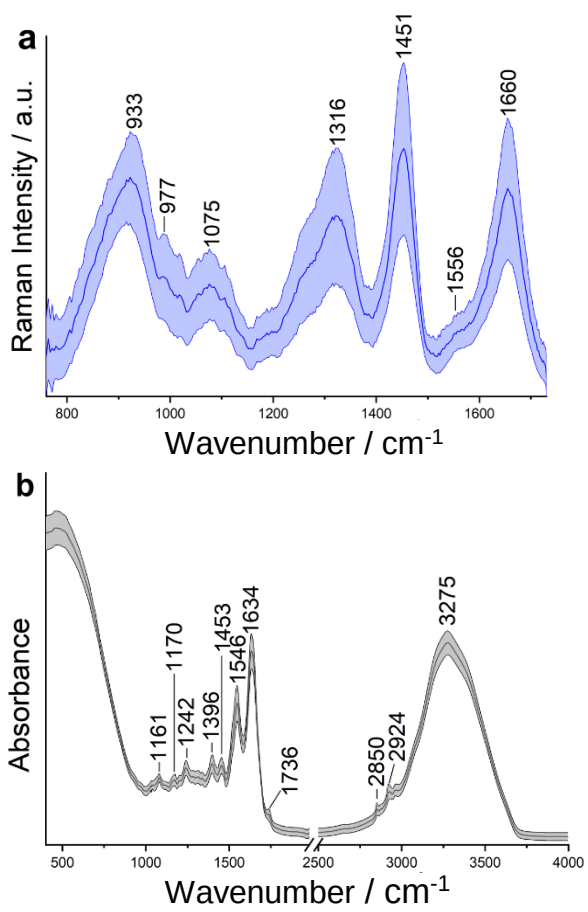

Figure 3.2: Mean spectra and standard deviation of preprocessed spectral data collected using (a) 1064 nm fibre optic – coupled Raman instrument (b) ATR – FTIR instrument

Table 3.2: Spectral bands assignment for the chemical functional groups in venison, lamb and beef according to literature

Raman (cm^{-1})	IR (cm^{-1})	Origin	Assignments	References
850 - 933		Amino acid (glutamic acid, lysine) and polypeptide	C-CN stretching	[131, 136]
977		Protein	=C-H wag	[134]
1316		Protein	Amide III fibrous α-helix	[86, 129, 144]
1451		Lipid	CH_2 and CH_3 bending	[74, 132, 137]
1556		Tryptophan	Bending vibrations of pyrrole and benzene rings	[74, 129, 134]
1660		Peptide backbone	Amide I	[86, 131, 133]
	1161	Lipids	C-O stretching and CH_2 bending	[145, 146]
	1170	Lipid	C-O-C asymmetric stretching (of esters)	[145]
	1242	Protein, Phospholipids	CN stretching and NH bending (amide III), PO_2^- asymmetric stretching	[147, 148]
	1396	Lipid	COO^- stretching and amino side chains	[146, 149]
	1453	Lipid	CH_2 and CH_3 scissoring	[36, 150]
	1546	Protein	NH bending and CN stretching (amide II)	[151]
	1634	Protein	CN stretching, C=C stretching (olefins), C=O stretching (amide I)	[146, 148, 152, 153]
	1736	Lipid	C=O stretching of ester, aldehyde, ketone: free fatty acids	[146, 148]
	2850	Lipid	CH_2 symmetric stretching	[145, 153]
	2924	Lipid	CH_2 asymmetric stretching	[145, 153]
	3275	Protein	NH asymmetric and symmetric stretching	[145, 151]

3.5.3 Correlation of spectral data to meat quality parameters

The output of the PLS models are presented as a plot of the predicted values versus the observed reference values. A summary of all PLS model performances are shown in Table 23. The number of factors introduced in each model was determined by the RMSECV. In all cases, no more than seven factors were necessary for a rational PLS model. The optimal PLS model was adopted after removal of spectral data points with potential bias (outliers), with model improvement observed in most cases.

3.5.4 Individual spectroscopic instrument

3.5.4.1 Estimation of pH

The pH of samples utilized in this study ranged from 5.1 – 6.8 pH units (SD = 0.05), covering a range common to various meat classifications including pale-soft-exudative (PSE), dark-firm-dry (DFD) and red-firm-non exudative (RFN) meat [178].

The Raman spectral region (750-1750 cm^{-1}) was deemed the most informative and was correlated to the observed pH and % IMF measurements. Figure 3.3 shows the correlation between the predicted pH values against the reference measurements using the Raman spectral data. The pH PLS model was created using seven factors accounting for 83 % of the variance in the data set. The model yielded a determination coefficient (R_C^2) of 0.84 and $RMSEC$ of 0.14. Results obtained on cross-validation of the model yielded R_{CV}^2 and $RMSEC$ values of 0.64 and 0.22, respectively. This shows a positive correlation between the predicted and observed pH values. Upon validation of the model against an independent test set, a $RMSEP$ of 0.36 corresponding to a $NRMSEP$ of 21.1 %, highlighting the model's ability to predict unknown samples. Based on data from previous studies on pH prediction using Raman, we observed two peaks that have been

associated with phosphate (875 cm^{-1}) and inosine monophosphate (1573 cm^{-1}) in the regression coefficient plot [179-181]. Other observable peaks at 1115, 1360, 1477, 1573, and 1618 cm^{-1} have been associated with tryptophan [137, 182]. The regression coefficient plot is shown in Figure 3.4.

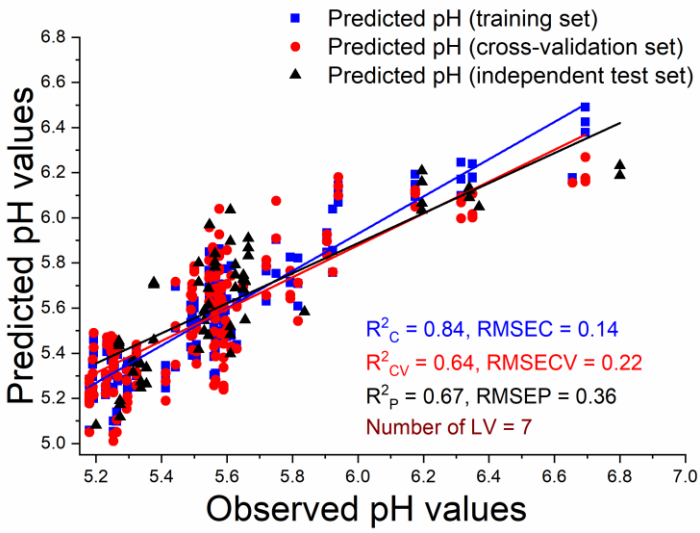

Figure 3.3: Regression line showing correlation of observed pH values with Raman spectral measurements.

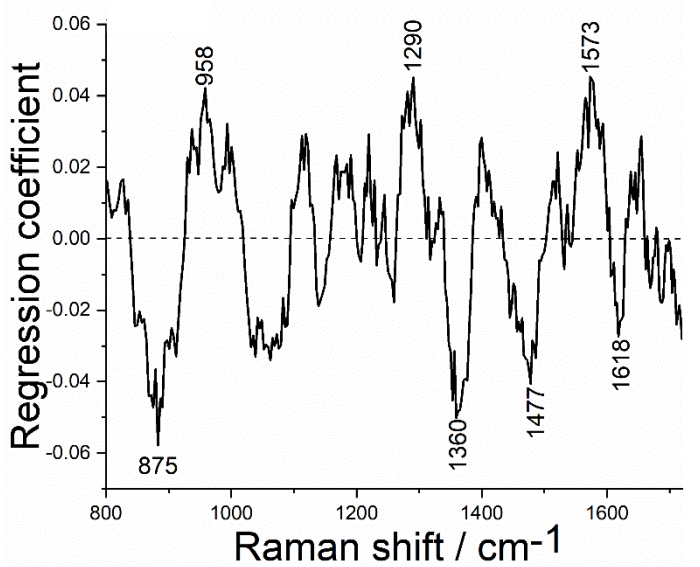

Figure 3.4: Regression coefficient showing spectral features influencing the pH prediction of red meat using Raman spectral data.

The performance of the model was consistent with a study by Scheier and Schmidt [181] for the prediction of pH in pork (R_C^2 = 0.87 and $RMSEC$ = 0.22), corresponding to a $NRMSEC$ of 18 % [175]. It is important to note that this study was performed on a smaller dataset and the model not validated against an independent test set. Similarly, a study involving the prediction of pH_{24} in pork on a larger dataset by [157] yielded an R_{CV}^2 = 0.68 and $RMSECV$ = 0.09, corresponding to a $NRMSECV$ of about 10 % using 8 factors [175].

The model obtained using the MIR spectral data showed good predictive values for pH, with a R_C^2 and $RMSEC$ of 0.78 and 0.20, respectively on the training set. Validation of the model against an independent test set yielded a R_P^2 and $RMSEP$ of 0.82 and 0.20, respectively. This error corresponds to a $NRMSEP$ of 11.8 %. These values show a positive relationship between predicted and observed pH measurements. The regression coefficient for the pH prediction shows major contribution from peaks at 1068, 1082, 1125, 1220, 1400, 1458 and 2918 cm^{-1}. The strongest peaks at 1082 and 1125 cm^{-1} have been assigned to PO_2^- symmetric stretching and C-O/C-C stretching vibrations, respectively [145] whereas the bands between 1440 to 1458 cm^{-1} are consistent with complex CH bending pattern associated with lactate [183] . In literature, both the phosphate and lactate bands have been correlated with actual pH content in meat [180, 183], which means that the variables used by these model are chemically meaningful. The regression plot and regression coefficient are shown in Figure 3.5 and 3.6, respectively.

Both the Raman and FTIR systems showed positive correlations in predicting the observed pH values, however, the MIR spectral data offered better prediction results upon validation against an independent test set which could partly be attributed to the reduced noise contribution ($RMSEP$ = 0.2, $NRMSEP$ = 11.8 %). A summary of the model performance parameters is shown in Table 3.3.

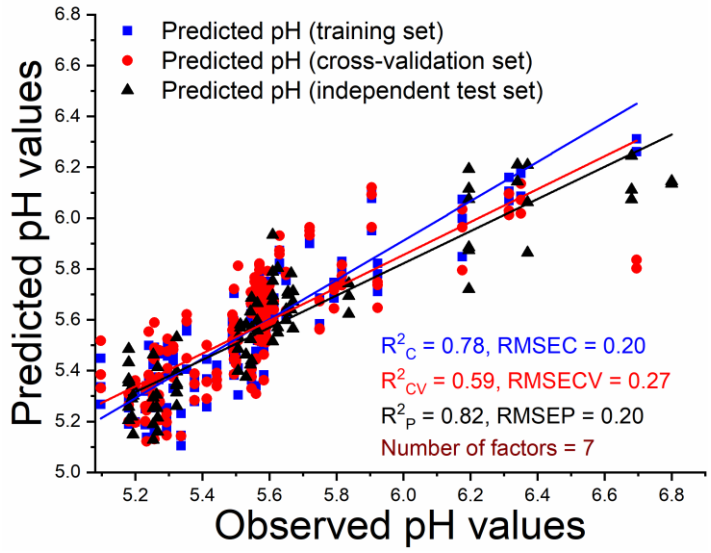

Figure 3.5: Regression line showing correlation of observed pH values with IR spectral measurements.

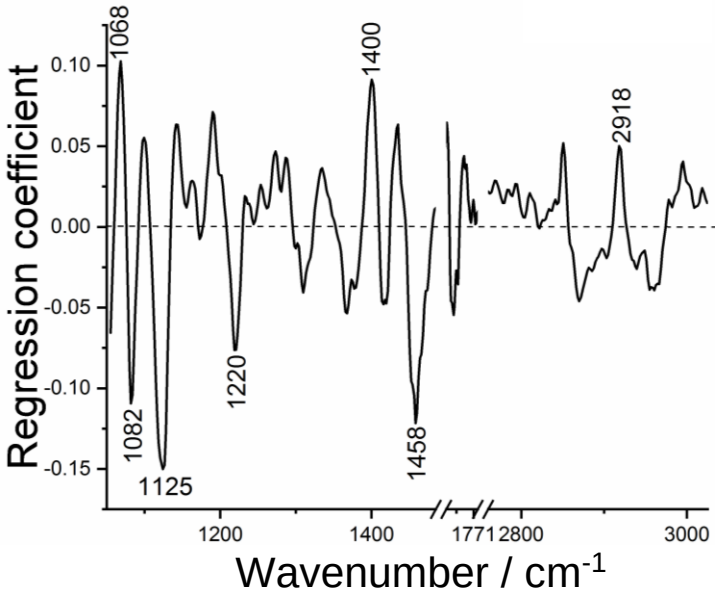

Figure 3.6: Regression coefficient showing spectral features influencing the pH prediction of red meat using IR spectral data

3.5.4.2 Percentage Intramuscular fat (% IMF)

The % IMF of samples used in this study ranged from 0.6 % - 4.2 %. As with the pH models, model calculation was performed using a k-fold cross-validation technique. Predicting the % IMF using the Raman data showed a strong positive correlation with the observed % IMF values (R_C^2 = 0.75), as shown in Figure 3.7. The R^2 value slightly reduced on cross-validation (R_{CV}^2 = 0.72) with a corresponding $RMSECV$ of 0.26 obtained. Validation on an independent test set yielded a reduced R_P^2 and $RMSEP$ of 0.60 and 0.30, respectively. The error corresponds to a $NRMSEP$ of 8.5 %. The performance of the model was consistent with a previous study for prediction of different fatty acid composition in intact adipose tissues (R^2 > 0.73, average $RMSEP$ = 11.9 %) by [128], but higher for % IMF prediction in lamb samples using 2 factors (R_{CV}^2 = 0.02, $RMSECV$ = 1.12) [156]. The improved performance could be due to the larger number of latent variables introduced in the current model. However, model overfitting was prevented by visualising the explained variance and loadings plot, ensuring that the additional factors contributed to the total variance. The regression coefficient features show that the main bands of influence appear around 886, 990, 1116, 1445 and 1640 cm^{-1}. Despite the subtle changes in the peaks, the band at 1445 cm^{-1} is consistent with lipid modelling in red meat [137].

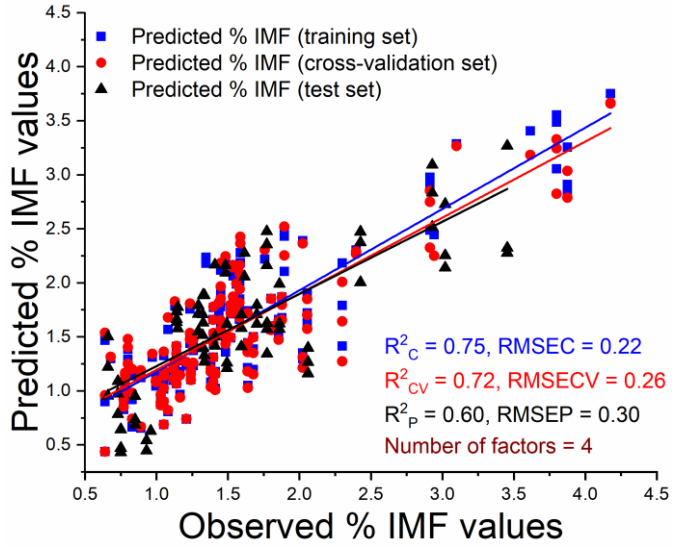

Figure 3.7: Regression line showing correlation of observed % IMF values with Raman spectral measurements.

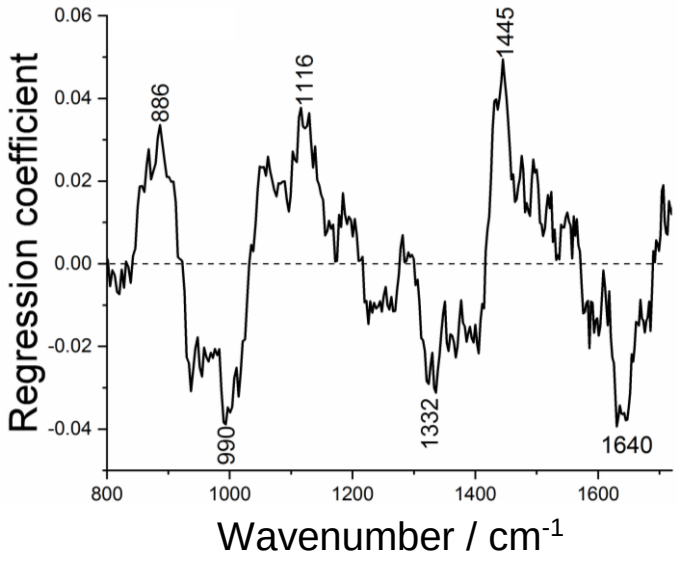

Figure 3.8: Regression coefficient showing spectral features influencing the % IMF content prediction in red meat using Raman spectral data.

A plot correlating the observed % IMF with IR spectral data is shown in Figure 3.9. The training model yielded a R_C^2 and $RMSEC$ value of 0.70 and 0.35, respectively. On

validation against an independent test set a R_P^2 and $RMSEP$ of 0.67 and 0.40, respectively. This corresponds to a $NRMSEP$ of 11.3%; this error is rather low considering it was obtained using an external test set. A similar study performed using MIR spectral regions at 3000-2800, 2300-1850 and 981-489 cm^{-1} yielded an R_{CV}^2 value of 0.92 and $RMSECV$ of 0.71 using 11 factors, which is rather high considering that a cross-validation was performed on same set of data belonging to the training set [114]. The regression coefficient plot shows that IR peaks at 1153, 1174, 1726, 1750, 2848, 2916 and 2931cm^{-1} are major contributors to the prediction plot (Figure 3.10). These bands despite slight shifts have been associated with lipid composition in the meat samples [114, 145], which means that the variables used by these models are chemically meaningful. The bands at 1153, 1750 and 2931 cm^{-1} were positively correlated to the % IMF content in the red meat.

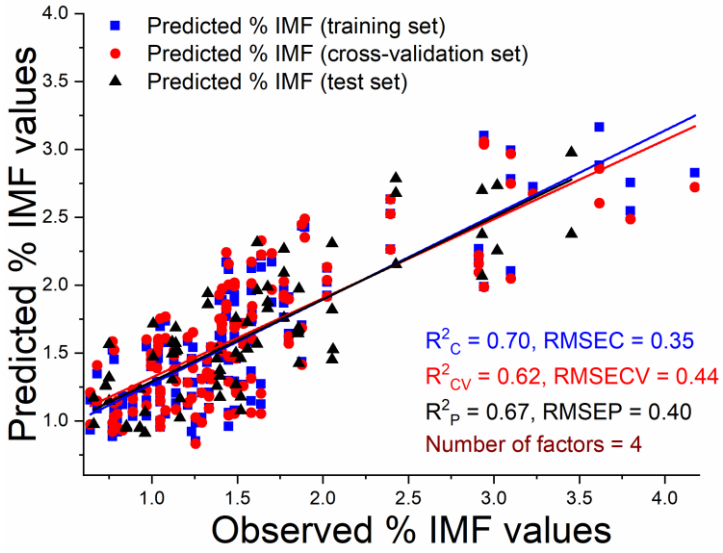

Figure 3.9: Regression line showing correlation of observed % IMF values with IR spectral measurements.

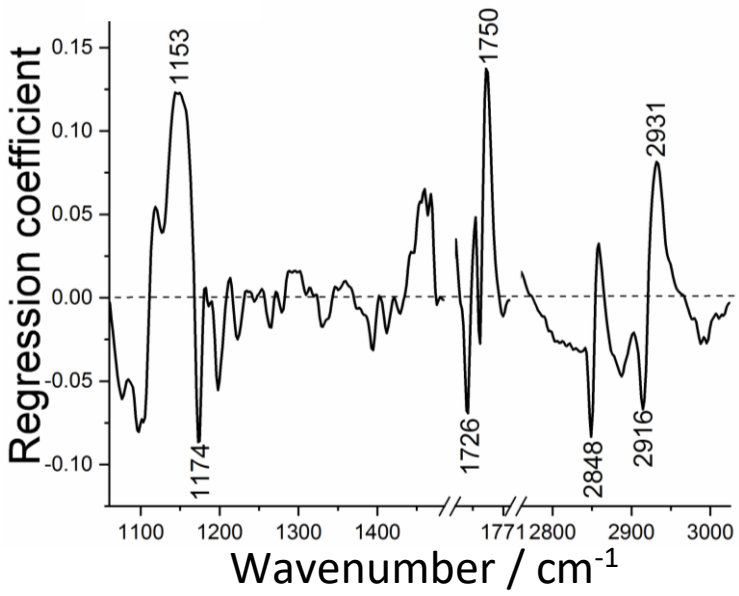

Figure 3.10: Regression coefficient showing spectral features influencing the % IMF content prediction in red meat using IR spectral data

3.5.5 Sensor (Data) fusion

3.5.5.1 Low-level fusion (LLF)

Preprocessed Raman and IR spectral data were concatenated to form a single matrix, then further utilized in building quantitative model for both pH and % IMF predictions. For pH assessment, the LLF model yielded a R_C^2 and $RMSEC$ value of 0.78 and 0.17, respectively. However, on validation against an independent test set, there was an increase in the $RMSEP$ to 0.30, which is equivalent to an $NRMSEP$ of 17.6 % (Figure 3.11). Comparison of the model performance of the LLF strategy to individual Raman and infrared models were performed to observe any possible improvement. The LLF model (R_P^2 = 0.65, $RMSEP$ = 0.30, $NRMSEP$ = 17.6 %) performed slightly better than the Raman technique (R_P^2 = 0.67, $RMSEP$ = 0.36, $RMSEP$ = 21.1 %) but not as well as the IR technique (R_P^2 = 0.82, $RMSEP$ = 0.20, $RMSEP$ = 11.8 %). This may be attributed to the presence of data not directly associated with the prediction, which in turn could worsen a model [173].

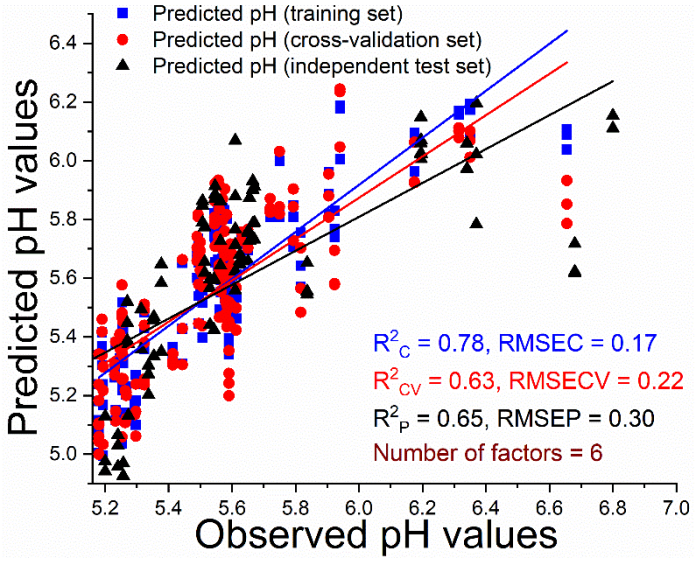

Figure 3.11: Regression line showing correlation of observed pH values with low-level fused Raman and IR spectral measurements.

For % IMF assessment, the combination of Raman and IR data at a low-level showed slight model improvement for the training set as shown in Table 3.3. Validation of the model against the test set yielded a $RMSEP$ of 0.30, corresponding to an $NRMSEP$ of 8.5 %. This marks an improved performance for predicting external samples. The low level fusion model performed similarly to the Raman data but better than the IR model in predicting % IMF on external samples. This fusion strategy looks promising for predicting pH and % IMF in intact meat samples.

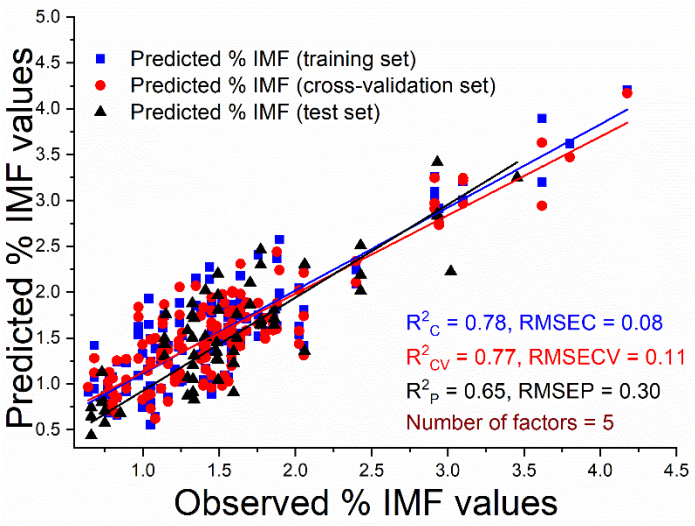

Figure 3.12: Regression line showing correlation of observed % IMF values with low-level fused Raman and IR spectral measurements.

3.5.5.2 Mid-level fusion (MLF)

As initially described, two data reduction techniques, PCA and MCR were utilized in extracting features for the mid-level data fusion. With the PCA data reduction step, a total of 11 PCs were extracted from the PCA scores plot of the individual instruments (6 and 5 PCs from the Raman and infrared instrument, respectively). This was concatenated into a single matrix and subsequently utilized in building PLS models. In the case of MCR, a total of 14 components were used as the extracted feature and utilized in building predictive models. Ideally, a larger number of latent variables should be utilised owing to the potential loss of information during the data reduction step, however, only 4 variables were selected based on the RMSECV to avoid overfitting of the models. Results obtained for the prediction of pH irrespective of the data reduction step did not yield a strong positive correlation with the observed reference values (R_C^2 = 0.52). Further validation of the models against the test set showed limited reliability with a R_P^2 values of 0.42 (PCA) and 0.33 (MCR). A plot of the MLF analysis is shown in Figure 3.13.

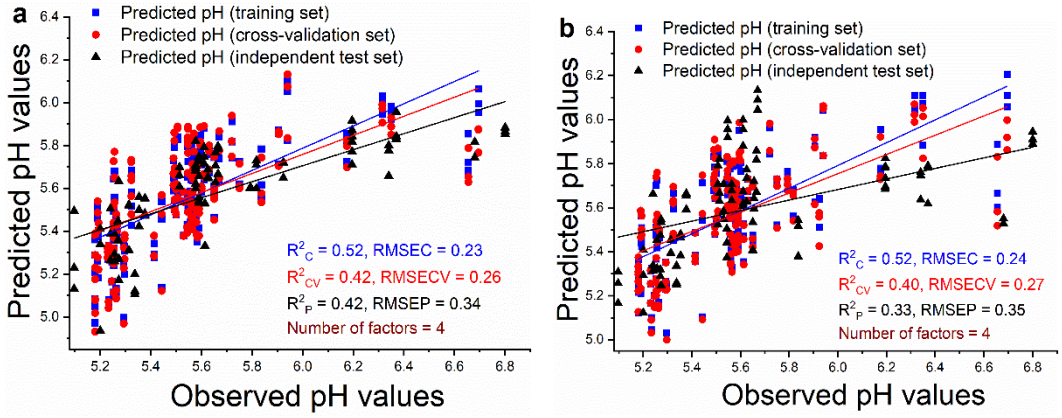

Figure 3.13: Regression line showing correlation of observed pH values with mid-level fused data extracted from Raman and IR spectral measurements using (a) PCA (b) MCR.

Similarly, the determination coefficient of the MLF models were not high enough to consider the model accurate for % IMF prediction (R_C^2 = 0.48 for PCA, R_C^2 = 0.48 for MCR). The outputs were worse on validation using both the cross-validation and test sets. The poor correlation exhibited by the MLF could be due to substantial loss of useful information during the feature extraction step. A summary of the model performance is presented in Table 3.3.

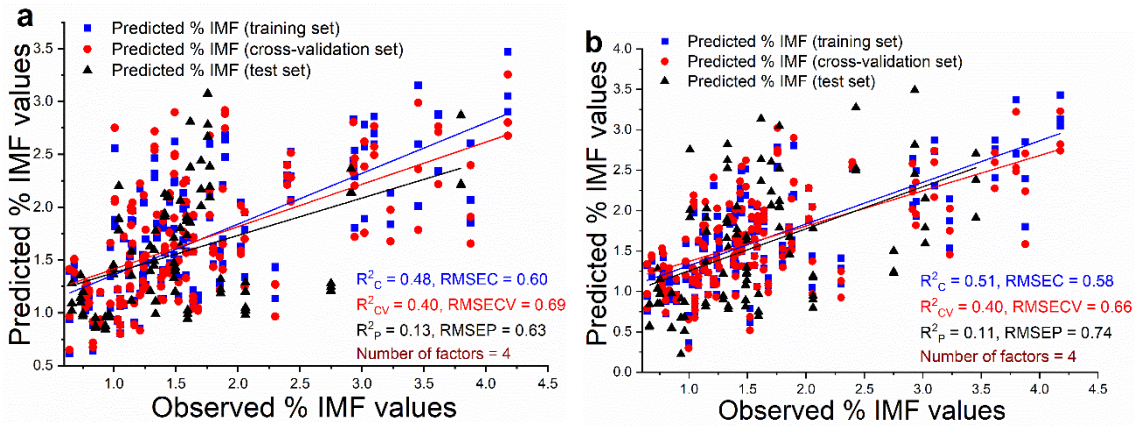

Figure 3.14: Regression line showing correlation of observed % IMF values with mid-level fused data extracted from Raman and IR spectral measurements using (a) PCA (b) MCR.

3.5.5.3 High-level fusion (HLF)

Independent model predictions were first calculated using the Raman and MIR spectral data, and the outputs for each sample averaged prior to performing the HLF. Results of the HLF strategy showed an improved predictive model particularly for pH predictions, with a R_P^2 and $RMSEP$ of 0.76 and 0.15, respectively. Interestingly, the R_P^2 of the HLF strategy was less than that of the independent IR predictions (R_P^2 = 0.82, $RMSEP$ = 0.20), yet had better $RMSEP$ for predicting external samples. This highlights that a larger R^2 doesn't always indicate a better model. The HLF strategy had the best performance of all data techniques employed for predicting pH, suggesting possible enhancement of information upon combination of the prediction outputs of the Raman and infrared systems for pH prediction.

The HLF strategy for predicting the % IMF in red meat yielded a R_P^2 of 0.46 ($NRMSEP$ = 17.2 %), which is a better performance over the MLF strategy (R_P^2 = 0.11, $NRMSEP$ = 20.9 %). However, it was worse on comparison to predictions by the individual instruments as well as LLF strategy.

Figure 3.15 suggests a possible non-linear relationship between the spectral data and the observed quality parameters. Description of this relationship would require the use of non-linear algorithms such as artificial neural network, support vector machine regression and local weighted regression. This was not carried out in the current study.

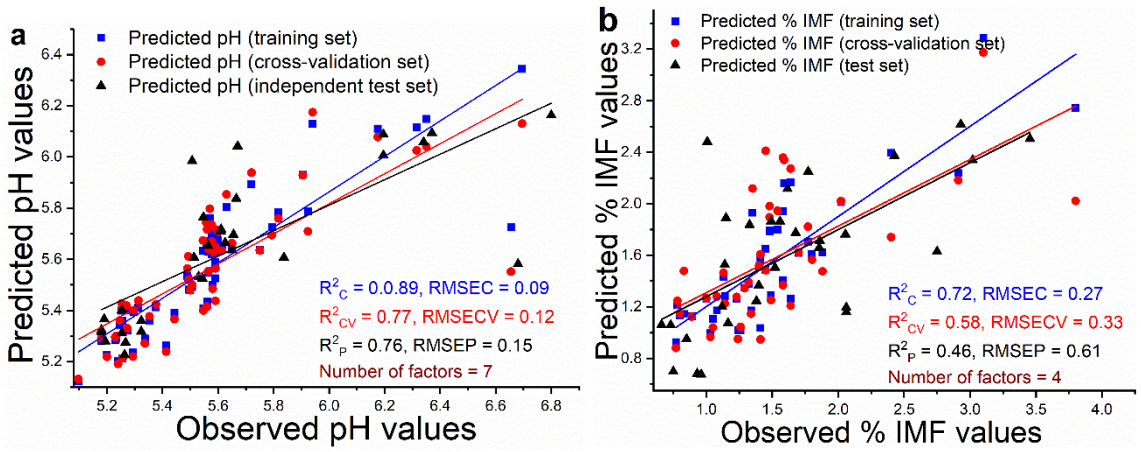

Figure 3.15: Regression line showing correlation of high-level fusion strategy of Raman and IR data to observed (a) pH values (b) % IMF values.

Attempts to improve the HLF model by improving individual instrument models was performed using a subset of the dataset. Ten (10) samples were selected at random and p models were developed using both Raman and IR spectral data. These models yielded improved performance statistics. HLF analysis was then performed using the improved predictions from both models, resulting in an improved performance (Figure 3.16). This suggests that HLF performance is only as good as the individual technique performances.

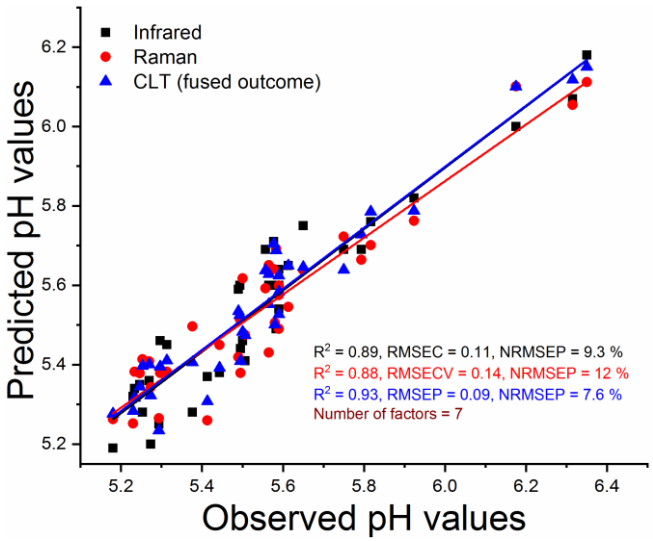

Figure 3.16: Illustration of improvement in the HLF strategy for prediction of pH. Improvement in individual spectroscopic models resulted in improvements in the HLF strategy

Overall, the data fusion strategies showed some improvements for quantifying pH and %IMF over single instruments particularly Raman. All possible fusion strategies were evaluated and their prediction performance summarised. Among the possible fusion strategies, the low-level fusion strategy worked the best, which could suggest modelling of more information. For example, it is difficult to detect the amide II Raman band in non-resonant conditions; however, this band can easily be detected using the ATR-FTIR instrument. Thus, this combination with additional information might lead to an improvement in model accuracies, although at the cost of a possible increase in noise

contribution. This can be avoided by selecting an adequate number of latent variables assessed using the RMSE.

Table 3.3: Results of PLSR models developed for predicting the pH and % IMF content of venison, beef and lamb using Raman, FTIR and data fusion strategies.

	Technique	R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2p	RMSEP	NRMSEP (%)	LV
pH	Raman	0.84	0.14	0.64	0.22	0.67	0.36	21.1	7
	Infrared	0.78	0.20	0.59	0.27	0.82	0.20	11.8	7
	Low – level	0.78	0.17	0.63	0.22	0.65	0.30	17.6	6
	Mid – level (PCA)	0.52	0.23	0.42	0.26	0.42	0.34	20.0	4
	Mid – level (MCR)	0.52	0.24	0.40	0.27	0.33	0.35	20.6	4
	High – level (CLT)	0.89	0.09	0.77	0.12	0.76	0.15	8.80	7
% IMF	Raman	0.75	0.22	0.72	0.26	0.60	0.30	8.47	4
	Infrared	0.70	0.35	0.62	0.44	0.67	0.40	11.3	4
	Low – level	0.78	0.08	0.77	0.11	0.65	0.30	8.47	5
	Mid – level (PCA)	0.48	0.60	0.40	0.69	0.13	0.63	17.8	4
	Mid – level (MCR)	0.51	0.58	0.40	0.66	0.11	0.74	20.9	4
	High level (CLT)	0.72	0.27	0.58	0.33	0.46	0.61	17.2	7

3.6 Conclusion

This study investigated the ability of Raman and infrared spectroscopy as well as data fusion strategies to predict the quality of red meat (pH and % IMF). Sample datasets representative of pH and % IMF ranges observed in New Zealand meat industry were utilized in building training models which were subsequently validated against an independent test set. Results obtained in this study yielded a $NRMSEP$ below 25 % for all techniques and parameters considered, despite some poor correlation particularly with the mid-level fusion strategy. The high-level fusion strategy had the best performance in predicting unknown red meat samples for pH, with a 25 % improvement in the $NRMSEP$ whereas the results for the low-level fusion strategy was comparable to the Raman results

for predicting the % IMF quality trait. Although significant model improvements were not observed upon concatenating the Raman and infrared data, the reduced $NRMSEP$ gives promise for the use of data fusion in future meat quality and discrimination studies.

Chapter 4

Isothermal dehydration of crystalline hydrates

4.1 Acknowledgments

The theoretical calculations presented in this study were performed by Kārlis Bērziņš of the Department of Chemistry, University of Otago, New Zealand.

4.2 Results publication

The results discussed in this chapter have been previously published:

Chima Robert, Sara J. Fraser-Miller, Kārlis Bērziņš, Peter O. Okeyo, Jukka Rantanen, Thomas Rades and Keith C. Gordon: Monitoring the isothermal dehydration of crystalline hydrates using low-frequency Raman spectroscopy. *Molecular Pharmaceutics*, 2021, *18*, 1264-1276.

My contribution to this paper included conceptualization, data collection, data curation, formal analysis, drafting of the original manuscript and correction of the reviewed manuscript.

4.3 Introduction

Active pharmaceutical ingredients (APIs) exist in different solid-state forms such as crystalline, amorphous, solvate, hydrate and salt forms [184]. These solid-state forms possess unique physiochemical and structural properties and are capable of undergoing transformations from one form to another. It is important to understand these transformations, as the presence of unwanted solid-state transformations in pharmaceutical products can lead to their poor therapeutic performance [185, 186].

Approximately one-third of existing crystalline APIs can incorporate water in their crystal lattice resulting in the formation of crystalline hydrates [185, 187]. Crystalline hydrates can possess different physiochemical properties from their anhydrous counterparts including density, free energy, thermodynamic activity, dissolution rate, stability and bioavailabilty [185, 188]. Also, the unique lattice framework in crystalline hydrates make them susceptible to changes including possible transformation to other solid-state forms [189]. The frequent use of water and heat in the drug manufacturing process facilitates solid-state transformation through hydrate formation and dehydration processes [43, 189-191]. It is therefore imperative to monitor these changes during the manufacturing process to ensure the target solid-state form in a specific drug product is present.

Several analytical techniques including powder X-ray diffraction (PXRD), solid-state nuclear magnetic resonance (ssNMR), differential scanning calorimetry (DSC), thermal gravimetric analysis (TGA), infrared spectroscopy (IR) and Raman spectroscopy (RS), have been utilized in investigating crystalline hydrates [192, 193]. Vibrational spectroscopy (infrared and Raman spectroscopy) can provide molecular level information, making these techniques potent for understanding and monitoring the hydration/dehydration processes. The sensitivity of Raman spectroscopy to changes in molecular arrangement and conformation as well as the negligible contribution of water to the Raman signal makes Raman ideal for application in hydrate formation and dehydration studies [194, 195]. Other advantages of Raman spectroscopy include the ease

of use, little to none sample preparation requirement, non-destructiveness and spectral specificity [196].

Conventional Raman instruments involve detection in the mid to high-frequency regions, typically around 200 – 4000 cm^{-1}, which are mainly associated with the intramolecular interactions within the molecule [197]. More subtle changes involving the intermolecular interactions occur at the low-frequency Raman region (<10 – 200 cm^{-1}) and cannot be observed using conventional Raman instruments. This is due to the over compensation of the optical method (e.g. notch filters) removing the low-frequency Raman signals during filtering of the extremely dominant Rayleigh scattering signal. However, advances in ultra-narrow optical filter technology (particularly volume Bragg gratings) have increased the application of low-frequency Raman (LFR) instruments, as they permit the Rayleigh scattering to be easily filtered without interfering with signals associated with vibrations in the LFR region [194, 195, 198, 199].

Vibrational spectra below 200 cm^{-1} posit great advantage for understanding phase transformation, crystallization and solid dosage formulation [197, 200]. The LFR region provides insight relating to intermolecular interactions like hydrogen bonding, $\pi - \pi$ stacking in aromatic rings and phonon modes that exist within the crystal lattice [197, 198, 201], explaining the growing interest in pharmaceutical applications including dehydration of crystalline hydrates [202, 203], solid-state dosage forms [13], phase transformation [204-207] and in-situ solubilisation [198, 208, 209].

Despite the increasing application of LFR, interpretation of the observed modes remains somewhat challenging and requires detailed theoretical calculations. This is further complicated by the molecular complexities and intricacies within the crystal structure as highlighted by Bērziņš et al. [210]. Notwithstanding these challenges, Raman spectroscopy and computational tools like density functional theory (DFT) continue to prove crucial in understanding physio-chemical and structural properties of compounds [211, 212]. The DFT computational models augment spectroscopic methods by providing assignments to vibrational modes below 200 cm^{-1} in organic molecules [213]. The

combination of the LFR and mid-frequency Raman (MFR) regions provides insight into solid-state transformation, owing to the availability of information relating to intra and intermolecular vibrations of the molecule.

In this study, piroxicam and theophylline were utilized as the model drugs in understanding dehydration of crystalline hydrates. Piroxicam (4-hydroxy-2-methyl-N-pyridin-2-yl-2H-1,2-benzothiazine-3-carboxamide 1,1-dioxide) is a poorly water-soluble nonsteroidal anti-inflammatory drug used in the treatment of arthritis [214]. Six solid-state forms of piroxicam have been identified in the Cambridge structural database (CSD) including five anhydrous and one monohydrate crystalline forms [215]. The published crystal structure of piroxicam monohydrate has been shown to have a more complex hydrogen-bonding network compared to the anhydrous forms. The solid-state forms used in this study were the piroxicam anhydrous form 1 (CSD: BIYSEH13) [216] (hereby denoted as PXA form I) and the monohydrate (CSD: CIDYAP02) [216] (hereby denoted as PXM).

Theophylline (1,3-dimethyl xanthine) is a widely prescribed respiratory drug for the treatment of asthma or chronic obstructive pulmonary disease [217] . Like piroxicam, theophylline exists in different solid-state forms, with four anhydrous and one monohydrate form already reported [218]. The solid-state forms used in this study were the anhydrous form II (CSD: BAPLOT01) [219] (hereby denoted as TPAH form II) and the monohydrate (CSD: THEOPH02) [220] (hereby denoted as TPMH). TPAH form II has been reported to be stable at room temperature whereas anhydrous form I is stable at high temperatures [221]. The structure of piroxicam and theophylline monohydrates is shown in Figure 4.1.

Analysis of large spectral data is made feasible with the use of various multivariate analytical tools including PCA [192], PLSR [13], window factor analysis [222] and MCR [50]. The MCR approach has garnered increased attention in spectral analysis as it offers the advantage of decomposing complex spectral mixtures into pure component spectra as well as providing the concentration profiles. This allows for both qualitative and

quantitative analysis [50]. MCR has been successfully applied in solid-state transformation studies involving powder X-ray powder diffraction [223] and Raman spectroscopy [224].

In the present study, the LFR and MFR regions as well as theoretical calculations were utilized to gain insight into the dehydration of crystalline hydrates. The simultaneous access to the LFR and MFR regions was made feasible by utilising a custom-built Raman instrument. Isothermal kinetic analysis was performed using MCR to highlight the solid-state forms (and/or metastable states) upon dehydration of the monohydrates, and DFT simulations were used to better understand the nature of the vibrational modes.

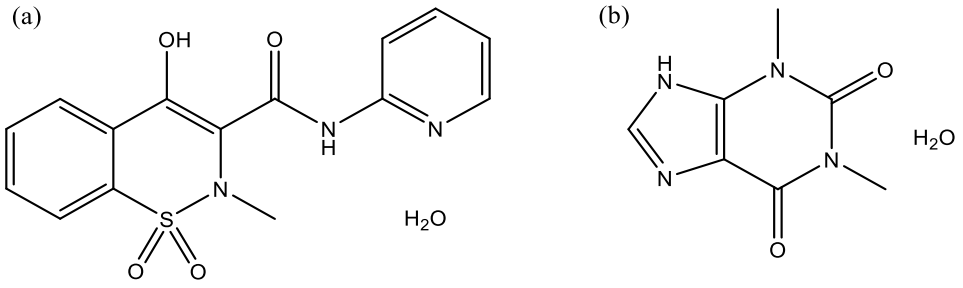

Figure 4.1: Molecular structures of (a) piroxicam monohydrate and (b) theophylline monohydrate.

4.4 Materials and methods

4.4.1 Sample preparation

Piroxicam (USP grade) was used as received (Hawkins Inc., Roseville, Minnesota, USA). Piroxicam monohydrate (PXM) was prepared by dissolving 1.5 g of anhydrous piroxicam in 1.5 L of distilled water and heating to 80 °C with stirring at 450 rpm for

approximately one hour, until fully dissolved. The solution was cooled overnight in a fume hood and vacuum filtration carried out to harvest the monohydrate. Characterisation of PXM was carried out by X-ray powder diffractometry (XRPD) as described by Lipiäinen et al. [13]. Further characterization was performed by comparing the MFR data to published references.

Theophylline monohydrate (TPMH) was prepared by dissolving 0.5 g of anhydrous theophylline in 50 mL distilled water at 60 ^{o}C using a stirring rate of 450 rpm for 30 min, until the anhydrous theophylline was fully dissolved. The solution was cooled overnight and the needle-shaped crystals of TPMH were harvested after 72 hours. Characterisation was carried out by comparing the conventional Raman spectrum to published references.

4.4.2 Isothermal dehydration of crystalline hydrates

Piroxicam and theophylline monohydrate samples, approximately 0.1 μm (PXM) and 10 μm (TPMH) sizes, were contained within a Tzero pan (TA Instruments Ltd, DE) and secured to the PE120 stage using a small amount of thermal paste. The temperature was altered using a PE95 controller, LinkPad PE120 variable-temperature stage and an ECP water supply (Linkam Scientific Instruments Ltd, UK). Isothermal dehydration was carried out at 95, 100, 105 and 110 ^{o}C for 6 hours (PXM) and at 50, 60, 70 and 80 ^{o}C for 90 minutes (TPMH); using a heating rate of 20 ^{o}C/min to reach the desired temperature.

4.4.3 Low-frequency Raman (LFR) spectroscopy

Raman spectra were acquired using a custom built LFR system (LFR – 785) that allows for simultaneous collection of data in both, the LFR and MFR [195]. The LFR setup included an excitation source from a 785 nm laser module (Ondax, Inc. Monrovia, CA, USA) which was filtered by BragGrate bandpass filters (OptiGrate Corp. Oviedo, FL, USA) to remove amplified spontaneous emission before irradiating the sample. Backscattered light from the sample was collected and filtered through a set of volume Bragg gratings (Ondax Inc., Monrovia, CA) and focused into an LS 785 spectrograph

(Princeton Instruments, Trenton, NJ, USA) where the light was dispersed onto a CCD detector (PIXIS 100 BR CCD, Princeton Instruments, Trenton, NJ, USA). The Raman data were calibrated using sulfur, 1,4 bis (2-methylstyryl) benzene (BMB), toluene and acetonitrile solvent (1:1) standards. Spectra were collected over the spectral window -365 to 2030 cm^{-1} with 5-7 cm^{-1} resolution, with a sample spot size of approximately 500 μm. Each spectrum consisted of 60 scans recorded with an integration time of 0.1 s each, accounting for a total readout time of 7.35 s. Dehydration measurements were carried out in triplicate for each dehydration temperature.

4.4.4 Reference measurements

Reference Raman data were collected using the LFR setup described above. An LNP95 liquid nitrogen cooler controlled variable temperature stage (THMS600; Linkam Scientific Instruments Ltd., UK) was used to horizontally hold the sample, which was secured in a quartz sample holder. Data collection was carried out at two different temperatures, initially at 20 ^{o}C then rapidly cooled to -190 ^{o}C. At each temperature point, the sample was held for at least 5 min to allow for temperature stabilization before acquiring the spectral data. Spectral acquisition was performed using the same setup described above and each spectrum consisted of 300 scans with a 1 s integration time.

4.5 Spectral data and multivariate analysis

4.5.1 Spectral pre-processing

Spectral data were pre-processed by performing cosmic spike removal across the full spectral region (-365 cm^{-1} to 2030 cm^{-1}) in the optical spectroscopy software Spectragryph v1.2.14 (Friedrich Menges, Germany); with further pre-processing carried out in The Unscrambler X 10.5 (CAMO, Norway). The LFR (20 - 300 cm^{-1}) and MFR (330 - 1800 cm^{-1}) regions were baseline corrected using linear baseline correction, with

standard normal variate (SNV) performed to normalise the intensity variations in the spectral measurements. Baseline offset was then performed on the normalised data to ensure non-negative variables were obtained for further analysis. The TPMH spectral data were further averaged (five spectra point) and Savitzky-Golay smoothing was performed using a first order polynomial and seven-point filter window to reduce the noise (particularly in the MFR region).

4.5.2 Multivariate analysis (chemometrics)

4.5.2.1 Principal component analysis (PCA)

Preliminary spectral data analysis was performed using PCA in The Unscrambler X 10.5 (CAMO, Norway). PCA was performed on data from both the LFR and MFR regions independently as well as in combination. PCA model was calculated using spectral data acquired at isothermal conditions.

4.5.2.2 Multivariate curve resolution (MCR)

Kinetic data analysis was carried out using multivariate curve resolution (MCR) in The Unscrambler X 10.5 (CAMO, Norway). MCR is a mathematical technique that resolves multicomponent mixture systems through a bilinear modelling of the pure component contributions. It resolves spectra mixtures (X) by decomposing them as a product of the component spectra (S) and contribution profiles (C).[225] MCR analysis was performed on the isothermal spectral data using different constraints including non-negativity and closure. These constraints were applied to optimize the MCR model and reduce the number of responses with chemical significance. Model specific settings (sensitivity to pure components: 50 and maximum number of alternating least square iterations: 100) were adopted.

4.6 Computational details

Theoretical solid-state density functional theory (DFT) calculations were performed using the fully periodic CRYSTAL17 [226] software package to model vibrational modes. A generalized gradient approximation Perdew-Burke-Ernzerhof (PBE) [227] functional with Van der Waals interactions treated according to the Grimme D3 [228] method was used, and all of the atoms were described with Ahlrichs' TZVP[229] basis set. The crystal structures of PXM (CSD ref: CIDYAP02),[216] PXA form I (CSD ref: BIYEH13) [216], TPMH (CSD ref: THEOPH02) [220] and TPAH form II (CSD ref: BAPLOT01) [219] were fully optimized with no implied restrictions on atom positions or lattice parameters, and the dielectric tensor and Raman intensities were calculated analytically using the coupled-perturbed Hartree-Fock/Kohn-Sham (CPHF/CPKS) approach [230]. Energy convergence criteria was set to $\Delta E \leq 10^{-8}$ and 10^{-10} Hartree for the geometry optimization and vibrational calculations, respectively. The calculated vibrational modes were visualized using the MOLDRAW 2.0 (version H1) software [231].

4.7 Results and discussion

The results and discussion are divided into four main components: exploratory analysis with PCA for PXM, MCR derived kinetics for dehydration of PXM, exploratory analysis with PCA for TPMH and MCR derived kinetics for dehydration of TPMH.

4.7.1 Exploratory analysis

An initial characterisation of PXM and TPMH was conducted in order to identify and monitor their solid-state forms. The commonly used MFR spectral data collected at ~20 ^{o}C were compared to published data in literature [13, 192]. Piroxicam showed characteristic Raman bands at 1400 cm^{-1}, 1464 cm^{-1} (PXM) and 1522 cm^{-1} (PXA form I) [232]. The observed signatures have been associated with various intramolecular

vibrational modes including conjugated CO moieties (1400 and 1464 cm^{-1}) and C-N stretch and N-H deformation (1522 cm^{-1}) (Figure A.2 in supplementary information). Comparison of the MFR spectral data of the different forms of theophylline showed bands at 1174, 1685 cm^{-1} (TPMH) and 1665, 1706 cm^{-1} (TPAH form II); which were consistent with those reported in literature [233, 234]. Peak assignment showed that the bands at 1653-1685 cm^{-1} is associated with carbonyl stretching modes in TPMH whereas the corresponding mode in TPAH form II appeared at 1665 and 1706 cm^{-1} [235]. Solid-state characterisation in the MFR region commonly involves peak shifts due to similarity in chemical composition of the different forms, however, geometry differences could result in the appearance of unique peaks which can facilitate characterisation [200].

In contrast to the MFR region, the LFR region probes lattice vibrational modes that are dependent on the long range order (i.e., crystallinity) and intermolecular interactions between participating molecules [193]. This interaction varies across different solid-state forms; as such, the LFR offers a unique way of identification and characterisation of different solid-state forms [197]. Comparison of the LFR and MFR reference data reveals that the bands in the LFR region had greater intensity and were distinctly different for the two crystalline forms of each of piroxicam and theophylline. The intense bands have been attributed to sizeable changes in the polarizability for these low-wavenumber vibrations [197, 236]. In the LFR region, the largest spectral difference in the four crystal forms were observed at Raman shift < 100 cm^{-1}. The lattice vibrations are expected to be observed in this region, and are dependent on a number of factors including the molecular-site symmetry and intermolecular forces. The number of symmetry-independent molecules in a crystal structure is represented as Z' [237, 238]. More defined peaks were observed for crystal phases with $Z' = 1$ (PXA, TPAH and TPMH) in contrast to broader peaks observed for PXM which has a $Z' = 2$ [216, 219, 220]. This could be due to the complex nature of the packing in PXM as shown Figure A.1 in supplementary information.

TPAH form II and TPMH have been reported to exhibit an ordered arrangement with the hydrogen-bonded network stacked over one another to form a π- π aromatic ring

stacking. In TPAH form II, the amine (NH) group and N atoms of the purine ring are involved in the hydrogen bond interaction whereas in TPMH, the NH···O bonds involving the NH group of the purine and one of the lactam CO groups result in dimer formation. These dimers are linked by the water molecules through an OH···N hydrogen bond with a N atom in the purine ring [197]. Conversely, the LFR signature of caffeine, a structural analogue of theophylline, is characterised by the absence of phonon peaks as reported by Hédoux et al [202]. This has been highlighted to be due to the orientational disorder around the C_6 molecular axis in caffeine [202, 204].

Investigation into the nature of vibrational frequencies and consequently Raman activities of piroxicam and theophylline solid-state forms was carried out in the CRYSTAL17 software environment. Comparison of the experimental and theoretical results are presented in Figure 4.2 and Figure 4.3. The theoretical results were in good agreement with the experimental data (especially with spectra obtained at -190 ^{o}C due to the nature of the theoretical simulations - full relaxation of atom positions and lattice parameters were carried out during the geometry optimization step which better represent the crystal structure near absolute zero). Information on the specific mode assignments for all four solid-state forms used in this study is presented in Table A1 – A4 in the supplementary information section.

Interestingly, the PXM model showed uncharacteristically low predicted Raman intensities for the modes in the LFR region, which could be related to their experimentally observed intrinsic similarity to the bands in the higher wavenumber domain. Nevertheless, it is important to note that while DFT-simulations for the optimized solid-state structures of PXM and TPAH yielded generally valid models. PXA and TPMH exhibited several (small) imaginary frequencies (one and three, respectively). Occurrence of such ambiguities for periodic calculations is not uncommon considering the complexity of these simulations (for example, PXM structure optimization and frequency analysis took approximately 51 days using 276 processor cores in parallel from Intel® Xeon® E5-2695 v4 CPUs), which typically appear due to numerical errors. Conceptually, these irregularities can be treated differently to *in vacuo* counterparts due

their often localized nature in periodic systems, *i.e.*, negligible impact on the rest of the data. It is important to note that attempts to further optimize the structures with tighter convergence criteria to resolve the negative frequencies were unsuccessful. Additionally, the use of an alternative crystal structure (for example, THEOPH01) [239] for the geometry input, not only generated similar problems for the simulation, but also inherited lesser accuracy for the respective peak positions and Raman intensities.

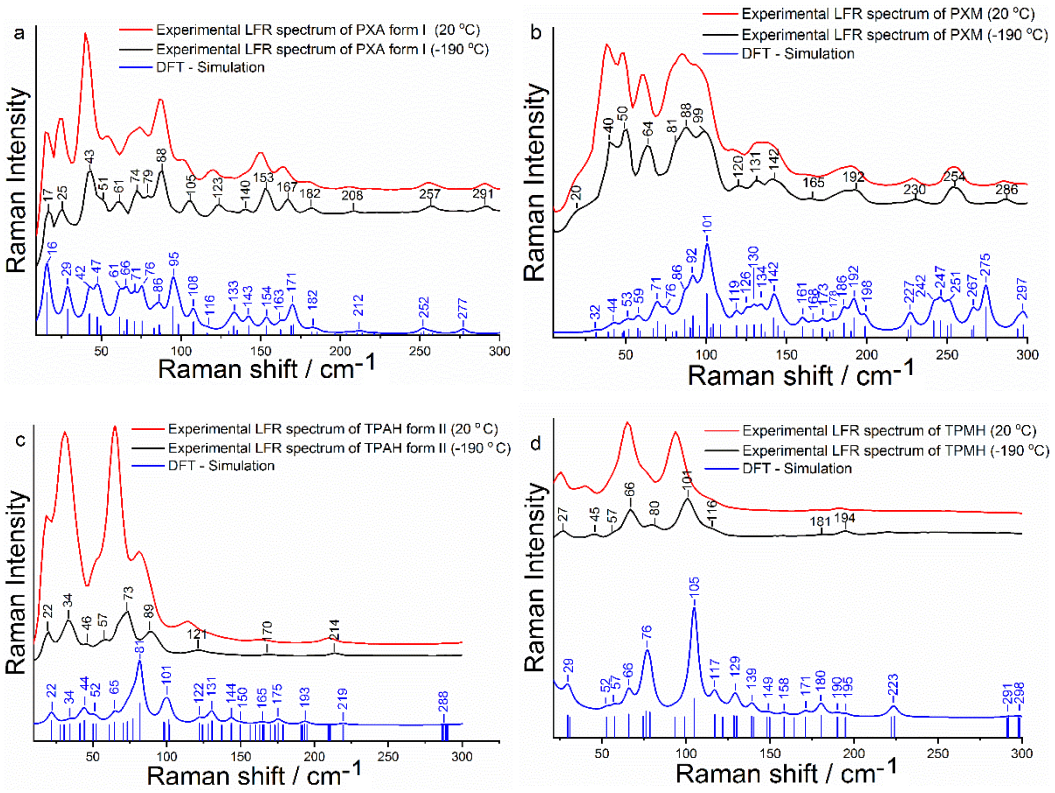

Figure 4.2: Experimental and DFT-simulated LFR spectra of (a) PXA form I (b) PXM (c) TPAH form II (d) TPMH. The theoretical data were in good agreement with experimental results particularly spectral data acquired at -190 °C.

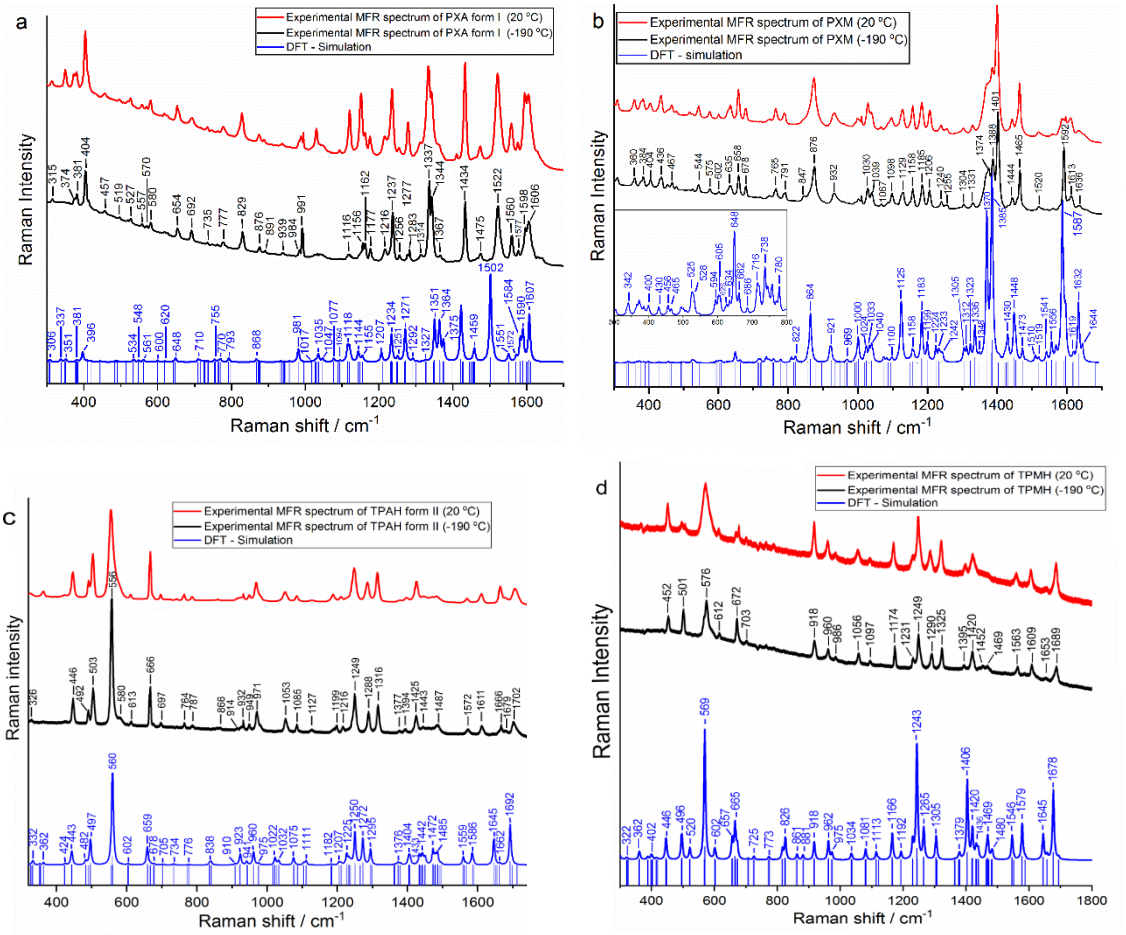

Figure 4.3: Experimental and DFT-simulated MFR spectra of (a) PXA form I (b) PXM (c) TPAH form II (d) TPMH. The experimental and theoretical results show good agreement particularly at -190 ºC.

4.7.2 Dehydration of piroxicam monohydrate

To investigate the dehydration of PXM, multiple isothermal experiments were carried out between 95 – 110 ºC for 6 hours using a low-frequency Raman setup. This system was capable of the simultaneous acquisition of the LFR and MFR regions, as such allowed investigation of both the intra- and intermolecular interactions within the crystal lattice of the solid-state forms. Complete dehydration of PXM was accompanied by a colour change of the starting material from yellow to white which was visually observed. Figure 4.4 shows representative Raman spectral data collected on dehydration of PXM at

110 ℃. Initial spectral assessment reveals that Raman signatures of the final spectral data for all temperatures were consistent with PXA form I (Figure A.2 in supplementary information). The presence of the characteristic PXA form I bands at 1522 cm^{-1} and a shoulder peak at 1475 cm^{-1} highlight that PXM dehydrates to PXA form I, which is in agreement with studies conducted by Sheth et al. and Kogermann et al. [192, 240].

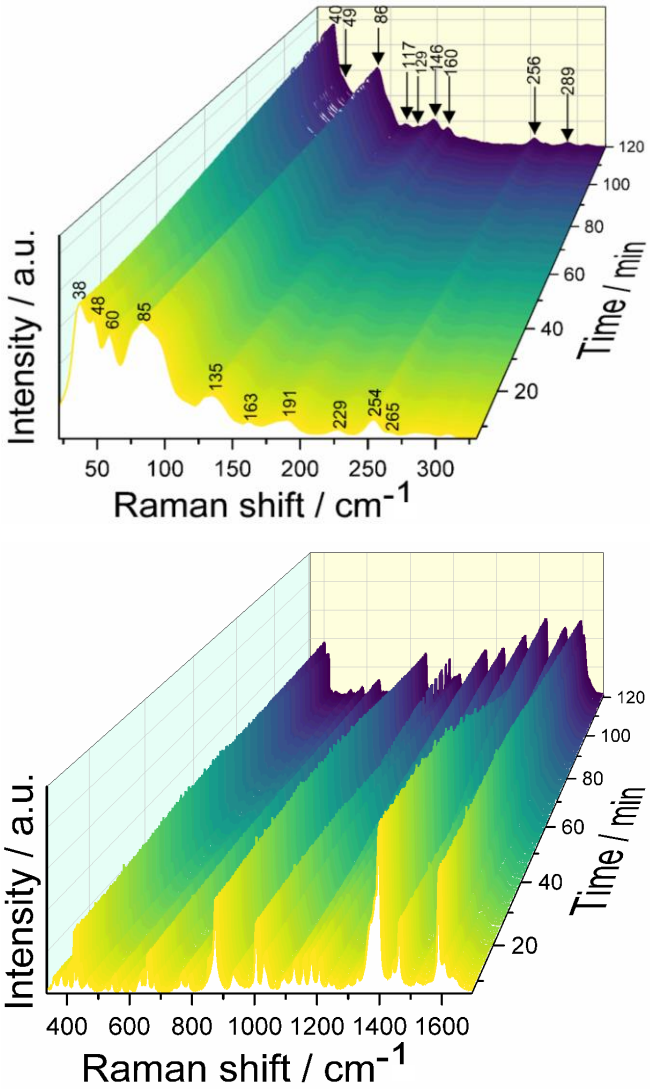

Figure 4.4: Representative LFR (top) and MFR (bottom) spectral data acquired during the isothermal dehydration of PXM at 95 ºC for 120 min.

In the low wavenumber domain, PXM showed strong peaks at 60 cm^{-1} and 85 cm^{-1}. Less intense peaks were observed at higher wavenumber domain (120 – 270 cm^{-1}). During dehydration, the intensity of the Raman band at 60 cm^{-1} decreased until it completely disappeared by the end of the experiment. This peak has no correspondence in the spectra of PXA form I, and could be attributed to vibrations involving water molecules.[241] Spectral changes in the MFR showed decrease in intensities of the PXM peaks at 1400 cm^{-1} and 1464 cm^{-1} until complete disappearance, with the formation of a shoulder peak at 1475 cm^{-1}.

PCA was carried out to better elucidate the dehydration behaviour of PXM and TPMH in isothermal conditions. PCA models were calculated using the LFR and MFR spectral data separately as well as in combination (LFR-MFR). As expected, faster dehydration was observed at higher temperatures. Examination of the loadings and scores plots using the conventional MFR region highlighted that 87 % of the total spectral variance was explained in the first principal component (PC1). The PC1 loadings plot had spectral contributions from both PXM (positive PC1 space) and PXA form I (negative PC1 space) (Figure 4.5), suggesting that the dehydration involves two solid-state forms. PC2 loadings was dominated by signals of PXA form I, which emphasises the final form of PXM dehydration. This finding was consistent on investigation of the LFR and the combined LFR-MFR spectral data. The loadings in the MFR was flipped on comparison to the LFR and LFR-MFR, with the PXM bands appearing in the negative PC space and the PXA form I bands appearing in the positive PC space. This is likely due to the way the algorithm models the different data regions. The loadings and scaled scores plots from the PCA analysis are presented in Figure 4.5. The PCA scores were scaled by performing mean normalisation on the data in The Unscrambler X software environment.

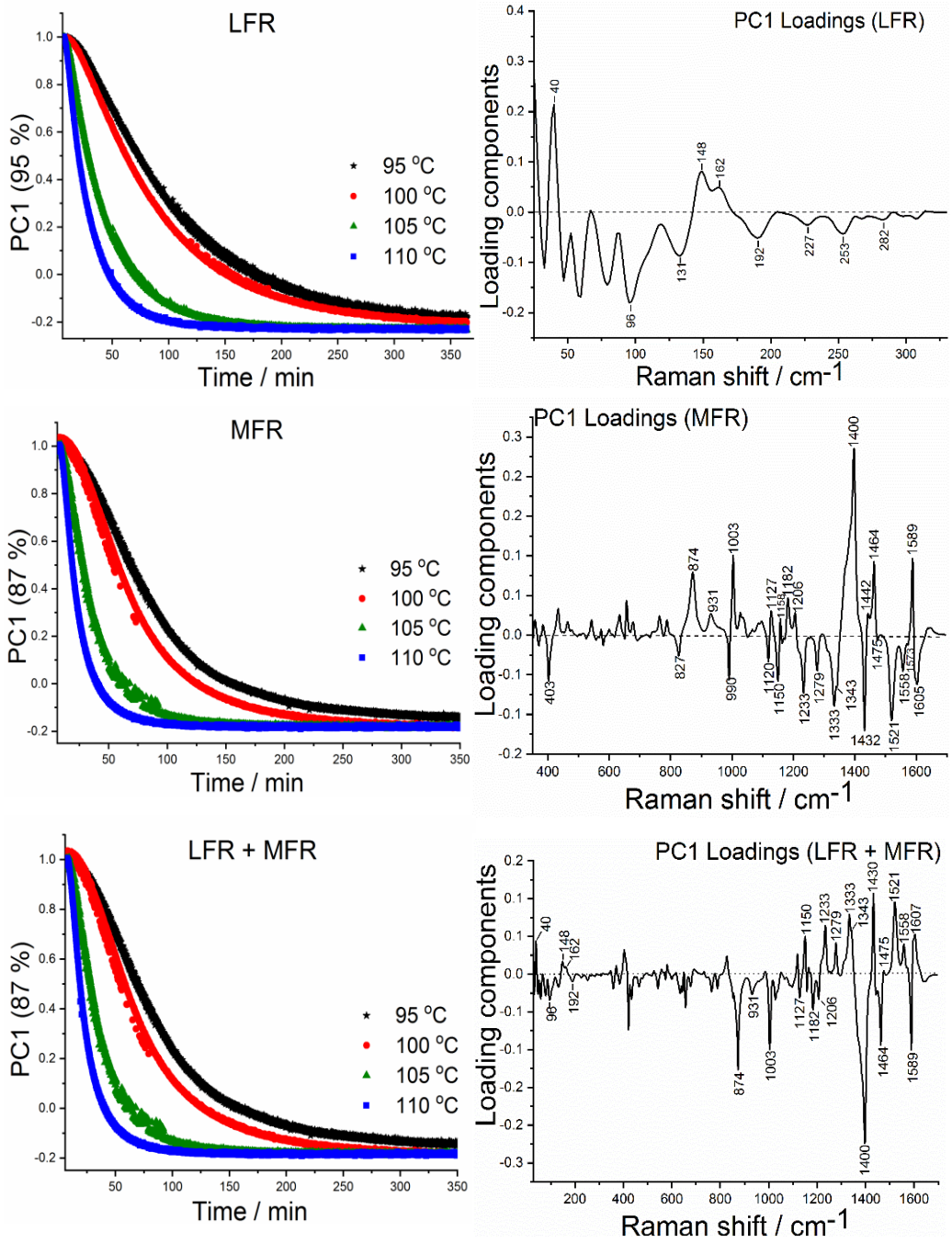

Figure 4.5: PC1 scores and loadings plot showing dehydration profiles of PXM in the LFR (top), MFR (middle) and both regions combined (bottom).

4.7.3 Multivariate curve resolution derived dehydration kinetics of piroxicam monohydrate

MCR analysis was performed to identify the contributing pure components within the acquired spectral data matrix as well as their concentration profiles. Like PCA, MCR was performed using the LFR and MFR data separately and in combination. MCR analysis was performed using different optimization techniques to facilitate complete resolution of spectra components. The MCR concentration profiles and component spectra are presented in Figure 4.6. The MCR model using the conventional MFR regime suggests the dehydration of PXM to be a two-component system. The MCR output shows the first spectral component (C1) to be consistent with Raman signatures of PXM whereas the second component (C2) resembles PXA form I. The concentration profile showed full resolution for the two components at all temperatures considered, with a decrease in C1 (PXM) from 1 to 0 associated with a corresponding increase in C2 (PXA form I) from 0 to 1. Similarly, the LFR MCR modelling revealed the dehydration to be a single-step process, with C1 having spectral features characteristic of PXM and C2 having features associated with PXA form I. However, the concentration profile was not sufficiently resolved at lower dehydration temperatures (particularly 95 and 100 ^{o}C). Optimization of the MCR analysis by introducing initial estimates of PXM and PXA form I yielded no improvement in the model output. Combination of LFR and MFR (LFR-MFR) spectral regions yielded improvements in the model output, with full resolution of C1 and C2 which could be indicative of an enhanced information obtained on combination of both regions.

Further investigation showed that the onset of dehydration of PXM was detected at different timescales by each of the LFR and MFR regions. The LFR region identified the onset of dehydration by approximately 2 min before the MFR. A representative plot is shown in Figure 4.7. This could be indicative of the increased sensitivity of the LFR region to solid-state changes compared to the MFR region. This sensitivity can be related to the complex hydrogen-bonding network present in PXM as shown in Figure A.1

(supplementary information), wherein the four molecules of piroxicam forms two hydrogen bonded dimers and the water molecules for a central tetrameric structure that connects the two dimers.

Many factors can influence the dehydration kinetics, with the crystal structure being the most important. Knowing the crystal structure and the energy of interactions between the water and host molecules facilitates the explanation of the dehydration process [242]. PXA form I exists in a dimer form, where the intermolecular hydrogen bonding occurs between an amide NH group and oxygen in sulfoxide group [243] whereas PXM forms a central tetrameric structure involving two piroxicam dimers bridged by a hydrogen bond network involving the water molecule and the amide carbonyl group. Two of these dimers are connected by the independent water molecules forming a tetramer [244]. It was thus considered that the low-frequency region is detecting the increasing initial mobility (and therefore lack of order) in the "columns" of H-bonded water molecules within the crystal framework. However, investigation into the crystal structure of PXM revealed the absence of channels, layers or tunnels that can accommodate water molecules, as such, preferential escape of water molecules of a crystalline hydrate along tunnels could not be used to explain the dehydration of PXM. The dehydration process of PXM has been reported to occur as a result of hydrogen-bond dissociation corresponding to inward advance of the phase boundary from the surface to the center of the crystal, explaining the isothermal fit to a two-dimensional phase boundary reaction model (R2) [240].

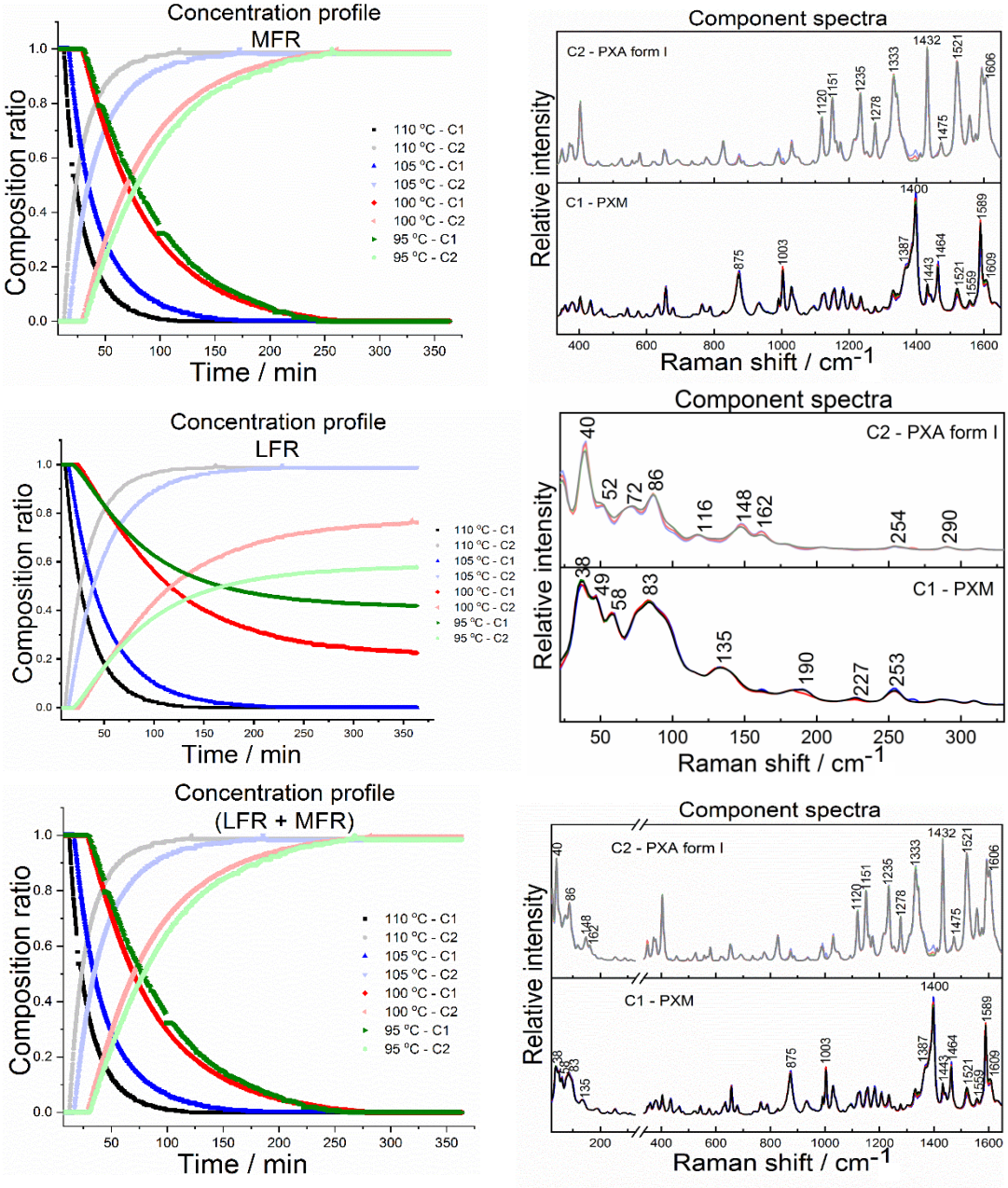

Figure 4.6: MCR analysis results showing the concentration profiles and component spectra on dehydration of PXM at different temperatures for LFR (top), MFR (middle) and both regions combined (bottom).

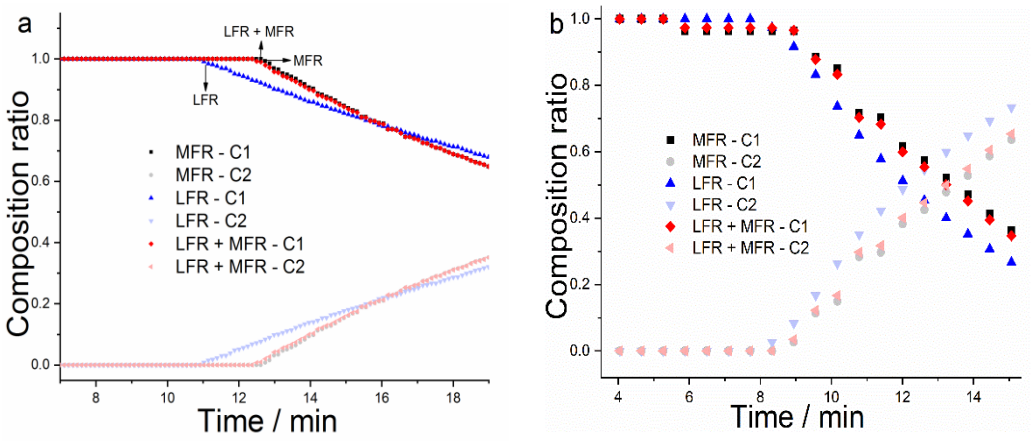

Figure 4.7: Representation of the dehydration kinetics for the low and mid-frequency regions showing detection of changes at different timescales (a) PXM (110 ^{o}C) (b) TPMH (80 ^{o}C).

4.7.4 Dehydration of theophylline monohydrate

Dehydration of TPMH was monitored at four different temperatures (50, 60, 70 and 80 ^{o}C). Figure 4.8 shows the acquired Raman spectral data in both the LFR and MFR regions for the dehydration of theophylline monohydrate at 60 ^{o}C. Evaluation of the acquired spectral data reveals the formation of TPAH form II on complete dehydration (as shown in Figure D3 in supplementary section). In the LFR region, TPMH showed bands at 26, 42 and 94 cm^{-1} which diminished in intensity and disappeared completely by the end of the experiment. This was associated with an appearance of a shoulder at 115 cm^{-1}, growing into a defined TPAH form II band. Other LFR bands were observed at 19, 33, 65 and 81 cm^{-1} which were consistent with the signatures of TPAH form II [245]. Observation into the conventional MFR region reveals the formation of characteristic TPAH form II carbonyl bands at 1667 and 1704 cm^{-1}, with the corresponding mode in the monohydrate form appearing at 1686 cm^{-1} [235].

The dehydration behaviour of TPMH was evaluated using PCA, with the scores and loadings plots presented in Figure 8. In the LFR domain, 95 % of the total variance was explained by the first principal component. PC1 loadings showed Raman signatures

corresponding to both TPMH and TPAH form II highlighting the composition of the dehydration process. PC1 scores plot shows the dehydration kinetics on conversion of TPMH. As expected, faster dehydration was observed at higher temperatures. In the MFR region, PC1 explained 64 % of the total spectral variance with the loadings plot having features associated with both TPMH (positive PC space) and TPAH form II (negative PC space) suggesting the dehydration of TPMH to be a two component system. A similar observation was made upon combination of both LFR and MFR regions. The PCA scores and loadings plots are shown in Figure 4.9.

4.7.5 Multivariate curve resolution derived dehydration kinetics of theophylline monohydrate

Isothermal dehydration kinetic modelling using MCR was utilized to resolve the spectral matrix into respective pure components. The MCR concentration profile (Figure 4.10) using the MFR and combined LFR-MFR regions suggest the dehydration of TPMH to be a multi-component system at 50 ^{o}C. It is of note that using only the LFR region, the dehydration kinetics implies a 2-component system for all temperatures considered. In the MFR and combined LFR-MFR regions, an intermediate state (denoted as TPMS) could be detected 50 ^{o}C but was absent for other temperatures. The absence of any intermediate state at higher temperatures could be attributed to a faster dehydration [246]. The loadings plot indicates that the first and third component can be associated with TPMH and TPAH form II, respectively. Whereas the second component corresponds to the intermediate state. Conversely, MCR resolved spectral data from the LFR region to consist of only two components. This might be indicative of the absence of adequate long-range order in the intermediate state as to be detected in the LFR spectral range.

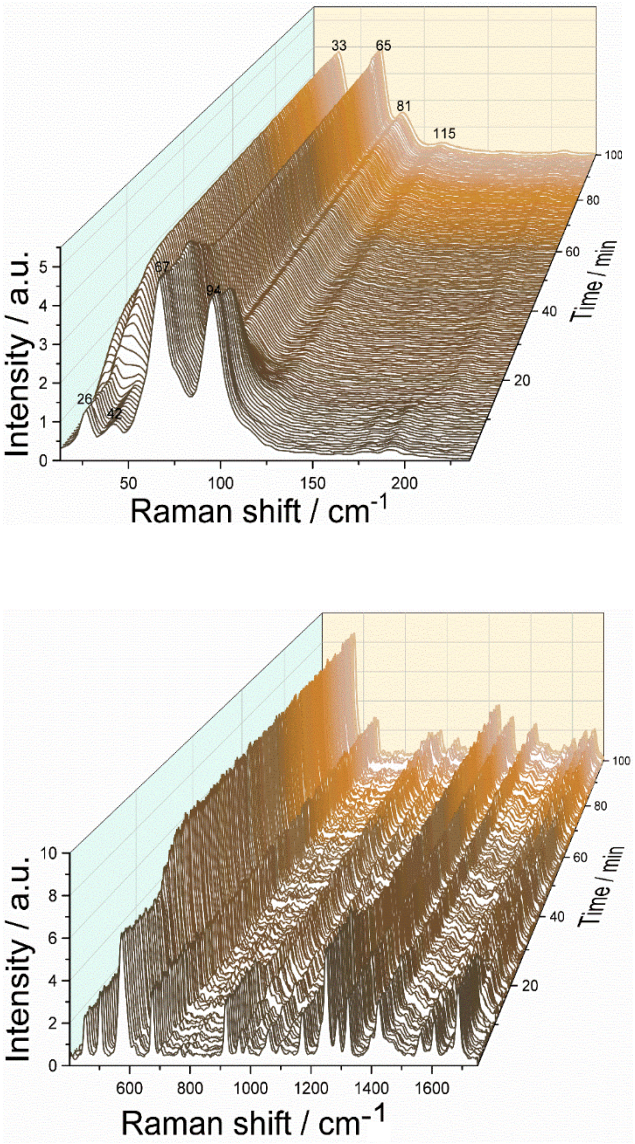

Figure 4.8: Representative Raman spectral data acquired during the isothermal dehydration of theophylline monohydrate at 60 ^{o}C for 100 min.

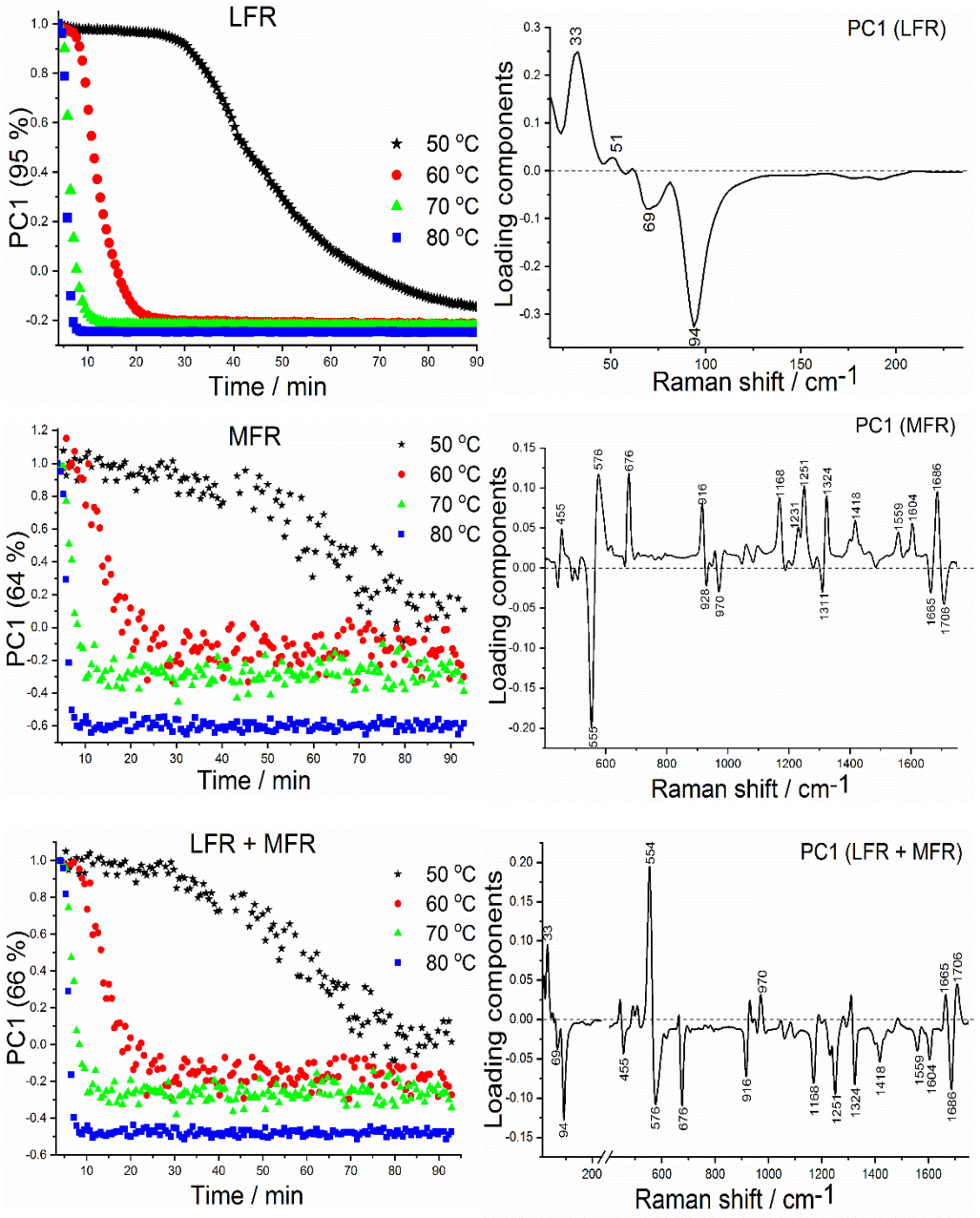

Figure 4.9: PC1 scores and loadings plot showing dehydration profile of TPMH in the LFR, MFR and both regions combined.

Similar to PXM, early onset of dehydration was first observed in the LFR region by approximately 1.5 min before the MFR (Figure 4.7b) which could be suggestive of the increased sensitivity to changes in the hydrogen bonding network and the disappearance of the water molecules. Studies involving direct process measurements during

fluidization with in-line near infrared (NIR) spectroscopy coupled with in-line Raman spectroscopy suggest dehydration of TPMH involving the disappearance of water detected with NIR a few minutes prior to structural rearrangements observed with Raman spectroscopy [247]. TPAH form II and TPMH have been reported to exhibit an ordered arrangement with the hydrogen-bonded network stacked over one another to form a π- π aromatic ring stacking. In TPAH form II, the amine (NH) group and N atoms of the purine ring is involved in the hydrogen bond interaction whereas in TPMH, the NH···O bonds involving the NH group of the purine and one of the lactam CO groups result in a dimer formation. These dimers are linked by the water molecules through an OH···N hydrogen bond with a N atom in the purine ring [197].

A different mechanism has been reported for the dehydration of TPMH which could be suggestive of the complexities of this process. Dehydration of TPMH has been reported to occur via a two-step process depending on the experimental conditions (including dehydration temperature) [234, 248-250]. Duddu et al. [251] reported the dehydration of TPMH to occur via a two-step process at 40 ^{o}C and a one-step process at 47 ^{o}C, while Karjalainen et al. (2005) reported the appearance of intermediate state for dehydration at temperatures above 40 ^{o}C [251, 252].

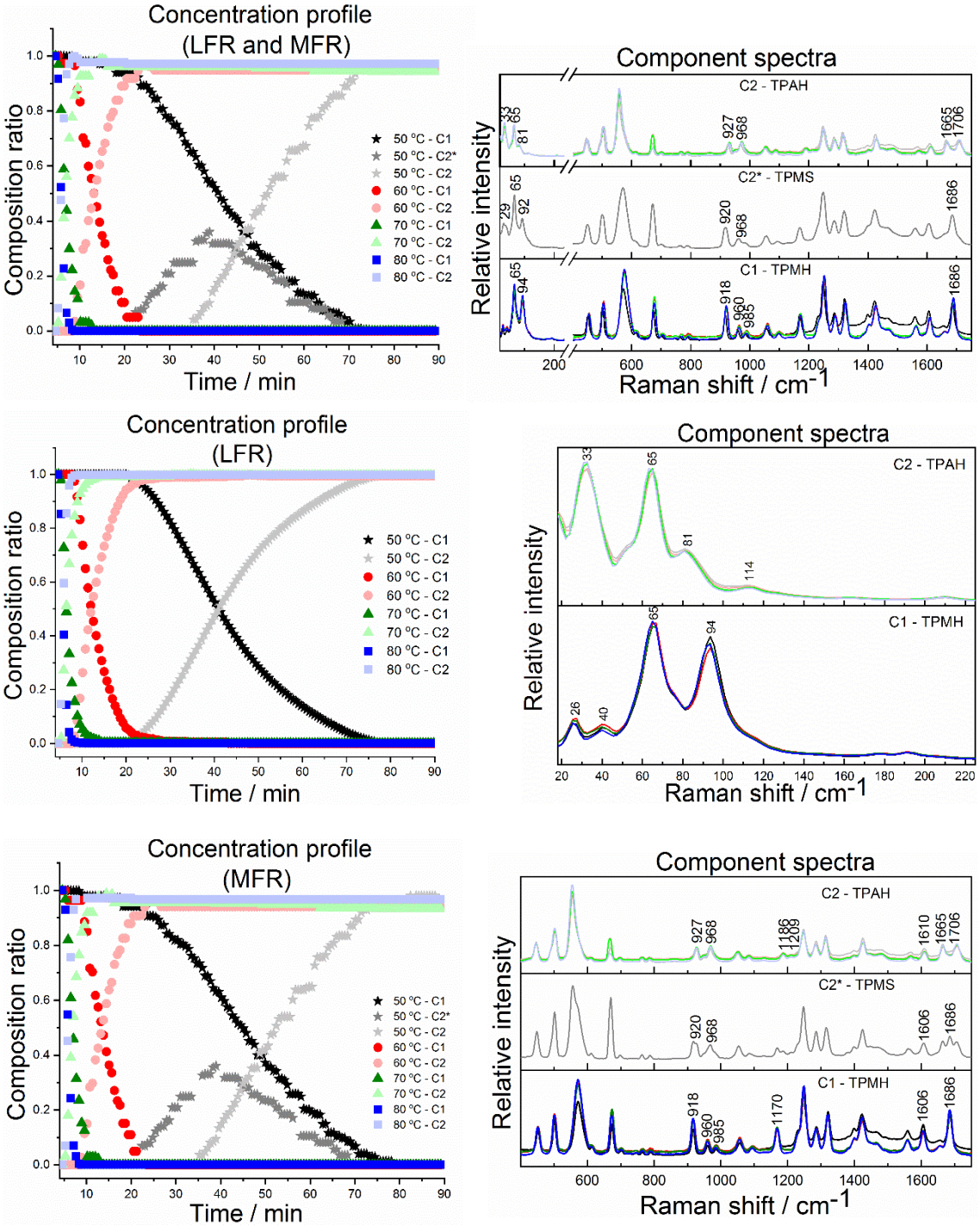

Figure 4.10: MCR analysis output showing the contributing concentration profiles and component spectra on dehydration of TPMH.

4.8 Conclusions

Isothermal dehydration of PXM and TPMH was performed using low frequency Raman spectroscopy, with multivariate analysis carried out using PCA and MCR. Dehydration changes showed the formation of anhydrous form I and anhydrous theophylline form II on complete dehydration of the monohydrates, which were consistent with literature. MCR kinetic analysis highlighted that the dehydration of PXM involves a two-component system (one-step dehydration) whereas the dehydration of TPMH suggests a multi-step process in the MFR at 50 ^{o}C and single-step process at higher temperatures may be indicative of the considerable influence of dehydration conditions as well as the heating method on the kinetics. In addition, the probing technique allowed for detection of different information with the LFR region identifying early onset of dehydration and loss of water molecules whereas the MFR highlighted structural rearrangement in the molecules. DFT calculations were informative on the nature of the vibrations and Raman bands particularly below 200 cm^{-1}. Based on the results, the use of both the LFR and MFR proved more informative and showed complete resolution upon MCR analysis as compared to the use of individual regions. This chapter highlights the importance of understanding the dehydration process in crystal hydrates, to ensure the presence of target solid-state form. The use of both the LFR and MFR regions facilitated the probing of both the intra- and intermolecular interactions within the molecule.

Chapter 5

Discrimination of phenotypes of immune cells

5.1 Acknowledgments

Cell culture and immunology studies on samples discussed in this chapter were performed by Julia Tsiampali (Neurosurgery Department, University Hospital Duesseldorf, Germany), Silke Neumann and Donata Maciaczyk, both of Department of pathology, University of Otago, Dunedin, New Zealand.

5.2 Results publication

The results discussed in this chapter have been previously published:

Chima Robert, Julia Tsiampali, Sara J. Fraser-Miller, Silke Neumann, Donata Maciaczyk, Sarah L. Young, Jaroslaw Maciaczyk and Keith C. Gordon: Molecular monitoring of glioblastoma's immunogenicity using a combination of Raman spectroscopy and chemometrics. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 2021, 252, 119534.

My contribution to this paper included data collection, data curation, formal analysis, drafting of the original manuscript and correction of the reviewed manuscript.

5.3 Introduction

Raman spectroscopy (RS) has garnered increased attention as a rapid and non-destructive diagnostic tool for uncovering molecular basis of diseases as well as for evaluating treatment therapies [253]. The contention for this method in biological and clinical spectroscopy is considerably justified owing to its non-invasiveness, label-free requirement and negligible interference with water molecules. RS probes the vibration of bonds in a sample yielding molecular specific information on the composition/structure of the molecule; which can be utilized for qualitative and quantitative analysis [5, 254]. RS is a two-photon process involving the inelastic scattering of an incident photon from a monochromatic light source (laser) on interaction with a sample. The resulting energy difference between the incident and scattered photon referred to as Raman shift provides information on the vibration of the chemical bonds present in a sample [113]. A typical Raman spectrum of a biological cell comprises different overlapping peaks relating to the composition of macromolecules like nucleic acids, lipids, protein, carbohydrates and other metabolites [255, 256].

RS has been widely applied in clinical spectroscopy, particularly in cancer diagnosis and discrimination, with several studies reporting the sensitivity of RS to changes in molecular conformation and composition in both cells and tissues owing to carcinogenesis [83, 84, 257, 258]. The biochemical changes resulting on transition of normal to malignant tissue (glioblastoma) have been investigated using RS [259-261], with increased cholesterol (-esters) levels reported in the necrotic tissue [261]. RS has also been utilized in the discrimination of normal and malignant cells (tissues) in brain tumour, with some studies reporting an accuracy of over 98 % [262, 263].

Despite the apparent applicability of RS in biological cell (tissue) studies, the inherent weak nature of Raman scattering limits its application in clinical settings. However, recent advancements in optical technologies allowing for integration of RS and confocal microscopy have led to improvements in spectral resolution as well as the possibility for live cell imaging [264, 265].

Glioblastomas (GBMs) are the most common and malignant primary brain tumours with an extremely aggressive clinical phenotype and very poor prognosis [266, 267]. One of the biggest challenges in GBM's effective treatment is the presence of highly invasive tumour cells called glioblastoma stem-cells (GSCs) that disseminate into the normal brain parenchyma and lead to tumour recurrence [268]. GSCs are enriched by the molecular pathway called epithelial to mesenchymal-like (EMT-like) transition and its activator, Zinc Finger E-Box Binding Homeobox 1 (ZEB1), plays a key role in glioma cell invasion [269, 270]. Most recently, an ecto-5'-nucleotidase (NT5E) known as CD73, that catalyses the conversion of adenosine mono phosphate (AMP) to adenosine (ADO), has been shown to regulate EMT in cancers [271]. Both CD73 and the EMT status have been shown to regulate cancer cell immunogenicity [272, 273].

The immune system and its responses can be broadly divided into two parts - innate and adaptive immunity with these two branches interacting closely to efficiently clear the body of invading pathogens and malignant cells. Innate immune cells, such as monocytes, are able to detect and phagocytose cancerous cells [274]. Monocytes can differentiate into macrophages (MΦs) and dendritic cells (DCs), which play important roles in priming antigen-specific adaptive immune responses [275]. Adaptive immune responses, mediated by B- and T-cells, develop more slowly than innate immune responses. They provide antigen-specific responses and are associated with immunological memory, leading to enhance response to subsequent encounter of same pathogen [276]. This is the basis of vaccination and an important mechanism to prevent cancer recurrence.

B- and T-cells are types of white blood cells, named after the organs where they develop, thymus gland (for T-cell) and bursa of Fabricius in birds (for B-cell) [277]. Two subtypes of T-cells carry out immune-mediated cell death, which is one of the main function of T-cells. These subtypes are CD8+ ("killer") and CD4+ ("helper") T-cells, with the names highlighting the cell surface protein present (CD8 or CD4). The CD8+ T-cells are cytotoxic, which entails that they directly kill virus-infected cells as well as cancer cells; whereas, CD4+ T-cells function by signalling other part of the immune system (by secreting cytokine) to a specific, perceived threat [278, 279].

In GBM patients, circulating monocytes express higher programmed death-ligand 1 (PD-L1) levels compared to healthy controls, which leads to suppression of T-cell responses [280]. This is the down-regulation of the response of the immune system to the cells of the human body by promoting self-tolerance and suppressing T-cell inflammation. It is responsible for preventing autoimmune diseases but can prevent the immune system from killing cancer cells. Furthermore, it has been shown that $CD8^+$ and $CD4^+$ T-cells are exhausted in GBMs and are incapable of driving an effector immune response [281].

Given the urgent need of developing techniques for cancer screening, diagnosis, and intraoperative surgical guidance, Raman spectroscopy has emerged as a non-invasive therapy guidance and diagnostic tool and has been used to define invasive margin of GBMs [282, 283]. The entire Raman spectra of a single cell referred to as "ramanome" is informative on the composition of all the molecules present in the cell. It is expected that different treatment of the cells will lead to changes in physiological and morphological features, such as distinct Raman spectrum [284]. Interpretation of Raman spectra for qualitative purposes have involved both univariate and multivariate approaches. The univariate approach entails visual inspection of one or few Raman peaks belonging to specific chemical components; and making comparison on the intensity changes in such peaks. This technique howbeit informative, results in loss of spectral information in the unselected regions. This challenge can be overcome by utilizing a global spectral (multivariate) approach using a single or combination of chemometric tools such as principal component analysis (PCA) [285], linear discriminant analysis (LDA) [286], K-means clustering [287] and support vector machine classification (SVMC) [288]. These tools allow for reduction in dimensionality in large datasets; facilitating interpretability while minimizing loss of information [289].

Linear discriminant analysis (LDA) is a supervised feature extraction tool utilized for qualitative studies. The LDA algorithms provide linear transformation of *n*-dimensional feature vectors (samples) into an m-dimensional space usually called discriminant variables (where $m < n$), such that the distance between-class variance is maximized relative to the within-class variance [38]. The requirement of samples in the training set being larger than the number of

variables poses a challenge to the use of LDA. This constraint however is overcome by utilizing PCA-LDA, where PCA calculations are performed for reduction of data dimensionality prior to LDA calculation. The first four PCs were utilized in performing the PCA-LDA calculations in this study. In PCA-LDA, samples are grouped into predefined classes prior to model creation. Similarly, SVM classification is supervised classification technique, with samples assigned to predefined groups prior to model creation.

In the present study, confocal Raman micro-spectroscopy in combination with two chemometric tools (PCA-LDA and SVMC) was utilized to detect molecular differences of monocytes and T-cells from a single blood donor after incubation with TCM of GSCs upon ZEB1 inhibition and, therefore, EMT modulation, or CD73 inhibition. Furthermore, since EMT and CD73 are critical in cancer progression and chemoresistance [270, 290], confocal Raman micro-spectroscopy in combination with two chemometric tools were utilized to detect differences of the cells upon ZEB1 and CD73 inhibition. The differences in the phenotype of the immune cells were also assessed by flow cytometry but not discussed, as it is outside the scope of this thesis [37].

5.4 Materials and methods

5.4.1 Cell culture

Cell samples were utilized as received from the Department of Pathology, University of Otago, New Zealand. Three human glioblastoma cell lines were used in this study. JHH520 cells were provided by G. Riggins (Baltimore, USA), GBM1 by A. Vescovi (Milan, Italy) and the paediatric GBM cell line SF188 was provided by E. Raabe (Baltimore, USA).

5.4.2 Raman instrument and cell measurements

Raman spectral data were collected using an Alpha 300R+ confocal Raman microscope controlled with Project 5.1 software (WITec GmbH, Ulm, Germany). Daily instrument calibration was performed using a silicon wafer sample with Raman peak position at

approximately 520.6 cm^{-1}. Measurement parameters included a laser excitation wavelength of 532 nm (Coherent, California) at 20 mW power, with the 50x dry objective (Zeiss, Oberkocken, Germany) resulting in a laser spot size of 412 nm (refer to Equation 1.3). Single cell (spot) measurements were randomly collected from the cytoplasm of live non-fixed cells dispersed in phosphate buffer saline (PBS) on a glass slide. The cells and PBS were kept on ice prior to acquisition of spectral data in order to avoid metabolic changes. Additionally, the drop of cells in PBS placed on the glass slide was renewed within 10 min of measurements. Each spectrum consisted of 60 scans recorded with an integration time of 1 s over a spectral window of -55 to 3789 cm^{-1} resulting in a read out time of 63 s. Twenty (20) cells were measured for each treatment condition. The laboratory temperature was maintained at ~20 ^{o}C.

5.4.3 Spectral preprocessing

Spectral data were pre-processed in the Project 5.1 software environment (WITec, Ulm, Germany). Spectral data were first corrected for cosmic spikes using the cosmic ray removal tool (filter size = 4, dynamic factor = 4). The obtained spectra were background subtracted using the dynamic 'shape' background algorithm (shape size = 300) to remove any fluorescence contribution. Standard normal variate (SNV) transformation was performed over the selected spectral region (725 to 1825 cm^{-1}) in The Unscrambler X v10.5 (CAMO, Norway) to account for intensity variations across the collected spectral data.

5.4.4 Multivariate data analysis (chemometrics)

In the present study, PCA was utilized as an unsupervised exploratory analysis tool whereas PCA-LDA and SVMC were employed as a supervised multivariate analysis tool for discrimination of the cells. All data analysis was performed in The Unscrambler X v10.5 (CAMO, Norway) using the pre-processed spectral data region (725 to 1825 cm^{-1}).

5.4.4.1 Exploratory analysis - Principal component analysis (PCA)

The PCA model was calculated using a k-fold (k=3) random cross validation procedure and the nonlinear iterative partial least squares (NIPALS) algorithm to enhance model optimization. PCA was carried out on T-cells and monocytes (after incubation with TCMs) independently.

Discrimination of phenotypes of immune cells were performed using two MVA tools including PCA-LDA and SVMC, which have been described in the introduction. Hence, only a brief review is presented herein.

5.4.4.2 Principal component-Linear discriminant analysis (PCA-LDA)

Principal component - Linear discriminant analysis (PCA-LDA) is a supervised feature extraction tool utilized for qualitative studies. The PCA-LDA model was created using 67 % of the dataset (training set) and the model validated against 33 % of the remaining dataset (test set). PCA-LDA was performed on each of the three cell lines with their respective control and knockdowns (pLKO.1, shCD73 and shZEB1) for both T-cells and monocytes. Model performance was evaluated with respect to the prediction accuracy, sensitivity and specificity.

5.4.4.3 Support vector machine classification

In this study, a linear kernel SVMC function was utilized in creating the model and the model performances were assessed as a function of accuracy, sensitivity and specificity. In all cases, the SVMC models were created using 67 % of the dataset (training set) and validated against 33 % of the dataset (test set). Like PCA-LDA, SVMC classification was performed on T-cells and monocytes after incubation with the genetically modified (pLKO.1, shCD73 and shZEB1) and WT GSCs.

Model performances were evaluated with respect to the prediction accuracy, sensitivity and specificity. Sensitivity is the ability of the model to detect true positives based on all

samples classified as positive whereas specificity is the ability of a model to detect true negatives based on all samples classified as negative in a class [291].

5.5 Results and discussion

To assess the ability of RS to distinguish between monocytes and T-cells after incubation with TCMs from GSCs with blocked ZEB1 or CD73, a minimum of twenty (20) cells were measured for each treatment condition. There were three human cell lines (GBM1, JHH520 and SF188) and four treatment conditions for each cell line (pLKO.1, WT, shCD73 and shZEB1). Spectra visualization was performed manually to select only spectra data with lesser noise contribution. Upon removal of poor spectral data, a total of 214, 156 and 220 spectral data was obtained for T-cell, monocyte and GSC groups, respectively. This was obtained as the sum of all spectral data collected across the treatment conditions and cell lines for a particularly cell type (say T-cell). Figure 5.1A-C shows the mean ± standard deviation of preprocessed Raman spectral data collected from T-cells, monocytes and GSC cells, respectively.

Reference spectrum obtained for PBS displayed a strong band around 3100 – 3500 cm^{-1} (O-H stretching) and a weaker band around 1640 cm^{-1} (O-H bending vibration). No visible band was observed for phosphate group (typically around 1093 cm^{-1}). The observed Raman spectra for the cells were dominated by bands at 754, 787, 938, 1003, 1093, 1258, 1303/1310, 1332, 1450, 1578, 1657, 1754, 2934, 3262 and 3416 cm^{-1}. Broadly speaking, the observed bands were consistent with Raman signatures reported for proteins, lipids and nucleic acids as shown in Table 5.1 [292-298]. The sharp band at 1003 cm^{-1} can be associated with phenylalanine ring breathing mode [285, 299] whereas other nucleic acid vibrations could be observed at 787 cm^{-1} (cytosine and uracil residues), 1093 and 1578 cm^{-1} (PO_2^- nucleic acids) [293, 296, 298]. The bands at 1258 cm^{-1} and 1303/1310 cm^{-1} are usually assigned to amide III (β-sheet) vibrations in protein and CH_2 twisting mode in lipids, respectively [292, 294, 295, 298]. The amide I band was observed at 1657 cm^{-1} [300]. Similar Raman bands were observed for the immune cells and GSCs as shown in Figure 1. The number of overlapping components in the spectral features for the three cell types limits the use of univariate analysis method; hence three

multivariate techniques including PCA, SVMC and LDA, was employed for better spectral visualisation and interpretation.

Table 5.1: Peak assignments for Raman spectra of T-cells and monocytes incubated with TCMs

Bands (cm^{-1})	Origin	Assignments	Reference
754	Nucleic acids	Tryp breathing	[297, 299]
781/787	Nucleic acids	PO_2^- symmetric stretching	[292, 294]
938	Protein	C-C backbone stretching α-helix	[292, 294]
1001/1003	Phenylalanine	Phe ring breathing	[285, 292, 294, 298, 299]
1093	DNA backbone	PO_2^- stretching	[293, 295, 301]
1258	Protein	Amide III (β-sheet)	[294, 298]
1303/1310	Lipid / protein / nucleic acids	CH_2 twisting	[292, 295, 302]
1332	Nucleic acids	CH_3 CH_2 wagging	[292, 295, 303]
1450	Lipids/proteins	CH_2 bending	[294-296]
1578	Protein / nucleic acids	Pyrimidine ring (nucleic acids) stretching and heme protein	[296, 298, 302]
1657	Peptide backbone	Amide I (α-helix), C=C stretch (lipids)	[292, 298, 304]
1754		C=O stretching	[292]
2934	Protein	CH_3 stretching	[292, 305]
3262	PBS	O-H symmetric stretching	[285]
3416	PBS	O-H antisymmetric stretching	[285]

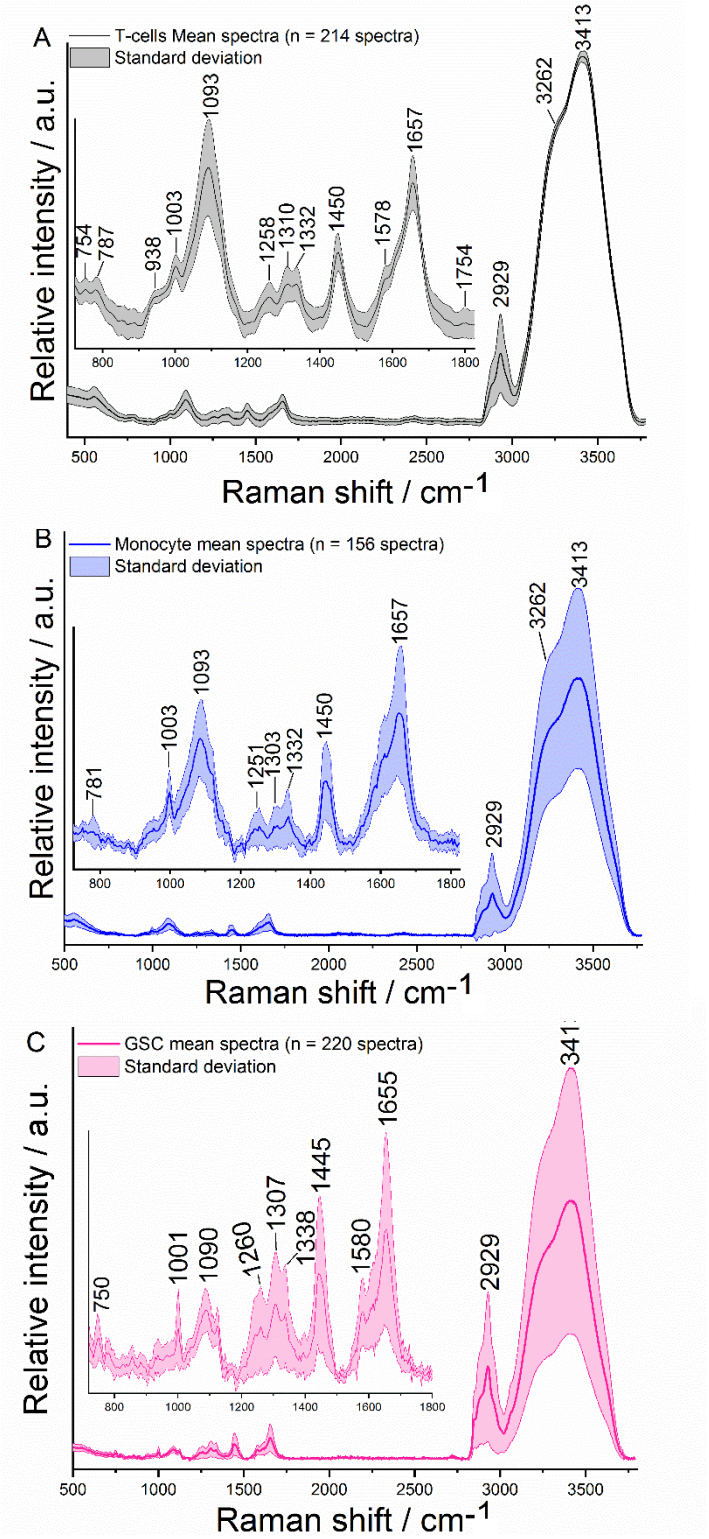

Figure 5.1: Mean Raman spectra with standard deviation of different cells after incubation with TCMs A) T-cells and B) monocytes and C) GSCs.

5.5.1 Exploratory analysis - PCA

PCA was carried out on both T-cells and monocytes incubated with TCMs of three GSC lines (GBM1, JHH520 and SF188) with their controls (WT and empty vector pLKO.1) and respective knockdowns (shZEB1 and shCD73). The PCA model was built using a random cross-validation technique on data collected from each the GSC cell lines separately. A breakdown of the groupings for the PCA is as follows: (GBM1-WT, GBM1-shCD73, GBM1-pLKO.1 and GBM1-shZEB1), (JHH520-WT, JHH520-shCD73, JHH520-pLKO.1 and JHH520-shZEB1) and (SF188-WT, SF188-shCD73, SF188-pLKO.1 and SF188-shZEB1).

In all instances, the early PCs describe variance associated with relative signal contribution from the PBS sample medium and cell samples. From herein the latter PCs, which describe variances between the cell-types and exposure conditions, are discussed.

5.5.1.1 T-cells

In the T-cells incubated with TCM GBM1, 98 % of the total variance was explained in the first four principal components (PCs). Both PC3 and PC4 show contributions relating to the separation of T-cells GBM1 group. Figure 5.2 shows the scores and loadings plot for the PCA on the T-cells. The cluster of T-cells with TCM GBM1-pLKO.1 is seen in the positive PC4 space, which describes 2 % of the variability within the dataset. T-cells incubated with TCM GBM1-WT, TCM GBM1-shZEB1 and GBM1-shCD73 clustered in the negative PC4 space. The PC4 loadings plot indicates that delineation of T-cells with TCM pLKO.1 was due to contributions from proteins (1010, 1260 and 1676 cm^{-1}) and nucleic acids (796, 956, 1108, 1134, 1322 and 1348 cm^{-1}) spectral features [292, 294, 296, 299, 306] as shown in Figure 5.2B. The negative PC4 loadings show higher lipid contribution (1294 and 1435 cm^{-1}) [292, 306] in T-cells with TCM GBM1-WT, TCM GBM1-shZEB1 and GBM-shCD73.

The separation of T-cells incubated with TCM JHH520 was seen in PC3 score, accounting for 11 % of the variability in the dataset. The cluster of T-cells incubated with

TCM JHH520-pLKO.1 and JHH520-shZEB1 was observed in the negative PC3 space whereas the T-cells with TCM JHH520-WT and JHH520-shCD73 clustered positive PC3 space (Figure 5.2C). The loadings plot showed similar trends as that of the GBM1 cell line, with separation of pLKO.1 and shZEB1 arising from proteins and lipids (1010, 1260, 1464 and 1676 cm^{-1}) and nucleic acids (800, 962, 1108, 1134, 1322 and 1348 cm^{-1}) spectral features [292, 294, 296, 299, 306] as shown in Figure 5.2D. Unlike T-cells with TCM GBM1 and JHH520, the PCA of T-cells with TCM SF188 showed separation along the PC2 axis, which describes 20 % of the explained variance within the dataset. T-cells with TCM SF188-pLKO.1 clustered in the positive PC2 space. Examination of the loadings plot show that major contributions from proteins and lipids (1450 and 1657 cm^{-1}) and nucleic acids (1089 cm^{-1}) spectral features were responsible for the delineation of the T-cells with TCM SF188-pLKO.1.

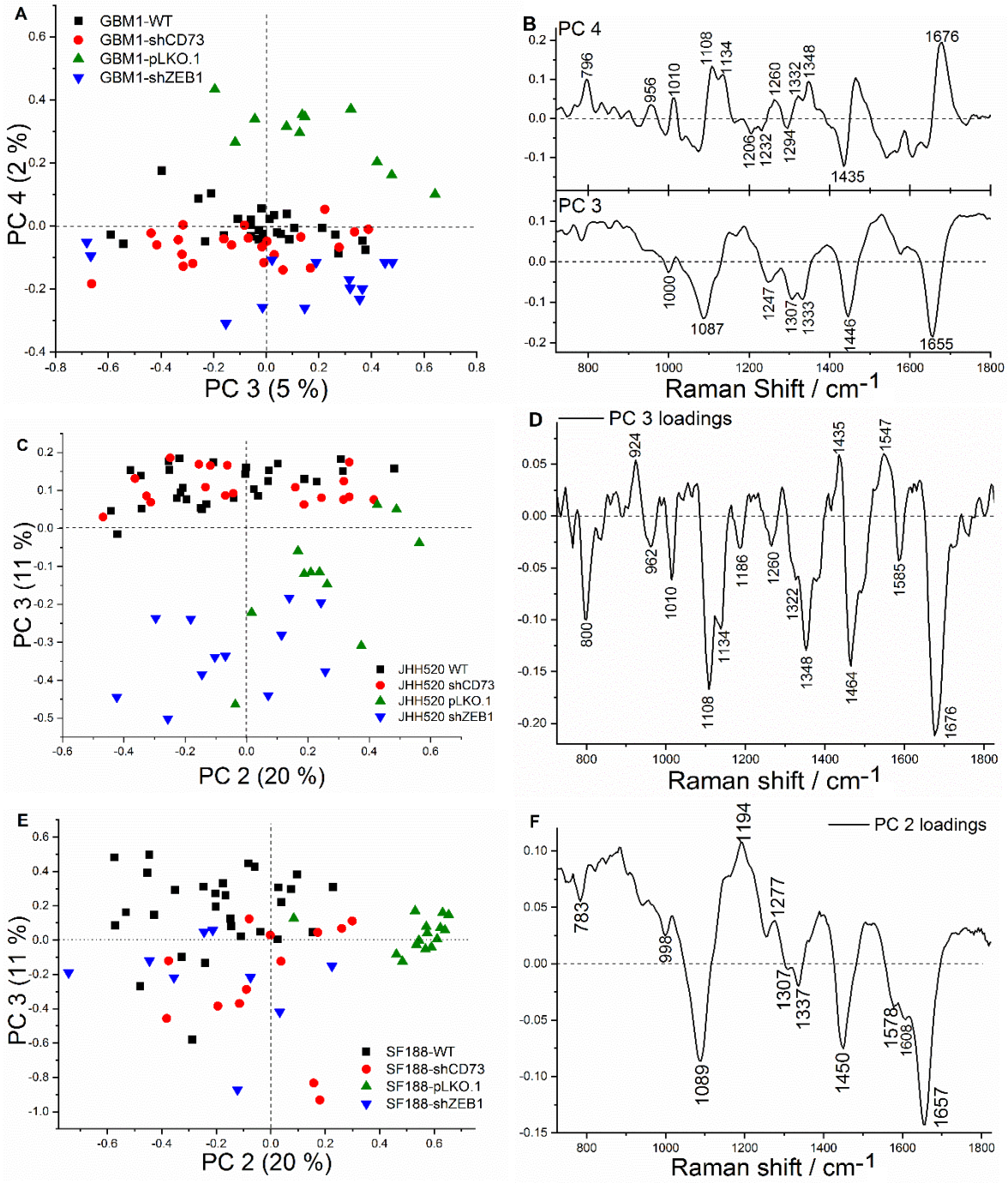

Figure 5.2: PCA scores and loadings plot for T-cells analysis after incubation with TCMs of three GSC lines with their controls and knockdowns (WT, pLKO.1, shCD73 and shZEB1) A) GBM1 scores plot B) GBM1 loadings plot C) JHH520 scores plot D) JHH520 loadings plot E) SF188 scores plot and F) SF188 loadings plot.

5.5.1.2 Monocytes

Figure 5.3 shows the scores and loadings for the PCA of monocytes. Data from the pLKO.1 knockdown were excluded from the analysis because they exhibited a very poor signal to noise ratio, which negatively affected the analysis. This was associated to poor extraction during the sample preparation, as such the spectral quality could not be improved using preprocessing. In Figure 5.3C, discrimination of monocytes was along the PC3 axis describing 3 % of the variance in the dataset. The majority of monocytes incubated with TCM JHH520-shZEB1 clustered in negative PC3 space whereas TCM JHH520-shCD73 and JHH520-WT clustered in positive PC3 space. The loading plot suggests major contribution of proteins and lipids (1122, 1251, 1468 and 1675 cm^{-1}) [130, 294, 306] in monocytes incubated with TCM JHH520-shZEB1 whereas monocytes incubated with TCM JHH520-shCD73 and JHH520-WT had more contributions from nucleic acids (806, 942 and 1175 cm^{-1}) [294, 306]. Discrimination in GBM1 cell line was observed along PC2 axis describing 27 % of the variance in the dataset. Clusters of monocytes treated with TCM GBM1-shZEB1 was seen in the PC2 positive space whereas majority of monocytes incubated with TCM GBM1-shCD73 and GBM1-WT clustered in the negative PC2 space. The loading features showed that samples clustering in the positive PC2 space (TCM GBM1-shZEB1) had higher relative contributions from bands at 1450 and 1657 cm^{-1}, which are consistent with lipids and protein content, respectively [130, 292, 294, 306]. In monocytes with SF188 cell line, spectral data from TCM SF188 shZEB1 and SF188 WT separated along PC3 axis describing 4 % of the variability in the sample set. Samples from TCM SF188 shCD73 were seen in the positive PC3 space. The overlap in the scores plot contributes to difficulty in identifying the chemical constituents responsible for the separation of the cell groups, hence the use of other classification technique such as PCA-LDA and SVMC for better interpretation.

The observed clusters in the above PCA models could be explained by the differences in cytokine profile of the immune cells indicating the presence of the mixed populations including pro-inflammatory and anti-inflammatory MΦs and DCs as well as $CD4^+$ and $CD8^+$ T-cells. The results indicate that GSCs with ZEB1 and CD73 inhibition can actively

influence the phenotype of T-cells and monocytes and these differences in cell state can be visualised by Raman spectral analysis.

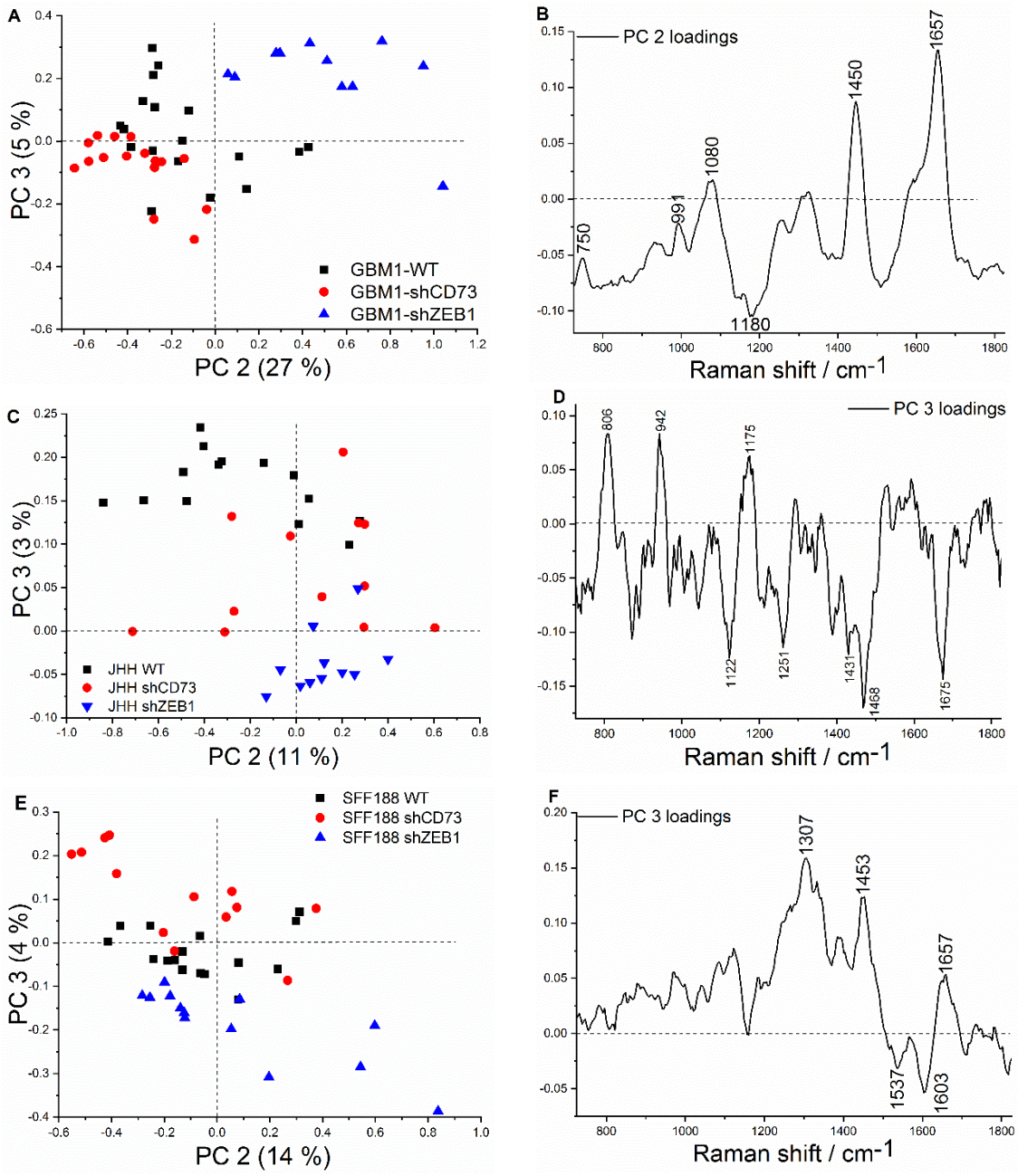

Figure 5.3: PCA scores and loadings plot for monocytes analysis after incubation with TCMs of three GSC lines with their controls and knockdowns (WT, shCD73 and shZEB1) A) GBM1 scores plot B) GBM1 loadings plot C) JHH520 scores plot D) JHH520 loadings plot E) SF188 scores plot and F) SF188 loadings plot.

5.5.1.3 Glioma cells

PCA scores plot of the glioma cells, do not show distinct separation for all three GSC lines considered (Figure 4), which might be suggestive of similarity in the largest source of variance in the cell lines and their respective knockdowns. The observable spectral feature for the GSC (GBM1 and JHH520) is seen at 1090/1098 and 1445 cm^{-1}, which is indicative of the lipids and nucleic acid contribution in the cells samples.

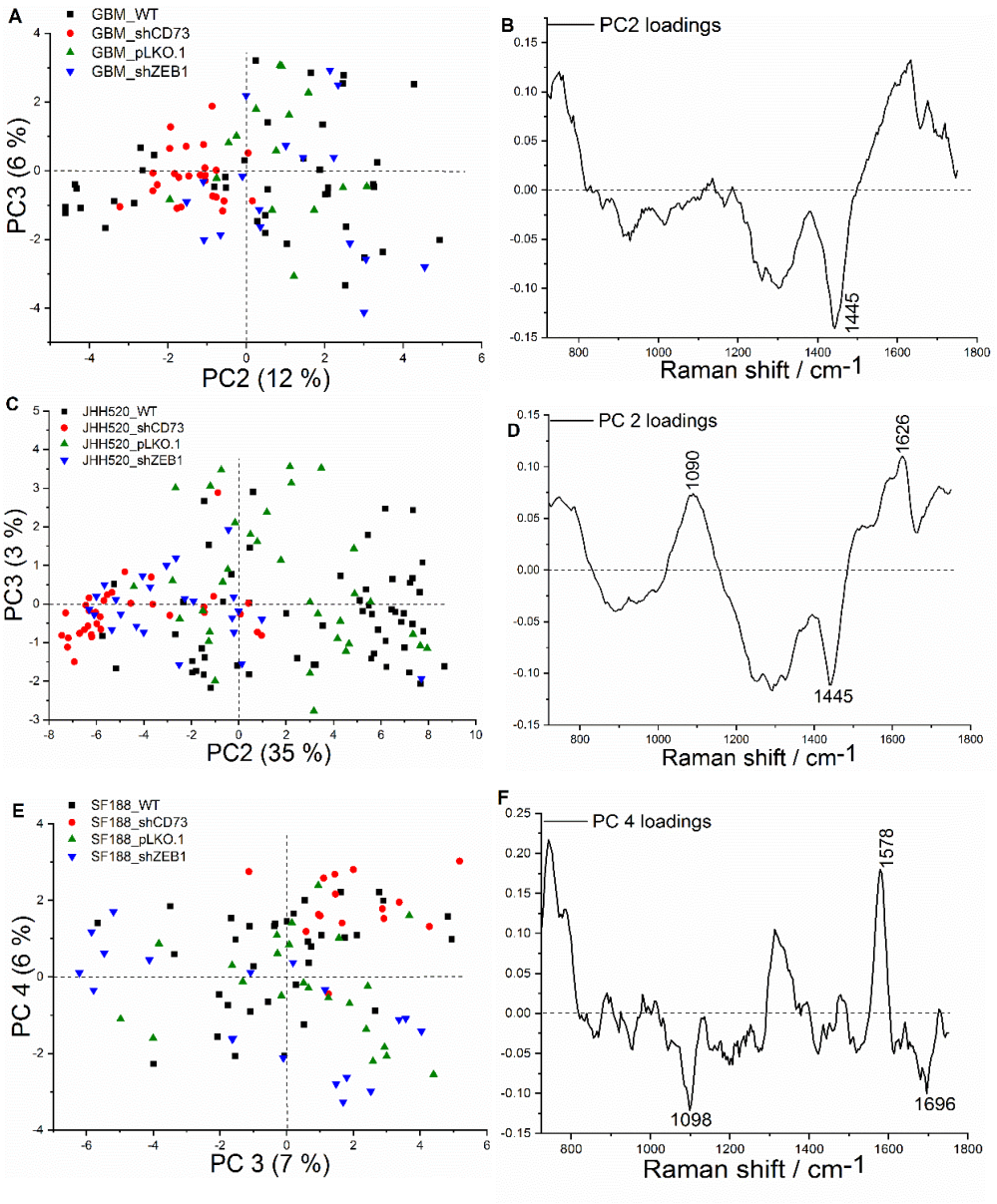

Figure 5.4: PCA scores for GSCs with their controls and knockdowns (WT, pLKO.1, shCD73 and shZEB1) A) GBM1 B) JHH520 and C) SF188 scores plot.

5.5.2 Discrimination of T-cells/monocytes after incubation with TCMs from GSCs with respective knockdowns

PCA-LDA and SVMC have successfully been applied to Raman data for the classification of cells [303, 307]. Unlike PCA, PCA-LDA and SVMC techniques are supervised classification techniques which require assignment of the dataset into predefined groups prior to model training. For T-cells incubated with TCM from the GBM1 cell line, the classes were defined as: class 1 (GBM1-WT), class 2 (GBM1-shCD73), class 3 (GBM1-pLKO.1) and class 4 (GBM1-shZEB1). Similar groupings were adopted for JHH50 and SF188 cell line. However, only three classes were required for monocytes for the three cell lines owing to the exclusion of dataset from the pLKO.1 knockdowns. These data had very poor signal: noise ratios, which negatively affected the analysis.

Building of the PCA-LDA model utilised the use of the first four PCs from the PCA. Model performances are presented in terms of accuracy, sensitivity and specificity in Table 3. The PCA-LDA model yielded a sensitivity and specificity between 78 - 100 % and 85 - 100 %, respectively for the T-cells after incubation with TCM of all three GSCs lines. Validation of this model against an independent test set yielded a maximum sensitivity of 100 %. The PCA-LDA model worked best for the T-cells treated with TCM of GBM1 with both sensitivity and specificity of 100 % for all cell groups considered upon validation against the test set. Improved sensitivity and specificity could be achieved using the SVMC model on the training set, with validation on an independent set yielding similar results as the PCA-LDA models. The SVMC models yielded a 100 % accuracy on the training set for all cell lines considered whereas cell group sensitivities ranged between 67 - 100 %. Specificity of over 78 % was achieved for individual cell groups. The T-cells with TCM SF188-shZEB1 group had the most misclassification with a sensitivity of ~67 % for both SVMC and PCA-LDA model. This is not surprising, as they had the least clear separation of the four groups in the PCA (Figure 2E).

Discriminatory analysis performed using the monocytes also highlighted the model's ability to classify cell groups. Both PCA-LDA and SVMC models yielded a classification

sensitivity over 67 % on validating the model against an independent test set, with specificity over 73 %. The SVMC models tend to have better performance for both T-cells and monocytes.

For GSC classification, the PCA-LDA model yielded sensitivity and specificity of between 40 - 78 % and 56 - 67 %, respectively on validation against an independent test set (Table 3). The PCA-LDA had highest sensitivity and specificity for the JHH520 cell line. Improved sensitivity and specificity was observed on the SVMC model using the training set. Validation against the test set yielded sensitivity and specificity of 33 - 88 % and 53 - 70 %, respectively. On comparison to the immune cells, the glioma cells had the worse classification performance. The PCA of the glioma cells showed the major overlap across all cell-line considered, as such not surprising that the classification of the spectral data did not work so well. Despite this, the discrimination of the cell lines and their knockdown treatments were not random with sensitivities over 25 % (4 classes) for all cell groups considered.

Overall, it is difficult to make comparison to the established literature owing to the differences in treatment conditions of the immune cells, however, the performance of our analyses were similar to the work exploring the discrimination of different immune cells with sensitivities in the range of 86-100% achieved [254]. A more closely related study using RS and PCA-LDA, was successfully employed in the discrimination of active and naïve T-cells, with an increased cytochrome c levels observed for the activated T-cells [308].

The combination of Raman and multivariate analysis techniques yielded good classification result for both the PCA-LDA and SVMC models in discriminating differences of monocytes and T-cells after incubation with TCMs. These differences could be present due to the involvement of ZEB1 and CD73 in monocytes and T-cells activation and differentiation. This can offer accurate and fast alternative for non-invasive screening in clinical diagnosis.

Results from the flow cytometry analysis were in agreement with the Raman spectroscopic analysis [37]. However, these results have not been presented in this thesis as they were outside the scope of the thesis.

Table 5.2: Summary of model performance illustrating the efficiency of RS to discriminate monocytes / T-cells for the training (in black) and test sets (in red).

Cell line	Accuracy (%)	Sensitivity (%)				Specificity (%)			
		WT	shCD73	pLKO.1	shZEB1	WT	shCD73	pLKO.1	shZEB1
T-cells PCA-LDA									
GBM1 WT v shCD73 v pLKO.1 v shZEB1	90.4 (100)	100 (100)	87.5 (100)	100 (100)	77.9 (100)	85.3 (100)	94.3 (100)	88.6 (100)	93.0 (100)
JHH520 WT v shCD73 v pLKO.1 v shZEB1	95.7 (91.7)	100 (100)	83.3 (71.4)	100 (100)	100 (100)	92.9 (85.7)	100 (100)	95.0 (90.0)	94.7 (90.5)
SF188 WT v shCD73 v pLKO.1 v shZEB1	88.6 (79.2)	78.9 (70.0)	88.9 (100)	100 (80.0)	100 (66.7)	96.0 (85.7)	88.6 (72.2)	85.3 (78.9)	86.8 (90.0)
SVMC									
GBM1 WT v shCD73 v pLKO.1 v shZEB1	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
JHH520 WT v shCD73 v pLKO.1 v shZEB1	100 (87.5)	100 (100)	100 (71.4)	100 (75.0)	100 (100)	100 (78.6)	100 (94.1)	100 (90)	100 (85.7)
SF188 WT v shCD73 v pLKO.1 v shZEB1	100 (91.7)	100 (100)	100 (100)	100 (80.0)	100 (66.7)	100 (85.7)	100 (88.9)	100 (94.7)	100 (95.7)
Monocytes PCA-LDA									
GBM1 WT v shCD73 v pLKO.1 v shZEB1	84.9 (87.5)	100 (100)	66.7 (71.4)	-	100 (100)	77.3 (81.8)	100 (100)	-	80.8 (87.5)
JHH520 WT v shCD73 v pLKO.1 v shZEB1	73.2 (85.7)	88.9 (100)	72.7 (100)	-	77.8 (100)	68.8 (93.8)	73.3 (93.3)	-	71.9 (92.9)
SF188 WT v shCD73 v pLKO.1 v shZEB1	86.7 (80)	76.9 (66.7)	100 (80)	-	88.9 (100)	94.1 (88.9)	81.8 (80)	-	85.7 (72.7)
SVMC									
GBM1 WT v shCD73 v pLKO.1 v shZEB1	100 (93.8)	100 (100)	100 (100)	-	100 (75)	100 (100)	100 (100)	-	100 (100)
JHH520 WT v shCD73 v pLKO.1 v shZEB1	100 (95.2)	100 (100)	100 (100)	-	100 (50)	100 (90.5)	100 (86.7)	-	100 (86.7)
SF188 WT v shCD73 v pLKO.1 v shZEB1	100 (88.6)	100 (83.3)	100 (80)	-	100 (100)	100 (87.5)	100 (88.9)	-	100 (81.8)

Table 5.3: Summary of model performance illustrating the efficiency of RS to discriminate GSCs for the training (in black) and test sets (in red)

		GSCs PCA-LDA							
Cell line	Accuracy	Sensitivity				Specificity			
		WT	CD73	pLKO.1	shZEB1	WT	CD73	pLKO.1	shZEB1
GBM1 WT vs CD73 vs pLKO.1 vs shZEB1	63.4 (61)	40 (56.3)	100 (75)	91.7 (66.7)	41.6 (50)	80.5 (65)	51.9 (57)	57.6 (60)	67.8 (63.3)
JHH520 WT vs CD73 vs pLKO.1 vs shZEB1	59.4 (73.5)	56.8 (77.8)	69.6 (63.6)	56 (53.8)	56.3 (50)	60.9 (56.3)	56.4 (64.1)	60.5 (67.6)	60 (66.7)
SF188 WT vs CD73 vs pLKO.1 vs shZEB1	69.2 (61.8)	61.5 (58.3)	80 (40)	62.5 (66.7)	84.6 (62.5)	74.4 (59)	67.3 (62)	71.4 (56)	65.4 (57.7)
		SVMC							
GBM1 WT vs CD73 vs pLKO.1 vs shZEB1	100 (64)	100 (62.5)	100 (87.5)	100 (66.7)	100 (33.3)	100 (65)	100 (57)	100 (63)	100 (70)
JHH520 WT vs CD73 vs pLKO.1 vs shZEB1	98 (62)	100 (77.8)	100 (45.5)	96 (46.2)	93.8 (75)	96.9 (53)	97.4 (66.7)	98.7 (67.6)	98.8 (59.5)
SF188 WT vs CD73 vs pLKO.1 vs shZEB1	96.9 (61.8)	96.2 (58.3)	100 (40)	93.8 (77.8)	100 (62.5)	97.4 (68.2)	96.4 (65.5)	98 (56)	96.2 (61.5)

5.6 Conclusion

This study shows the ability of RS in combination with chemometrics to discriminate T-cells and monocytes upon incubation with TCM of GSCs in different conditions, therefore, to detect glioma associated neuroinflammation caused by molecular differences. A previous study has shown the ability of RS to detect even minor changes such as variations in keratin expression levels in cancer cells [309]. Using both the PCA-LDA and SVMC techniques the different phenotype of the immune cells could be

discriminated and the results were in agreement with the flow cytometry analysis of the same samples [37]. The flow cytometry analysis suggests that the GSCs with ZEB1 and CD73 inhibition can actively influence the phenotype of T-cells and monocytes driving their differentiation into a population of mixed pro-inflammatory and anti-inflammatory MΦs and DCs.

Chapter 6

Conclusions and Future Work

The turning point in the development of any vibrational spectroscopic method for routine analysis is in the art of building a robust and accurate chemometric model to predict the sample properties. Prior to the development of a prediction model, several key factors need to be considered as they directly influence on the overall performance quality and applicability. Specification of the responses alongside the range of variation should be the first step; this would determine the choice of instrument, configuration and sampling method. Subsequent steps would include selecting a representative calibration set covering the expected variation range, collection of spectral data, selection of appropriate preprocessing techniques based on the quality of spectral data, variable selection based on the chemical properties to be modelled and development of the chemometric model that is tested using cross-validation and an independent test set. Each of the identified steps play a crucial role in the model performance, however, a well-selected calibration set is key to a well performing model, as it determines the spectral data that is further utilised in building the model.

Selection of the calibration set should incorporate the expected variability of future prediction sets. This is not limited to the prediction response (like concentration range) but also physical and environmental variability. Several strategies have been proposed for the capture of variability in the calibration set including using a pilot-plant samples reproducing full scale measurement conditions, however, this become problematic on increasing number of responses. Another strategy would be preparation of laboratory samples in which concentration ranges of the intended responses are varied simultaneously within ranges to avoid correlation. The calibration set development

strategy for quantitative analysis like the ones applied in this thesis depends on factors like sample complexity (number of interferences and responses) and intended application (off-line or real-time monitoring) [310]. Assuming the expected variability changes in the future, the calibration model would need to be updated to expand the variability and scope of the method.

This thesis involved the application of different chemometric modelling techniques to vibrational spectroscopic data from samples like meat, pharmaceuticals and immune cells. The nature of information probed determined the choice of chemometric techniques; however, some techniques might give better approximation (performance) for one sample than others. For example, PCA-LDA gave better performance for the cell study than PLSDA.

Prior to model development, selection of the appropriate calibration range for the intended responses is crucial. For the meat study, this was challenging as quality attributes (pH and %IMF) in single red meat type of similar breed, sex, age will have a narrow variation. However, this was resolved by including samples across three different meat types, increasing the variability as well as the robustness of the model. Spectral preprocessing techniques was selected based on the signal quality, for example, smoothing and rubberband baseline correction was applied to Raman data whereas linear baseline correction and no smoothing was performed on the IR data. Issues of subsampling associated with the IR configuration was reduced by incorporating a raster system in the Raman method. Improvement in spectral quality was performed by collecting co-added spectra that were averaged. Predictive models were developed using PLS (quantitative and classification) and SVM (classification), which are linear and non-linear algorithms, respectively. This was to allow for modelling of linear and non-linear relationship between the spectral data and reference measurements. For the meat classification studies in chapter two, the non-linear method yielded better predictive performance on validation against the independent test sets. Future work in this field should involve the development of a robust calibration set, creation of the model and deployment of live prediction model alongside vibrational spectroscopy (NIR and Raman) for routine meat assessment. This can be incorporated in-line/at-line in meat abattoirs, alongside automatic large area

sampling to allow for collection of representative data. Capturing the wide variability in meat including animal types, breeds, age, sex, rearing condition, sample presentation and environmental conditions would be important to build a stable prediction model. Creation of spectral library using different meat types can facilitate characterisation of the meat species during routine monitoring of meat at supermarket and other outlets. Future chemometric modelling should incorporate non-linear algorithms like artificial neural network, local weighted regression in place of partial least square regression to account for non-linear relationship between the spectral data and responses.

Classification models developed using PCA-LDA and SVM were successful in discriminating immune cells incubated with different tumour condition media. The choice of PCA-LDA and SVM was to allow comparison between the performance of linear and non-linear algorithms. Phenotype of T-cells and monocytes were incubated with media conditioned by different glioblastoma stem-cells (GSCs) showing different molecular background. Multivariate analysis performed using PCA-LDA and SVMC yielded sensitivities and specificities $\geq$ 70 % and $\geq$ 67 %, respectively. The results were in agreement with findings obtained using flow cytometry analysis. Future work in this field should entail increase in the number of participants and possible imaging of the cell samples. It would be interesting to see if more information could be obtained as compared to spot measurements.

The choice of multivariate curve resolution for the dehydration studies was to allow the identification of original spectra as well as their quantification. The dehydration process was monitored at different temperatures in both the low- and mid-frequency Raman regions. The low-frequency Raman region probes the intermolecular interactions whereas the mid-frequency Raman region probes the intramolecular interaction in the crystals. DFT calculation provided insight into the vibrational modes particularly in the low frequency region. Kinetic analysis performed using multivariate curve resolution and principal component analysis revealed the formation of piroxicam anhydrous form I and theophylline anhydrous form II upon dehydration of their respective monohydrates. Detection of solid-state forms and their transformation is important from a drug performance and patent point of view. Low frequency Raman spectroscopy was very

sensitive in identifying the changes on transformation of the monohydrates across different timescales observed in the low and mid-frequency regions.

Improvement in the performance of individual models can be achieved by fusing data from multiple instruments. This allows for modelling of greater information, for example, amide II band was not observed in the Raman spectra but was observed in the IR spectra, thus combining both data will lead to greater information. However, the use of data from multiple instruments could lead problems of redundancy. This can be resolved by preprocessing the spectra, use of less number of instruments and adoption of mid- or high-level fusion strategies.

To develop this work further, the application of vibrational spectroscopy for in-line monitoring should be trialled in the industry, especially for meat. This would entail the use of fiber optic probes connected at the interface of the different stages of the process to acquire spectral data at defined intervals. The response values obtained through the reference methods are correlated with the spectral data, considering the process time as a factor. Impact of sample presentation, changing process and interferences can be monitored to challenge the robustness of the methods. For meat assessment, fibre optic coupled Raman or NIR can be deployed at different stages in the process for monitoring of quality parameters such as %IMF, moisture content, pH and shear force, whereas handheld device can be utilised for routine analysis. For inline analysis in meat industry, automated large area scanning would facilitate acquisition of representative data, while chemometric models can be developed using non-linear algorithms on a wide range of meat samples. Live prediction models can be held in cloud storage to ensure that instrument failure does not lead to loss. It would be interesting to compare the performance of the inline prediction to laboratory studies.

Bibliography

1. Yadav, L.D.S., *Introduction to Spectroscopy (Spectrometry)*, in *Organic Spectroscopy*, L.D.S. Yadav, Editor. 2005, Springer Netherlands: Dordrecht. p. 1-6.
2. Skoog, D.A., F.J. Holler, and S.R. Crouch, *Principles of instrumental analysis*. 2017: Cengage learning.
3. Arnold, M.A., *IR Spectroscopy: An Introduction Helmut Gunzler and Hans-Ulrich. Weinheim, Germany: Wiley, 2002, 376 pp., $45.00. ISBN 3-527-28896-1*. Clinical Chemistry, 2003. **49**(8): p. 1423-1423.
4. Rayleigh, L., *X. On the electromagnetic theory of light.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1881. **12**(73): p. 81-101.
5. Raman, C.V. and K.S. Krishnan, *A New Type of Secondary Radiation.* Nature, 1928. **121**(3048): p. 501-502.
6. Bumbrah, G.S. and R.M. Sharma, *Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse.* Egyptian Journal of Forensic Sciences, 2016. **6**(3): p. 209-215.
7. Fraser, S.J., *Raman spectroscopy for analytical applications* in *Chemistry*. 2014, University of Otago.
8. Busca, G. and C. Resini, *Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials*, in *Encyclopedia of Analytical Chemistry*. 2006.
9. Banwell, C.N., *Fundamentals of molecular spectroscopy [by] C. N. Banwell*. 1972, London, New York: McGraw-Hill.
10. Ferraro, J.R., K. Nakamoto, and C.W. Brown, *Chapter 1 - Basic Theory*, in *Introductory Raman Spectroscopy (Second Edition)*, J.R. Ferraro, K. Nakamoto, and C.W. Brown, Editors. 2003, Academic Press: San Diego. p. 1-94.
11. Moser, C. and F. Havermeyer, *Ultra-narrow-band tunable laserline notch filter*. Applied Physics B, 2009. **95**(3): p. 597-601.
12. Bērziņš, K., S.J. Fraser-Miller, and K.C. Gordon, *Recent advances in low-frequency Raman spectroscopy for pharmaceutical applications.* International Journal of Pharmaceutics, 2021. **592**: p. 120034.
13. Lipiäinen, T., et al., *Direct comparison of low- and mid-frequency Raman spectroscopy for quantitative solid-state pharmaceutical analysis.* Journal of Pharmaceutical and Biomedical Analysis, 2018. **149**: p. 343-350.
14. Herzberg, G., *Molecular Spectra and molecular structure-Vol I*. Vol. 1. 2013: Read Books Ltd.
15. Polavarapu, P.L., *Ab initio vibrational Raman and Raman optical activity spectra.* The Journal of Physical Chemistry, 1990. **94**(21): p. 8106-8112.
16. Settle, F., *Handbook of instrumental techniques for analytical chemistry. 1997*. National Science Foundation, Arlington, 1997.
17. Richard, L.M., *Introduction and Scope*, in *Raman Spectroscopy for Chemical Analysis*. 2000. p. 1-14.
18. Everall, N.J., *Confocal Raman Microscopy: Why the Depth Resolution and Spatial Accuracy Can Be Much Worse Than You Think.* Applied Spectroscopy, 2000. **54**(10): p. 1515-1520.

19. Abramowitz, M., et al., *Basic Principles of Microscope Objectives*. BioTechniques, 2002. **33**(4): p. 772-781.
20. Bhargava, R., T. Ribar, and J.L. Koenig, *Towards Faster FT-IR Imaging by Reducing Noise.* Applied Spectroscopy, 1999. **53**(11): p. 1313-1322.
21. Einax, J.W., *M.J. Adams: Chemometrics in Analytical Spectroscopy, 2nd edition.* Analytical and Bioanalytical Chemistry, 2005. **382**(4): p. 861-862.
22. Richard, L.M., *Signal-to-Noise in Raman Spectroscopy*, in *Raman Spectroscopy for Chemical Analysis*. 2000. p. 49-71.
23. Gautam, R., et al., *Review of multidimensional data processing approaches for Raman and infrared spectroscopy*. EPJ Techniques and Instrumentation, 2015. **2**(1): p. 8.
24. Gallagher, N., *Savitzky-Golay Smoothing and Differentiation Filter*. 2020.
25. Press, W.H. and S.A. Teukolsky, *Savitzky-Golay Smoothing Filters.* Computers in Physics, 1990. **4**(6): p. 669-672.
26. Qian, F., Y. Wu, and P. Hao, *A fully automated algorithm of baseline correction based on wavelet feature points and segment interpolation.* Optics & Laser Technology, 2017. **96**: p. 202-207.
27. Camo, S.A. *The Unscrambler® Appendices: Method References*. Available from: https://www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf.
28. Beebe, K.R., et al., *Chemometrics: A Practical Guide*. 1998: Wiley.
29. Lieber, C.A. and A. Mahadevan-Jansen, *Automated method for subtraction of fluorescence from biological Raman spectra.* Appl Spectrosc, 2003. **57**(11): p. 1363-7.
30. Gan, F., G. Ruan, and J. Mo, *Baseline Correction by Improved Iterative Polynomial Fitting with Automatic Threshold.* Chemometrics and Intelligent Laboratory Systems, 2006. **82**: p. 59-65.
31. Pirzer;, M. and J. Sawatzki;, *Method and device for correcting a spectrum*. 2008, Bruker Optik GmbH, Ettlingen (DE) United States.
32. Randolph, T., *Scale-based normalization of spectral data.* Cancer biomarkers : section A of Disease markers, 2006. **2**: p. 135-44.
33. Zeaiter, M. and D. Rutledge, *3.04 - Preprocessing Methods*, in *Comprehensive Chemometrics*, S.D. Brown, R. Tauler, and B. Walczak, Editors. 2009, Elsevier: Oxford. p. 121-231.
34. Wold, S., *Chemometrics; what do we mean with it, and what do we want from it?* Chemometrics and Intelligent Laboratory Systems, 1995. **30**(1): p. 109-115.
35. Djuris, J., S. Ibric, and Z. Djuric, *4 - Chemometric methods application in pharmaceutical products and processes analysis and control*, in *Computer-Aided Applications in Pharmaceutical Technology*, J. Djuris, Editor. 2013, Woodhead Publishing. p. 57-90.
36. Al-Jowder, O., et al., *Mid-Infrared Spectroscopy and Chemometrics for the Authentication of Meat Products.* Journal of Agricultural and Food Chemistry, 1999. **47**(8): p. 3210-3218.
37. Robert, C., et al., *Molecular monitoring of glioblastoma's immunogenicity using a combination of Raman spectroscopy and chemometrics.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021: p. 119534.
38. Zhihua, Q., Z. Lan, and J. Huang, *Effective Linear Discriminant Analysis for High Dimensional, Low Sample Size Data.* Lecture Notes in Engineering and Computer Science, 2008. **2**.
39. Fredericks, P.M., et al., *Materials Characterization Using Factor Analysis of FT-IR Spectra. Part 1: Results.* Applied Spectroscopy, 1985. **39**(2): p. 303-310.
40. Li, H., et al. *Cluster analysis method and Near-infrared spectroscopy applied to the identification of food*. in *2010 International Conference on Machine Learning and Cybernetics*. 2010.

41. Boyaci, I.H., et al., *A novel method for discrimination of beef and horsemeat using Raman spectroscopy.* Food Chem, 2014. **148**: p. 37-41.
42. Brereton, R.G. and G.R. Lloyd, *Partial least squares discriminant analysis: taking the magic away.* Journal of Chemometrics, 2014. **28**(4): p. 213-225.
43. Kogermann, K., et al., *Qualitative in situ analysis of multiple solid-state forms using spectroscopy and partial least squares discriminant modeling.* J Pharm Sci, 2007. **96**(7): p. 1802-20.
44. Li, Q., et al., *An improved k-nearest neighbour method to diagnose breast cancer.* Analyst, 2018. **143**(12): p. 2807-2811.
45. Yu, P., *Applications of Hierarchical Cluster Analysis (CLA) and Principal Component Analysis (PCA) in Feed Structure and Feed Molecular Chemistry Research, Using Synchrotron-Based Fourier Transform Infrared (FTIR) Microspectroscopy.* Journal of Agricultural and Food Chemistry, 2005. **53**(18): p. 7115-7127.
46. Awad, M. and R. Khanna, *Support Vector Machines for Classification*, in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Editors. 2015, Apress: Berkeley, CA. p. 39-66.
47. Næs, T. and H. Martens, *Principal component regression in NIR analysis: Viewpoints, background details and selection of components.* Journal of Chemometrics, 1988. **2**(2): p. 155-167.
48. Boulesteix, A.L. and K. Strimmer, *Partial least squares: a versatile tool for the analysis of high-dimensional genomic data.* Brief Bioinform, 2007. **8**(1): p. 32-44.
49. Awad, M. and R. Khanna, *Support Vector Regression*, in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Editors. 2015, Apress: Berkeley, CA. p. 67-80.
50. De Luca, M., et al., *A single MCR-ALS model for drug analysis in different formulations: Application on diazepam commercial preparations.* Journal of Pharmaceutical and Biomedical Analysis, 2017. **134**: p. 346-351.
51. De Juan, A., J. Jaumot, and R. Tauler, *Multivariate Curve Resolution (MCR). Solving the mixture analysis problem.* Analytical Methods, 2014. **6**(14): p. 4964-4976.
52. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016. **374**(2065): p. 20150202.
53. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and Intelligent Laboratory Systems, 1987. **2**(1): p. 37-52.
54. Mardia, K., J. Kent, and J. Bibby, *Multivariate Analysis, 1979.* Probability and mathematical statistics. Academic Press Inc.
55. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine Learning, 1995. **20**(3): p. 273-297.
56. Eriksson, L., et al., *Multi- and Megavariate Data Analysis. Part I Basic Principles and Applications. Second revised and enlarged edition.* Ume Sweden: MKS Umetrics AB, 2006: p. 1-103.
57. Hanrahan, G., F. Udeh, and D.G. Patil, *CHEMOMETRICS AND STATISTICS | Multivariate Calibration Techniques*, in *Encyclopedia of Analytical Science (Second Edition)*, P. Worsfold, A. Townshend, and C. Poole, Editors. 2005, Elsevier: Oxford. p. 27-32.
58. Geladi, P. and E. Dåbakk, *Computational Methods and Chemometrics in Near Infrared Spectroscopy*, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, J.C. Lindon, G.E. Tranter, and D.W. Koppenaal, Editors. 2017, Academic Press: Oxford. p. 350-355.

59. Trygg, J. and T. Lundstedt, *Chapter 6 - Chemometrics Techniques for Metabonomics*, in *The Handbook of Metabonomics and Metabolomics*, J.C. Lindon, J.K. Nicholson, and E. Holmes, Editors. 2007, Elsevier Science B.V.: Amsterdam. p. 171-199.
60. Parastar, H., *Chapter 6 - Multivariate Curve Resolution Methods for Qualitative and Quantitative Analysis in Analytical Chemistry*, in *Data Handling in Science and Technology*, A.M. de la Peña, et al., Editors. 2015, Elsevier. p. 293-345.
61. Liu, Z., W. Cai, and X. Shao, *Outlier detection in near-infrared spectroscopic analysis by using Monte Carlo cross-validation.* Science in China Series B: Chemistry, 2008. **51**(8): p. 751.
62. Li, Z.-F., et al., *Outlier Detection for Multivariate Calibration in Near Infrared Spectroscopic Analysis by Model Diagnostics.* Chinese Journal of Analytical Chemistry, 2016. **44**(2): p. 305-309.
63. Yang, L., et al., *Rapid Identification of Pork Adulterated in the Beef and Mutton by Infrared Spectroscopy.* Journal of Spectroscopy, 2018. **2018**: p. 2413874.
64. Wang, W., et al., *Spectral Detection Techniques for Non-Destructively Monitoring the Quality, Safety, and Classification of Fresh Red Meat.* Food Analytical Methods, 2018. **11**(10): p. 2707-2730.
65. Wilson, R.H. and H.S. Tapp, *Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods.* TrAC Trends in Analytical Chemistry, 1999. **18**(2): p. 85-93.
66. Su, W.-H. and D.-W. Sun, *Mid-infrared (MIR) Spectroscopy for Quality Analysis of Liquid Foods.* Food Engineering Reviews, 2019. **11**(3): p. 142-158.
67. McGoverin, C.M., et al., *Near Infrared and Mid-Infrared Spectroscopy for the Quantification of Adulterants in Ground Black Pepper.* Journal of Near Infrared Spectroscopy, 2012. **20**(5): p. 521-528.
68. Heise, H.M. *Medical Applications of Infrared Spectroscopy*. in *Progress in Fourier Transform Spectroscopy*. 1997. Vienna: Springer Vienna.
69. De Bruyne, S., M.M. Speeckaert, and J.R. Delanghe, *Applications of mid-infrared spectroscopy in the clinical laboratory setting.* Critical Reviews in Clinical Laboratory Sciences, 2018. **55**(1): p. 1-20.
70. Franck, P., P. Nabet, and B. Dousset, *Applications of infrared spectroscopy to medical biology.* Cell Mol Biol (Noisy-le-grand), 1998. **44**(2): p. 273-5.
71. Kalinkova, G.N., *Infrared spectroscopy in pharmacy.* Vibrational Spectroscopy, 1999. **19**(2): p. 307-320.
72. Song, Y., et al., *Applications of Fourier transform infrared spectroscopy to pharmaceutical preparations.* Expert Opin Drug Deliv, 2020. **17**(4): p. 551-571.
73. Robert, C., et al., *Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy.* Food Chemistry, 2021. **343**: p. 128441.
74. Herrero, A.M., *Raman spectroscopy a promising technique for quality assessment of meat and fish: A review.* Food Chemistry, 2008. **107**(4): p. 1642-1651.
75. Li, Y.-S. and J.S. Church, *Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials.* Journal of Food and Drug Analysis, 2014. **22**(1): p. 29-48.
76. Yang, D. and Y. Ying, *Applications of Raman Spectroscopy in Agricultural Products and Food Analysis: A Review.* Applied Spectroscopy Reviews, 2011. **46**(7): p. 539-560.
77. Scarmo, S., et al., *Skin carotenoid status measured by resonance Raman spectroscopy as a biomarker of fruit and vegetable intake in preschool children.* Eur J Clin Nutr, 2012. **66**(5): p. 555-60.

78. Jiang, Y., et al., *Quantitative determination of peroxide value of edible oil by algorithm-assisted liquid interfacial surface enhanced Raman spectroscopy.* Food Chemistry, 2021. **344**: p. 128709.
79. Bērziņš, K., et al., *Low-wavenumber Raman spectral database of pharmaceutical excipients.* Vibrational Spectroscopy, 2020. **107**: p. 103021.
80. Salim, M., et al., *Application of Low-Frequency Raman Scattering Spectroscopy to Probe in Situ Drug Solubilization in Milk during Digestion.* The Journal of Physical Chemistry Letters, 2019. **10**(9): p. 2258-2263.
81. Vankeirsbilck, T., et al., *Applications of Raman spectroscopy in pharmaceutical analysis.* TrAC Trends in Analytical Chemistry, 2002. **21**: p. 869-877.
82. Robert, C., et al., *Monitoring the Isothermal Dehydration of Crystalline Hydrates Using Low-Frequency Raman Spectroscopy*. Molecular Pharmaceutics, 2021. **18**(3): p. 1264-1276.
83. Kong, K., et al., *Raman spectroscopy for medical diagnostics — From in-vitro biofluid assays to in-vivo cancer detection.* Advanced Drug Delivery Reviews, 2015. **89**: p. 121-134.
84. Pence, I. and A. Mahadevan-Jansen, *Clinical instrumentation and applications of Raman spectroscopy*. Chemical Society reviews, 2016. **45**(7): p. 1958-1979.
85. Ralbovsky, N.M. and I.K. Lednev, *Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning.* Chemical Society Reviews, 2020. **49**(20): p. 7428-7453.
86. De Biasio, M., et al., *Micro-Raman Spectroscopy for Meat Type Detection*. 2015.
87. FAO. *http://www.fao.org/3/i9286en/I9286EN.pdf*. 2018.
88. Bhat, M., et al., *Fraudulent Adulteration/Substitution of Meat: A Review.* International Journal of Recent Research and Applied Studies, 2015. **2**: p. 22-33.
89. Kim, G.D., et al., *Protein markers for discrimination of meat species in raw beef, pork and poultry and their mixtures.* Food Chem, 2017. **217**: p. 163-170.
90. Kumar, Y. and S. Chandrakant Karne, *Spectral analysis: A rapid tool for species detection in meat products.* Trends in Food Science & Technology, 2017. **62**: p. 59-67.
91. Monahan, F.J., O. Schmidt, and A.P. Moloney, *Meat provenance: Authentication of geographical origin and dietary background of meat.* Meat Science, 2018. **144**: p. 2-14.
92. Montowska, M., *Using Peptidomics to Determine the Authenticity of Processed Meat*, in *Proteomics in Food Science*. 2017. p. 225-240.
93. Montowska, M. and E. Fornal, *Absolute quantification of targeted meat and allergenic protein additive peptide markers in meat products.* Food Chem, 2019. **274**: p. 857-864.
94. Hou, B., et al., *Development of a sensitive and specific multiplex PCR method for the simultaneous detection of chicken, duck and goose DNA in meat products*. Meat Sci, 2015. **101**: p. 90-4.
95. Kitpipit, T., K. Sittichan, and P. Thanakiatkrai, *Direct-multiplex PCR assay for meat species identification in food products.* Food Chem, 2014. **163**: p. 77-82.
96. Safdar, M., et al., *A highly sensitive and specific tetraplex PCR assay for soybean, poultry, horse and pork species identification in sausages: development and validation.* Meat Sci, 2014. **98**(2): p. 296-300.
97. Hellberg, R.S., B.C. Hernandez, and E.L. Hernandez, *Identification of meat and poultry species in food products using DNA barcoding.* Food Control, 2017. **80**: p. 23-28.
98. José Fernando, V., et al., *Practical Application of DNA Fingerprinting To Trace Beef.* Journal of Food Protection, 2004. **67**(5): p. 972-979.
99. Arslan, A., et al., *Identification of meats using Random Amplified Polymorphic Dna (RAPD) technique.* Journal of Muscle Foods, 2005. **16**(1): p. 37-45.

100. Montowska, M. and E. Fornal, *Label-free quantification of meat proteins for evaluation of species composition of processed meat products.* Food Chem, 2017. **237**: p. 1092-1100.
101. Pavlidis, D.E., et al., *A volatilomics approach for off-line discrimination of minced beef and pork meat and their admixture using HS-SPME GC/MS in tandem with multivariate data analysis.* Meat Sci, 2019. **151**: p. 43-53.
102. Dayhuff, L.E. and M.J. Wells, *Identification of fatty acids in fishes collected from the Ohio River using gas chromatography-mass spectrometry in chemical ionization and electron impact modes.* J Chromatogr A, 2005. **1098**(1-2): p. 144-9.
103. Ayaz, Y., N.D. Ayaz, and I. Erol, *Detection of species in meat and meat products using enzyme-linked immunosorbent assay.* Journal of Muscle Foods, 2006. **17**(2): p. 214-220.
104. Yang, L., et al., *Rapid Identification of Pork Adulterated in the Beef and Mutton by Infrared Spectroscopy.* Journal of Spectroscopy, 2018. **2018**: p. 1-10.
105. Montowska, M., et al., *Authentication of processed meat products by peptidomic analysis using rapid ambient mass spectrometry.* Food Chem, 2015. **187**: p. 297-304.
106. Boyaci, İ.H., et al., *A rapid method for determination of the origin of meat and meat products based on the extracted fat spectra by using of Raman spectroscopy and chemometric method.* European Food Research and Technology, 2014. **238**(5): p. 845-852.
107. Andersen, P.V., E. Veiseth-Kent, and J.P. Wold, *Analyzing pH-induced changes in a myofibril model system with vibrational and fluorescence spectroscopy.* Meat Science, 2017. **125**: p. 1-9.
108. Argyri, A.A., et al., *A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage.* Food Control, 2013. **29**(2): p. 461-470.
109. Bauer, A., et al., *Assessment of tenderness of aged bovine gluteus medius muscles using Raman spectroscopy.* Meat Science, 2016. **115**: p. 27-33.
110. Berhe, D.T., et al., *Accurate determination of endpoint temperature of cooked meat after storage by Raman spectroscopy and chemometrics.* Food Control, 2015. **52**: p. 119-125.
111. Fowler, S.M., et al., *Raman spectroscopy compared against traditional predictors of shear force in lamb m. longissimus lumborum.* Meat Science, 2014. **98**(4): p. 652-656.
112. Tomasevic, I., et al., *Authenticity assessment of cooked emulsified sausages using Raman spectroscopy and chemometrics.* Fleischwirtschaft -Frankfurt-, 2016.
113. Smith, G.P.S., et al., *Raman imaging of drug delivery systems.* Advanced Drug Delivery Reviews, 2015. **89**: p. 21-41.
114. Lozano, M., et al., *Mid-Infrared Spectroscopy (MIR) for Simultaneous Determination of Fat and Protein Content in Meat of Several Animal Species.* Food Analytical Methods, 2017. **10**(10): p. 3462-3470.
115. Restaino, E., A. Fassio, and D. Cozzolino, *Discrimination of meat patés according to the animal species by means of near infrared spectroscopy and chemometrics Discriminación de muestras de paté de carne según tipo de especie mediante el uso de la espectroscopia en el infrarrojo cercano y la quimiometria.* CyTA - Journal of Food, 2011. **9**(3): p. 210-213.
116. Niu, X.Y., et al., *[Discrimination of donkey meat by NIR and chemometrics]*. Guang Pu Xue Yu Guang Pu Fen Xi, 2014. **34**(10): p. 2737-42.
117. Al-Jowder, O., E. Kemsley, and R. Wilson, *Detection of Adulteration in Cooked Meat Products by Mid-Infrared Spectroscopy.* Journal of agricultural and food chemistry, 2002. **50**: p. 1325-9.
118. Rannou, H. and G. Downey, *Discrimination of Raw Pork, Chicken and Turkey Meat by Spectroscopy in the Visible, Near- and Mid-infrared Ranges.* Analytical Communications, 1997. **34**(12): p. 401-404.

119. Pérez, N.F., J. Ferré, and R. Boqué, *Calculation of the reliability of classification in discriminant partial least-squares binary classification.* Chemometrics and Intelligent Laboratory Systems, 2009. **95**(2): p. 122-128.
120. Swift, A., R. Heale, and A. Twycross, *What are sensitivity and specificity?* Evidence Based Nursing, 2020. **23**(1): p. 2.
121. Craigie, C.R., et al., *Application of Hyperspectral imaging to predict the pH, intramuscular fatty acid content and composition of lamb M. longissimus lumborum at 24h post mortem.* Meat Science, 2017. **132**: p. 19-28.
122. Demsar, J., et al., *Orange: Data Mining Toolbox in Python.* Journal of Machine Learning Research, 2013. **14**(Aug): p. 2349−2353.
123. Lonergan, S.M., D.G. Topel, and D.N. Marple, *Chapter 5 - Fat and fat cells in domestic animals*, in *The Science of Animal Growth and Meat Technology (Second Edition)*, S.M. Lonergan, D.G. Topel, and D.N. Marple, Editors. 2019, Academic Press. p. 51-69.
124. Ortiz, C., et al., *Analysis of insulin amyloid fibrils by Raman spectroscopy.* Biophys Chem, 2007. **128**(2-3): p. 150-5.
125. Fabian, H. and P. Anzenbacher, *New developments in Raman spectroscopy of biological systems.* Vibrational Spectroscopy, 1993. **4**(2): p. 125-148.
126. Sane, S.U., S.M. Cramer, and T.M. Przybycien, *A holistic approach to protein secondary structure characterization using amide I band Raman spectroscopy.* Anal Biochem, 1999. **269**(2): p. 255-72.
127. Dong, J., et al., *Insulin assembly damps conformational fluctuations: Raman analysis of amide I linewidths in native states and fibrils.* J Mol Biol, 2003. **330**(2): p. 431-42.
128. Beattie, J.R., et al., *Prediction of adipose tissue composition using raman spectroscopy: Average properties and individual fatty acids.* Lipids, 2006. **41**(3): p. 287-294.
129. Beattie, J.R., S.E.J. Bell, and B.W. Moss, *A critical evaluation of Raman spectroscopy for the analysis of lipids: Fatty acid methyl esters.* Lipids, 2004. **39**(5): p. 407-419.
130. Ellis, D.I., et al., *Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool.* Analyst, 2013. **138**(14): p. 3871-3884.
131. Herrero, A.M., et al., *Chapter 10 - Vibrational Spectroscopy for Quality Assessment of Meat*, in *Poultry Quality Evaluation*, M. Petracci and C. Berri, Editors. 2017, Woodhead Publishing. p. 247-276.
132. Li-Chan, E., S. Nakai, and M. Hirotsuka, *Raman spectroscopy as a probe of protein structure in food systems*, in *Protein Structure-Function Relationships in Foods*, R.Y. Yada, R.L. Jackman, and J.L. Smith, Editors. 1994, Springer US: Boston, MA. p. 163-197.
133. Li-Chan, E.C.Y., *The applications of Raman spectroscopy in food science.* Trends in Food Science & Technology, 1996. **7**(11): p. 361-370.
134. Ostovar pour, S., et al., *Investigation of chemical composition of meat using spatially off-set Raman spectroscopy.* Analyst, 2019. **144**(8): p. 2618-2627.
135. Pelton, J.T. and L.R. McLean, *Spectroscopic Methods for Analysis of Protein Secondary Structure.* Analytical Biochemistry, 2000. **277**(2): p. 167-176.
136. Tu, A.T. *Raman spectroscopy in biology: Principles and applications*. 1982.
137. Herrero, A.M., *Raman Spectroscopy for Monitoring Protein Structure in Muscle Food Systems.* Critical Reviews in Food Science and Nutrition, 2008. **48**(6): p. 512-523.
138. Krimm, S. and J. Bandekar, *Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins*, in *Advances in Protein Chemistry*, C.B. Anfinsen, J.T. Edsall, and F.M. Richards, Editors. 1986, Academic Press. p. 181-364.
139. Careche, M., et al., *Structural Changes of Hake (Merluccius merluccius L.) Fillets: Effects of Freezing and Frozen Storage.* Journal of Agricultural and Food Chemistry, 1999. **47**(3): p. 952-959.

140. Lednev, I., *Protein Structures, Methods in Protein Structures and Stability Analysis.* Nova Sci, 2007: p. 1-26.
141. Baeten, V., et al., *Detection of Virgin Olive Oil Adulteration by Fourier Transform Raman Spectroscopy.* Journal of Agricultural and Food Chemistry, 1996. **44**(8): p. 2225-2230.
142. Gelder, J., et al., *Reference database of Raman spectra of biological molecules.* Journal of Raman Spectroscopy, 2007. **38**: p. 1133-1147.
143. Boyaci, I.H., et al., *Dispersive and FT-Raman spectroscopic methods in food analysis.* RSC Advances, 2015. **5**(70): p. 56606-56624.
144. Pelton, J.T. and L.R. McLean, *Spectroscopic methods for analysis of protein secondary structure.* Anal Biochem, 2000. **277**(2): p. 167-76.
145. Hernández-Martínez, M., et al., *Prediction of total fat, fatty acid composition and nutritional parameters in fish fillets using MID-FTIR spectroscopy and chemometrics.* LWT - Food Science and Technology, 2013. **52**(1): p. 12-20.
146. Meza-Márquez, O.G., T. Gallardo-Velázquez, and G. Osorio-Revilla, *Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef.* Meat Science, 2010. **86**(2): p. 511-519.
147. Carbonaro, M. and A. Nucara, *Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region.* Amino Acids, 2010. **38**(3): p. 679-690.
148. Papadopoulou, O., et al., *Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage.* Food Research International, 2011. **44**(10): p. 3264-3271.
149. Kohler, A., et al., *Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1143-1153.
150. Sinelli, N., et al., *Evaluation of freshness decay of minced beef stored in high-oxygen modified atmosphere packaged at different temperatures using NIR and MIR spectroscopy.* Meat Science, 2010. **86**(3): p. 748-752.
151. Alamprese, C., et al., *Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy.* LWT - Food Science and Technology, 2013. **53**(1): p. 225-232.
152. Shiroma, C. and L. Rodriguez-Saona, *Application of NIR and MIR spectroscopy in quality control of potato chips.* Journal of Food Composition and Analysis, 2009. **22**(6): p. 596-605.
153. Rohman, A., et al., *Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy.* Meat Science, 2011. **88**(1): p. 91-95.
154. Guan, Y., C.J. Wurrey, and G.J. Thomas, *Vibrational analysis of nucleic acids. I. The phosphodiester group in dimethyl phosphate model compounds: (CH3O)2PO2-, (CD3O)2PO2-, and (13CH3O)2PO2.* Biophysical Journal, 1994. **66**(1): p. 225-235.
155. Prache, S., et al., *Comparison of meat and carcass quality in organically reared and conventionally reared pasture-fed lambs.* Animal, 2011. **5**(12): p. 2001-9.
156. Fowler, S.M., et al., *Prediction of intramuscular fat content and major fatty acid groups of lamb M. longissimus lumborum using Raman spectroscopy.* Meat Science, 2015. **110**: p. 70-75.
157. Nache, M., et al., *Prediction of the pH as indicator of porcine meat quality using Raman spectroscopy and metaheuristics.* Chemometrics and Intelligent Laboratory Systems, 2016. **154**: p. 45-51.
158. Ferguson, D.M. and D.E. Gerrard, *Regulation of post-mortem glycolysis in ruminant muscle.* Animal Production Science, 2014. **54**(4): p. 464-481.

159. Stevenson-Barry, J.M., et al., *Incidence of high pH in venison: Implications for quality*, in *Proceedings of the New Zealand Society of Animal Production*. 1999, New Zealand Society of Animal Production. p. 145-147.
160. Węglarz, A., *Meat quality defined based on pH and colour depending on cattle category and slaughter season.* Czech Journal of Animal Science - UZEI (Czech Republic), 2010. **55**(12): p. 8.
161. Irshad, A., et al., *Factors Influencing Carcass Composition of Livestock.* Journal of Animal Production Advances, 2013. **3**: p. 1.
162. Smet, S., K. Raes, and D. Demeyer, *Meat fatty acid composition as affected by fatness and genetic factors: a review.* Animal Research, 2004. **53**: p. 81-98.
163. Troy, D.J., B.K. Tiwari, and S.-T. Joo, *Health Implications of Beef Intramuscular Fat Consumption.* Korean journal for food science of animal resources, 2016. **36**(5): p. 577-582.
164. Valsta, L.M., H. Tapanainen, and S. Männistö, *Meat fats in nutrition.* Meat Science, 2005. **70**(3): p. 525-530.
165. Berg, H., L. Dahlberg, and L. Mathiasson, *Determination of fat content and fatty acid composition in meat and meat products after supercritical fluid extraction.* J AOAC Int, 2002. **85**(5): p. 1064-9.
166. Korkeala, H., et al., *Determination of pH in meat.* Meat Science, 1986. **18**(2): p. 121-132.
167. Chen, Q., et al., *Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety.* TrAC Trends in Analytical Chemistry, 2013. **52**: p. 261-274.
168. Nunes, K.M., et al., *Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy.* Food Chem, 2016. **205**: p. 14-22.
169. Kiani, S., S. Minaei, and M. Ghasemi-Varnamkhasti, *Fusion of artificial senses as a robust approach to food quality assessment.* Journal of Food Engineering, 2016. **171**: p. 230-239.
170. Mitchell, H.B., *Multi-Sensor Data Fusion*. 1 ed. 2007: Springer-Verlag Berlin Heidelberg. XIV, 282.
171. Marquez, C., et al., *FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud.* Talanta, 2016. **161**: p. 80-86.
172. Borras, E., et al., *Data fusion methodologies for food and beverage authentication and quality assessment - a review.* Anal Chim Acta, 2015. **891**: p. 1-14.
173. Ballabio, D., et al., *Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey.* Food Chemistry, 2018. **266**: p. 79-89.
174. Castanedo, F., *A Review of Data Fusion Techniques.* The Scientific World Journal, 2013. **2013**: p. 704504.
175. Beganovic, A., et al., *Critical Review on the Utilization of Handheld and Portable Raman Spectrometry in Meat Science.* Foods, 2019. **8**(2).
176. Fourati, H., *Multisensor Data Fusion: From Algorithms and Architectural Design to Applications*. 2017: CRC Press.
177. Robert, C., et al., *Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy.* Food Chemistry, 2020: p. 128441.
178. Feiner, G., *4 - Definitions of terms used in meat science and technology*, in *Meat Products Handbook*, G. Feiner, Editor. 2006, Woodhead Publishing. p. 46-71.
179. Fowler, S.M., et al., *Predicting meat quality traits of ovine m. semimembranosus, both fresh and following freezing and thawing, using a hand held Raman spectroscopic device.* Meat Science, 2015. **108**: p. 138-144.

180. Scheier, R., A. Bauer, and H. Schmidt, *Early Postmortem Prediction of Meat Quality Traits of Porcine Semimembranosus Muscles Using a Portable Raman System.* Food and Bioprocess Technology, 2014. **7**(9): p. 2732-2741.
181. Scheier, R. and H. Schmidt, *Measurement of the pH value in pork meat early postmortem by Raman spectroscopy.* Applied Physics B, 2013. **111**(2): p. 289-297.
182. Eremenko, A., et al., *Effect of gold nanoparticles on an aerosil surface on the fluorescence and Raman spectra of adsorbed tryptophan.* Theoretical and Experimental Chemistry, 2012. **48**: p. 54-61.
183. Pedersen, D.K., et al., *Early prediction of water-holding capacity in meat by multivariate vibrational spectroscopy.* Meat Science, 2003. **65**(1): p. 581-592.
184. Aitipamula, S., et al., *Polymorphs, Salts, and Cocrystals: What's in a Name?* Crystal Growth & Design, 2012. **12**(5): p. 2147-2152.
185. Khankari, R.K. and D.J.W. Grant, *Pharmaceutical hydrates.* Thermochimica Acta, 1995. **248**: p. 61-79.
186. Bauer, J., et al., *Ritonavir: An Extraordinary Example of Conformational Polymorphism.* Pharmaceutical Research, 2001. **18**(6): p. 859-866.
187. Haleblian, J.K., *Characterization of Habits and Crystalline Modification of Solids and Their Pharmaceutical Applications.* Journal of Pharmaceutical Sciences, 1975. **64**(8): p. 1269-1288.
188. Hsieh, W.-h., et al., *Non-isothermal dehydration kinetic study of aspartame hemihydrate using DSC, TGA and DSC-FTIR microspectroscopy.* Asian Journal of Pharmaceutical Sciences, 2018. **13**(3): p. 212-219.
189. Khoo, J.Y., D.R. Williams, and J.Y.Y. Heng, *Dehydration Kinetics of Pharmaceutical Hydrate: Effects of Environmental Conditions and Crystal Forms.* Drying Technology, 2010. **28**(10): p. 1164-1169.
190. Kogermann, K., et al., *Establishing quantitative in-line analysis of multiple solid-state transformations during dehydration.* J Pharm Sci, 2008. **97**(11): p. 4983-99.
191. Fucke, K. and J.W. Steed, *X-ray and Neutron Diffraction in the Study of Organic Crystalline Hydrates.* Water, 2010. **2**(3): p. 333-350.
192. Kogermann, K., et al., *Investigating Dehydration from Compacts Using Terahertz Pulsed, Raman, and Near-Infrared Spectroscopy.* Applied Spectroscopy, 2007. **61**(12): p. 1265-1274.
193. Larkin, P.J., J. Wasylyk, and M. Raglione, *Application of Low- and Mid-Frequency Raman Spectroscopy to Characterize the Amorphous-Crystalline Transformation of Indomethacin.* Applied Spectroscopy, 2015. **69**(11): p. 1217-1228.
194. Walker, G., et al., *Probing Pharmaceutical Mixtures during Milling: The Potency of Low-Frequency Raman Spectroscopy in Identifying Disorder.* Molecular Pharmaceutics, 2017. **14**(12): p. 4675-4684.
195. Mah, P.T., et al., *Use of low-frequency Raman spectroscopy and chemometrics for the quantification of crystallinity in amorphous griseofulvin tablets.* Vibrational Spectroscopy, 2015. **77**: p. 10-16.
196. Paudel, A., D. Raijada, and J. Rantanen, *Raman spectroscopy in pharmaceutical product design.* Advanced Drug Delivery Reviews, 2015. **89**: p. 3-20.
197. Larkin, P.J., et al., *Polymorph Characterization of Active Pharmaceutical Ingredients (APIs) Using Low-Frequency Raman Spectroscopy.* Applied Spectroscopy, 2014. **68**(7): p. 758-776.
198. Salim, M., et al., *Application of Low-Frequency Raman Scattering Spectroscopy to Probe in Situ Drug Solubilization in Milk during Digestion.* The journal of physical chemistry letters, 2019. **10**(9): p. 2258-2263.

199. Smith, G.P.S., G.S. Huff, and K.C. Gordon, *Investigating Crystallinity Using Low Frequency Raman Spectroscopy: Applications in Pharmaceutical Analysis.* Spectroscopy (Santa Monica), 2016. **31**(2): p. 42-50.
200. Roy, S., B. Chamberlin, and A.J. Matzger, *Polymorph Discrimination Using Low Wavenumber Raman Spectroscopy.* Organic Process Research & Development, 2013. **17**(7): p. 976-980.
201. Bērziņš, K.r., et al., *Low-Frequency Raman Spectroscopic Study on Compression-Induced Destabilization in Melt-Quenched Amorphous Celecoxib.* Molecular Pharmaceutics, 2019. **16**(8): p. 3678-3686.
202. Hédoux, A., et al., *Dehydration mechanism of caffeine hydrate and structural description of driven metastable anhydrates analyzed by micro Raman spectroscopy.* International Journal of Pharmaceutics, 2015. **486**(1): p. 331-338.
203. Urabe, H., et al., *Low-Frequency Raman Spectra of Lysozyme Crystals and Oriented DNA Films: Dynamics of Crystal Water.* Biophysical Journal, 1998. **74**(3): p. 1533-1540.
204. Hédoux, A., et al., *Low- and high-frequency Raman investigations on caffeine: polymorphism, disorder and phase transformation.* The journal of physical chemistry. B, 2011. **115**(19): p. 5746-5753.
205. Hédoux, A., et al., *Using the low-frequency Raman spectroscopy to analyze the crystallization of amorphous indomethacin.* Eur J Pharm Sci, 2009. **38**(2): p. 156-64.
206. Hédoux, A., Y. Guinet, and M. Descamps, *The contribution of Raman spectroscopy to the analysis of phase transformations in pharmaceutical compounds.* International Journal of Pharmaceutics, 2011. **417**(1-2): p. 17-31.
207. Inoue, M., et al., *In Situ Monitoring of Crystalline Transformation of Carbamazepine Using Probe-Type Low-Frequency Raman Spectroscopy.* Organic Process Research & Development, 2017. **21**(2): p. 262-265.
208. Salim, M., et al., *Low-Frequency Raman Scattering Spectroscopy as an Accessible Approach to Understand Drug Solubilization in Milk-Based Formulations during Digestion.* Molecular Pharmaceutics, 2020. **17**(3): p. 885-899.
209. Bordos, E., et al., *Use of Terahertz-Raman Spectroscopy to Determine Solubility of the Crystalline Active Pharmaceutical Ingredient in Polymeric Matrices during Hot Melt Extrusion.* Molecular Pharmaceutics, 2019. **16**(10): p. 4361-4371.
210. Bērziņš, K.r., et al., *Solving the Computational Puzzle: Toward a Pragmatic Pathway for Modeling Low-Energy Vibrational Modes of Pharmaceutical Crystals.* Crystal Growth & Design, 2020. **20**(10): p. 6947-6955.
211. Mary, Y.S., et al., *DFT and molecular docking investigations of oxicam derivatives.* Heliyon, 2019. **5**(7): p. e02175.
212. Braun, D.E., et al., *Computational and Experimental Characterization of Five Crystal Forms of Thymine: Packing Polymorphism, Polytypism/Disorder and Stoichiometric 0.8-Hydrate.* Crystal growth & design, 2016. **16**(6): p. 3480-3496.
213. Tomerini, D. and G.M. Day, *Computational Methods for the Assignment of Vibrational Modes in Crystalline Materials*, in *Terahertz Spectroscopy and Imaging*, K.-E. Peiponen, A. Zeitler, and M. Kuwata-Gonokami, Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 151-190.
214. Rainsford, K.D., *Introduction and historical aspects of the side-effects of anti-inflammatory analgesic drugs*, in *Side-Effects of Anti-Inflammatory Drugs: Part One Clinical and Epidemiological Aspects*, K.D. Rainsford and G.P. Velo, Editors. 1987, Springer Netherlands: Dordrecht. p. 3-26.
215. Upadhyay, P.P. and A.D. Bond, *Crystallization and disorder of the polytypic α1 and α2 polymorphs of piroxicam.* CrystEngComm, 2015. **17**(28): p. 5266-5272.

216. Shi, X., et al., *Experimental electron densities of neutral and zwitterionic forms of the drug piroxicam.* CrystEngComm, 2016. **18**(18): p. 3289-3299.
217. Barnes, P.J., *Theophylline.* Pharmaceuticals, 2010. **3**(3): p. 725-747.
218. Khamar, D., et al., *Polymorphs of anhydrous theophylline: stable form IV consists of dimer pairs and metastable form I consists of hydrogen-bonded chains.* Acta Crystallographica Section C, 2011. **67**(12): p. o496-o499.
219. Ebisuzaki, Y., P.D. Boyle, and J.A. Smith, *Methylxanthines. I. Anhydrous Theophylline.* Acta Crystallographica 1997. **C53**: p. 777-779.
220. Fucke, K., et al., *New Insights into an Old Molecule: Interaction Energies of Theophylline Crystal Forms.* Crystal Growth & Design, 2012. **12**(3): p. 1395-1401.
221. Debnath, S. and R. Suryanarayanan, *Influence of processing-induced phase transformations on the dissolution of theophylline tablets.* AAPS PharmSciTech, 2004. **5**(1): p. E8-E8.
222. Malinowski, E.R., *Window factor analysis: Theoretical derivation and application to flow injection analysis data.* Journal of Chemometrics, 1992. **6**(1): p. 29-40.
223. Christensen, N.P.A., et al., *Rapid Insight into Heating-Induced Phase Transformations in the Solid State of the Calcium Salt of Atorvastatin Using Multivariate Data Analysis.* Pharmaceutical Research, 2013. **30**(3): p. 826-835.
224. Arnfast, L., et al., *Exploring the Complexity of Processing-Induced Dehydration during Hot Melt Extrusion Using In-Line Raman Spectroscopy.* Pharmaceutics, 2020. **12**(2): p. 116.
225. Kumar, K. and A.K. Mishra, *Multivariate curve resolution alternating least square (MCR-ALS) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets: Comparing certain ways of arranging TSFS-based three-way array.* Chemometrics and Intelligent Laboratory Systems, 2015. **147**: p. 66-74.
226. Dovesi, R., et al., *Quantum-mechanical condensed matter simulations with CRYSTAL.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2018. **8**(4): p. e1360.
227. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple.* Physical Review Letters, 1996. **77**(18): p. 3865-3868.
228. Grimme, S., et al., *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu.* The Journal of Chemical Physics, 2010. **132**(15): p. 154104.
229. Schäfer, A., H. Horn, and R. Ahlrichs, *Fully optimized contracted Gaussian basis sets for atoms Li to Kr.* The Journal of Chemical Physics, 1992. **97**(4): p. 2571-2577.
230. Ferrero, M., et al., *Coupled perturbed Hartree-Fock for periodic systems: The role of symmetry and related computational aspects.* The Journal of Chemical Physics, 2008. **128**(1): p. 014110.
231. Ugliengo, P., G. Borzani, and D. Viterbo, *MOLDRAW - program for the graphical manipulation of molecules on personal computers.* Journal of Applied Crystallography, 1988. **21**: p. 75.
232. Redenti, E., et al., *Raman and solid state 13C-NMR investigation of the structure of the 1 : 1 amorphous piroxicam : β-cyclodextrin inclusion compound.* Biospectroscopy, 1999. **5**(4): p. 243-251.
233. Beard, M.A., et al., *Monitoring dehydration kinetics using simultaneous thermal and spectral methods.* Journal of Raman Spectroscopy, 2010. **41**(10): p. 1283-1288.
234. Nunes, C., A. Mahendrasingam, and R. Suryanarayanan, *Investigation of the Multi-Step Dehydration Reaction of Theophylline Monohydrate Using 2-Dimensional Powder X-ray Diffractometry.* Pharmaceutical research, 2006. **23**: p. 2393-404.

235. Nolasco, M.M., A.M. Amado, and P.J.A. Ribeiro-Claro, *Computationally-Assisted Approach to the Vibrational Spectra of Molecular Crystals: Study of Hydrogen-Bonding and Pseudo-Polymorphism.* ChemPhysChem, 2006. **7**(10): p. 2150-2161.
236. Ohno, K., J. Kimura, and Y. Yamakita, *Strong Raman activities of low frequency vibrational modes in alkylbenzenes: conformation specific σ–π interactions between alkyl chain and benzene ring.* Chemical Physics Letters, 2001. **342**(1): p. 207-219.
237. Born, M. and K. Huang, *Dynamical theory of crystal lattices*. 1954: New York: Oxford University Press.
238. Steed, K.M. and J.W. Steed, *Packing Problems: High Z′ Crystal Structures and Their Relationship to Cocrystals, Inclusion Compounds, and Polymorphism.* Chemical Reviews, 2015. **115**(8): p. 2895-2933.
239. Khamar, D., et al., *Polymorphs of anhydrous theophylline: stable form IV consists of dimer pairs and metastable form I consists of hydrogen-bonded chains.* Acta crystallographica. Section C, Crystal structure communications, 2011. **67**: p. o496-9.
240. Sheth, A.R., et al., *Dehydration kinetics of piroxicam monohydrate and relationship to lattice energy and structure.* Journal of Pharmaceutical Sciences, 2004. **93**(12): p. 3013-3026.
241. Galvin, M. and D. Zerulla, *The Extreme Low-Frequency Raman Spectrum of Liquid Water.* ChemPhysChem, 2011. **12**(5): p. 913-914.
242. Perrier, P. and S.R. Byrn, *Influence of crystal packing on the solid-state desolvation of purine and pyrimidine hydrates: loss of water of crystallization from thymine monohydrate, cytosine monohydrate, 5-nitrouracil monohydrate, and 2'-deoxyadenosine monohydrate.* The Journal of Organic Chemistry, 1982. **47**(24): p. 4671-4676.
243. Vrečer, F., M. Vrbinc, and A. Meden, *Characterization of piroxicam crystal modifications.* International Journal of Pharmaceutics, 2003. **256**(1-2): p. 3-15.
244. Fucke, K., et al., *Unexpected Low-Temperature Behaviour of Piroxicam Monohydrate.* ChemPhysChem, 2013. **14**(4): p. 675-679.
245. Larkin, P., et al., *Polymorph Characterization of Active Pharmaceutical Ingredients (APIs) Using Low-Frequency Raman Spectroscopy.* Applied Spectroscopy, 2014. **68**.
246. Chowdhury, A.U., et al., *Kinetic Trapping of Metastable Amino Acid Polymorphs.* Journal of the American Chemical Society, 2014. **136**(6): p. 2404-2412.
247. Aaltonen, J., et al., *In-line monitoring of solid-state transitions during fluidisation.* Chemical Engineering Science, 2007. **62**(1): p. 408-415.
248. Amado, A.M., M.M. Nolasco, and P.J.A. Ribeiro-Claro, *Probing Pseudopolymorphic Transitions in Pharmaceutical Solids using Raman Spectroscopy: Hydration and Dehydration of Theophylline.* Journal of Pharmaceutical Sciences, 2007. **96**(5): p. 1366-1379.
249. Larsen, A.S., et al., *Determining short-lived solid forms during phase transformations using molecular dynamics.* CrystEngComm, 2019. **21**(27): p. 4020-4024.
250. Suihko, E., et al., *Dehydration of theophylline monohydrate—a two step process.* International Journal of Pharmaceutics, 1997. **158**(1): p. 47-55.
251. Duddu, S.P., et al., *Microcalorimetric investigation of phase transitions: I. Is water desorption from theophylline · HOH a single-step process?* International Journal of Pharmaceutics, 1995. **114**(2): p. 247-256.
252. Karjalainen, M., et al., *Characterization of polymorphic solid-state changes using variable temperature X-ray powder diffraction.* Journal of Pharmaceutical and Biomedical Analysis, 2005. **39**(1): p. 27-32.
253. Auner, G.W., et al., *Applications of Raman spectroscopy in cancer diagnosis.* Cancer metastasis reviews, 2018. **37**(4): p. 691-717.

254. McReynolds, N., et al., *Multimodal discrimination of immune cells using a combination of Raman spectroscopy and digital holographic microscopy.* Scientific Reports, 2017. **7**(1): p. 43631.
255. De Gelder, J., et al., *Reference database of Raman spectra of biological molecules.* Journal of Raman Spectroscopy, 2007. **38**(9): p. 1133-1147.
256. Ong, Y.H., M. Lim, and Q. Liu, *Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells.* Optics express, 2012. **20**: p. 22158-71.
257. Haka, A.S., et al., *Diagnosing breast cancer using Raman spectroscopy: prospective analysis.* Journal of biomedical optics, 2009. **14**(5): p. 054023-054023.
258. D'Acunto, M., et al., *Contribution of Raman Spectroscopy to Diagnosis and Grading of Chondrogenic Tumors.* Scientific Reports, 2020. **10**(1): p. 2155.
259. Beljebbar, A., et al., *Ex vivo and in vivo diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe.* Anal Bioanal Chem, 2010. **398**(1): p. 477-87.
260. Kalkanis, S.N., et al., *Raman spectroscopy to distinguish grey matter, necrosis, and glioblastoma multiforme in frozen tissue sections.* Journal of Neuro-Oncology, 2014. **116**(3): p. 477-485.
261. Koljenović, S., et al., *Discriminating Vital Tumor from Necrotic Tissue in Human Glioblastoma Tissue Samples by Raman Spectroscopy.* Laboratory Investigation, 2002. **82**(10): p. 1265-1277.
262. Zhang, J., et al., *Accuracy of Raman spectroscopy in differentiating brain tumor from normal brain tissue.* Oncotarget, 2017. **8**(22): p. 36824-36831.
263. Gajjar, K., et al., *Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis.* Analytical methods : advancing methods and applications, 2012. **5**: p. 89-102.
264. Zoladek, A., et al., *Development of Raman Imaging System for time-course imaging of single living cells.* Spectroscopy, 2010. **24**: p. 521962.
265. Smith, R., K.L. Wright, and L. Ashton, *Raman spectroscopy: an evolving technique for live cell studies.* Analyst, 2016. **141**(12): p. 3590-3600.
266. Ostrom, Q.T., et al., *CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014.* Neuro Oncol, 2017. **19**(suppl_5): p. v1-v88.
267. Stupp, R., et al., *Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial.* Lancet Oncol, 2009. **10**(5): p. 459-66.
268. Hatoum, A., R. Mohammed, and O. Zakieh, *The unique invasiveness of glioblastoma and possible drug targets on extracellular matrix.* Cancer Manag Res, 2019. **11**: p. 1843-1855.
269. Kahlert, U.D., G. Nikkhah, and J. Maciaczyk, *Epithelial-to-mesenchymal(-like) transition as a relevant molecular event in malignant gliomas.* Cancer Lett, 2013. **331**(2): p. 131-8.
270. Siebzehnrubl, F.A., et al., *The ZEB1 pathway links glioblastoma initiation, invasion and chemoresistance.* EMBO Mol Med, 2013. **5**(8): p. 1196-212.
271. Lupia, M., et al., *CD73 Regulates Stemness and Epithelial-Mesenchymal Transition in Ovarian Cancer-Initiating Cells.* Stem Cell Reports, 2018. **10**(4): p. 1412-1425.
272. Antonioli, L., et al., *CD39 and CD73 in immunity and inflammation.* Trends Mol Med, 2013. **19**(6): p. 355-67.
273. Chockley, P.J. and V.G. Keshamouni, *Immunological Consequences of Epithelial-Mesenchymal Transition in Tumor Progression.* J Immunol, 2016. **197**(3): p. 691-8.

274. Janeway, et al., *Immunebiology: the immune system in health and disease*. 6th ed. 2005, New York: Garland Science Publishing.
275. Olingy, C.E., H.Q. Dinh, and C.C. Hedrick, *Monocyte heterogeneity and functions in cancer.* Journal of Leukocyte Biology, 2019. **106**(2): p. 309-322.
276. Ahlers, J.D. and I.M. Belyakov, *Memories that last forever: strategies for optimizing vaccine T-cell memory.* Blood, 2010. **115**(9): p. 1678-89.
277. Alberts, B., et al., *Molecular biology of the cell*. 2017: WW Norton & Company.
278. Luckheeram, R.V., et al., *CD4(+)T cells: differentiation and functions.* Clin Dev Immunol, 2012. **2012**: p. 925135.
279. Zhang, N. and M.J. Bevan, *CD8(+) T cells: foot soldiers of the immune system.* Immunity, 2011. **35**(2): p. 161-8.
280. Bloch, O., et al., *Gliomas promote immunosuppression through induction of B7-H1 expression in tumor-associated macrophages.* Clin Cancer Res, 2013. **19**(12): p. 3165-75.
281. Mohme, M., et al., *Immunophenotyping of Newly Diagnosed and Recurrent Glioblastoma Defines Distinct Immune Exhaustion Profiles in Peripheral and Tumor-infiltrating Lymphocytes.* Clinical Cancer Research, 2018. **24**(17): p. 4187-4200.
282. Auner, G.W., et al., *Applications of Raman spectroscopy in cancer diagnosis.* Cancer Metastasis Rev, 2018. **37**(4): p. 691-717.
283. Jermyn, M., et al., *Intraoperative brain cancer detection with Raman spectroscopy in humans.* Sci Transl Med, 2015. **7**(274): p. 274ra19.
284. He, Y., et al., *Ramanome technology platform for label-free screening and sorting of microbial cell factories at single-cell resolution.* Biotechnology Advances, 2019. **37**(6): p. 107388.
285. Surmacki, J.M., et al., *Raman micro-spectroscopy for accurate identification of primary human bronchial epithelial cells.* Scientific Reports, 2018. **8**(1): p. 12604.
286. Vargas-Obieta, E., et al., *Breast cancer detection based on serum sample surface enhanced Raman spectroscopy.* Lasers in Medical Science, 2016. **31**(7): p. 1317-1324.
287. Daniel, A., et al., *Raman mapping of oral tissues for cancer diagnosis.* Journal of Raman Spectroscopy, 2014. **45**(7): p. 541-549.
288. Pyrgiotakis, G., et al., *Cell death discrimination with Raman spectroscopy and support vector machines.* Annals of biomedical engineering, 2009. **37**(7): p. 1464-1473.
289. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 2016. **374**(2065): p. 20150202-20150202.
290. Azambuja, J.H., et al., *CD73 Downregulation Decreases In Vitro and In Vivo Glioblastoma Growth.* Mol Neurobiol, 2019. **56**(5): p. 3260-3279.
291. Trevethan, R., *Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice.* Front Public Health, 2017. **5**: p. 307.
292. Aksoy, C. and F. Severcan, *Role of Vibrational Spectroscopy in Stem Cell Research.* Journal of Spectroscopy, 2012. **27**.
293. Galli, R., et al., *Rapid Label-Free Analysis of Brain Tumor Biopsies by Near Infrared Raman and Fluorescence Spectroscopy—A Study of 209 Patients.* Frontiers in Oncology, 2019. **9**(1165).
294. Lee, Y.J., et al., *Biocompatibility of a Novel Cyanoacrylate Based Tissue Adhesive: Cytotoxicity and Biochemical Property Evaluation.* PloS one, 2013. **8**: p. e79761.
295. Notingher, I., et al., *Spectroscopic study of human lung epithelial cells (A549) in culture: Living cells versus dead cells.* Biopolymers, 2003. **72**(4): p. 230-240.
296. Perez-Guaita, D., et al., *Multimodal vibrational imaging of cells.* Vibrational Spectroscopy, 2016.

297. Singh, G., et al., *The lag phase and G1 phase of a single yeast cell monitored by Raman microspectroscopy.* Journal of Raman Spectroscopy, 2006. **37**: p. 858-864.
298. Stone, N., et al., *Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers.* Journal of Raman Spectroscopy, 2002. **33**(7): p. 564-573.
299. Hobro, A.J., et al., *Deconstructing RNA: optical measurement of composition and structure.* Physical Chemistry Chemical Physics, 2013. **15**(31): p. 13199-13208.
300. Hobro, A.J., et al., *Raman spectroscopy as a tool for label-free lymphocyte cell line discrimination.* Analyst, 2016. **141**(12): p. 3756-3764.
301. George J. Thomas, J., *Raman spectroscopy of protein and nucleic acid assemblies* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**(1): p. 1-27.
302. Short, K.W., et al., *Raman Spectroscopy Detects Biochemical Changes Due to Proliferation in Mammalian Cell Cultures.* Biophysical Journal, 2005. **88**(6): p. 4274-4288.
303. Tang, M., et al., *Distinguishing Different Cancerous Human Cells by Raman Spectroscopy Based on Discriminant Analysis Methods.* Applied Sciences, 2017. **7**(9): p. 900.
304. Caponi, S., et al., *Raman micro-spectroscopy study of living SH-SY5Y cells adhering on different substrates.* Biophysical Chemistry, 2016. **208**: p. 48-53.
305. Pully, V.V., A.T.M. Lenferink, and C. Otto, *Time-lapse Raman imaging of single live lymphocytes.* Journal of Raman Spectroscopy, 2011. **42**(2): p. 167-173.
306. Delfino, I., et al., *Multivariate Analysis of Difference Raman Spectra of the Irradiated Nucleus and Cytoplasm Region of SH-SY5Y Human Neuroblastoma Cells.* Sensors, 2019. **19**: p. 3971.
307. Pyrgiotakis, G., et al., *Cell death discrimination with Raman spectroscopy and support vector machines.* Ann Biomed Eng, 2009. **37**(7): p. 1464-73.
308. Ichimura, T., et al., *Non-label immune cell state prediction using Raman spectroscopy.* Scientific Reports, 2016. **6**(1): p. 37562.
309. Singh, S.P., et al., *Identification of morphological and biochemical changes in keratin-8/18 knock-down cells using Raman spectroscopy.* J Biophotonics, 2017. **10**(10): p. 1377-1384.
310. Tomuţă, I., et al., *Multivariate Calibration for the Development of Vibrational Spectroscopic Methods*. 2018.
311. Burnham, A.K. and L.N. Dinh, *A comparison of isoconversional and model-fitting approaches to kinetic parameter estimation and application predictions.* Journal of Thermal Analysis and Calorimetry, 2007. **89**(2): p. 479-490.

Appendix 1

Supplementary information for isothermal dehydration of crystalline hydrates

The dehydration analysis of crystalline hydrates is given here. The initial and final spectra are presented to highlight solid-state forms post dehydration. The vibrational mode assignments and model free kinetic calculation are also presented to provide insight into the nature of the vibration as well as dehydration of the monohydrates.

A.1 Crystal structure of piroxicam

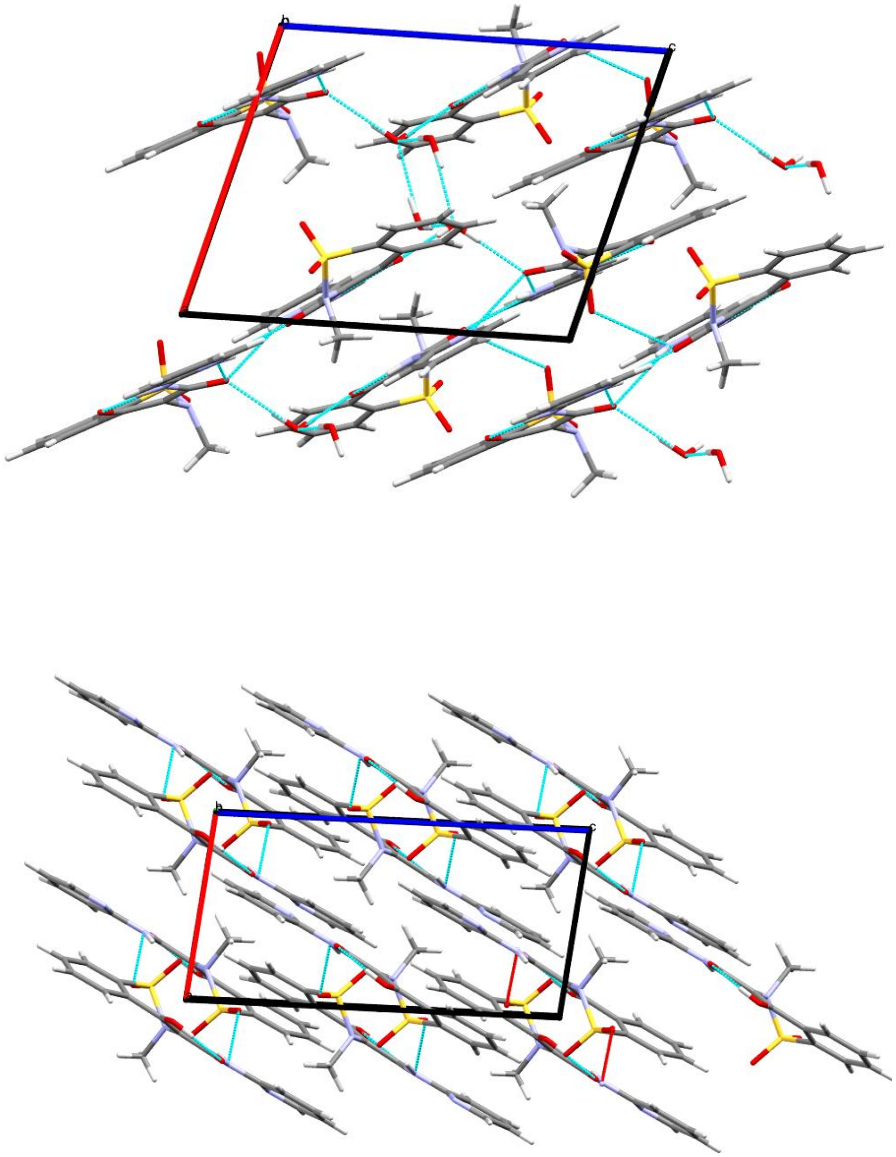

Figure A.1**:** Crystal structures of piroxicam monohydrate (top) and piroxicam anhydrous form I (bottom).

A.1.1 Dehydration of piroxicam monohydrate (PXM)

The initial and final spectra obtained during the dehydration of PXM was utilized to characterise the starting and end material over the course of dehydration. The Raman band assignments showed that the starting material was piroxicam monohydrate and the final form was anhydrous piroxicam form I.

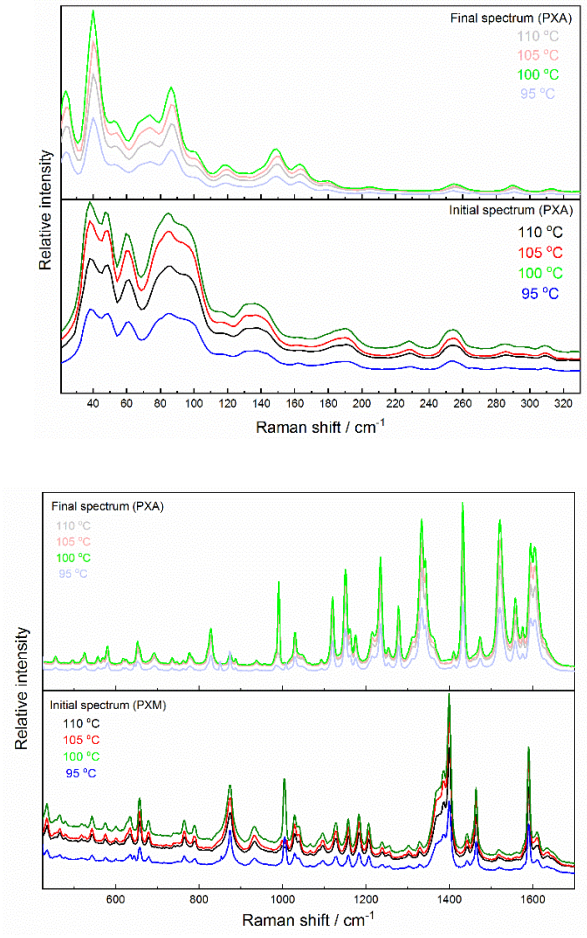

Figure A.2: Initial and final spectral data collected during dehydration of piroxicam monohydrate in the LFR (top) and MFR (bottom) at different temperatures. The initial (first) spectra acquired were consistent with Raman signatures of piroxicam monohydrate whereas the final spectra were consistent with the signatures of anhydrous piroxicam form I.

A.2 Dehydration of theophylline monohydrate (TPMH)

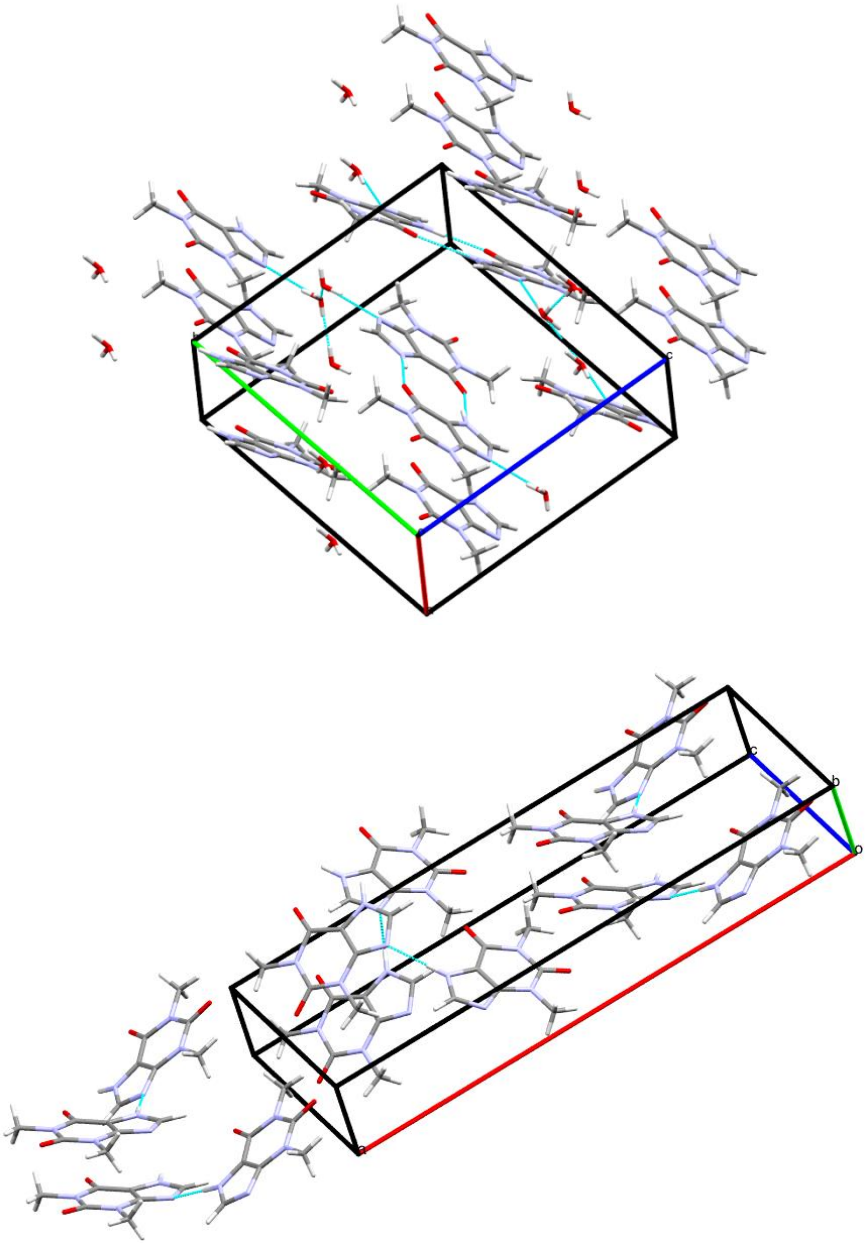

Figure A.3**:** Crystal structures of theophylline monohydrate (top) theophylline anhydrous form II (bottom)

A.2.1 Dehydration of theophylline monohydrate (TPMH)

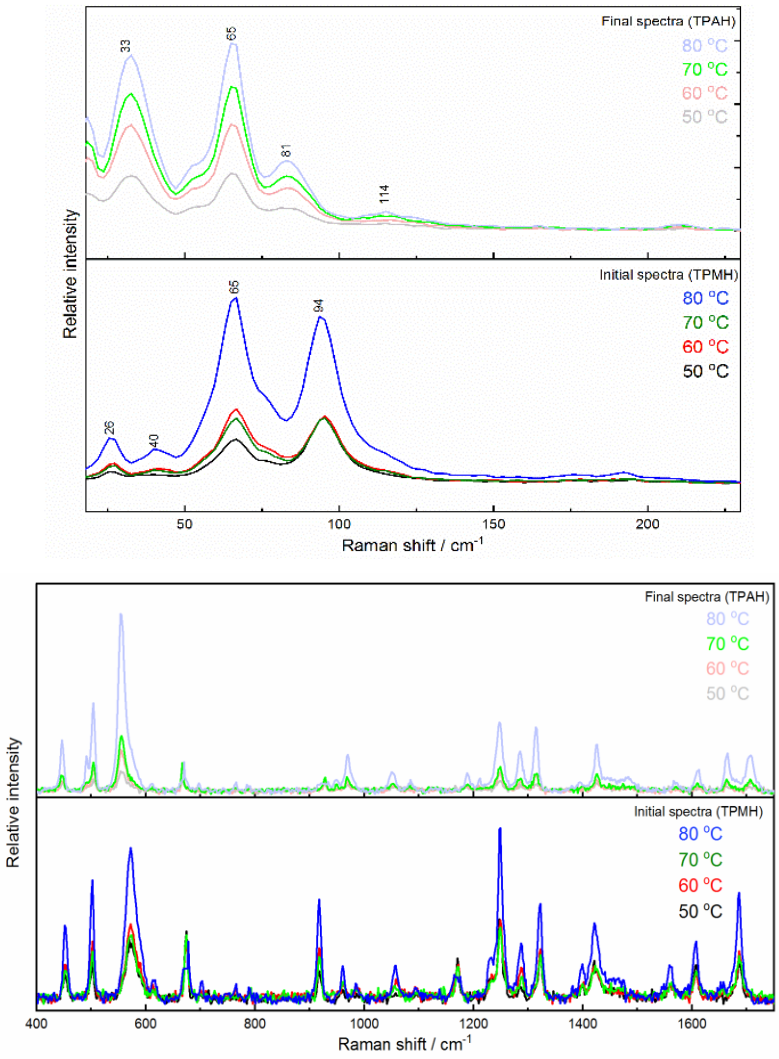

Figure A.4**:** Initial and final spectral data collected during dehydration of theophylline monohydrate in the LFR (top) and MFR (bottom) at different temperatures. The initial (first) spectra acquired were consistent with Raman signatures of theophylline monohydrate whereas the final spectra were consistent with the signatures of anhydrous theophylline form II.

A.3 Model free kinetic analysis

The complexity of the model fitting approach necessitated that the kinetic analysis be performed using a model free approach. The model free approach allows for kinetic analysis without prior knowledge of the reaction model as well as allowing for change in activation energy, E_a on changing conversion, α. Friedman analysis model free approach, which is based on differential data as shown in the equation, was utilized for the analysis [240, 311]. The kinetic analysis was calculated on data from the MCR output of the combined LFR and MFR regions.

$$\ln\left[\frac{d\alpha}{dt}\right] = \ln[Af(\alpha)] - \frac{E_a}{RT}$$

where $f(\alpha)$ is the differential kinetic function and A is the preexponential factor.

A representative plot of α versus t curves for the dehydration of piroxicam monohydrate is shown in the Figure A.5. The shape of the curve highlights increasing dehydration on increasing temperature under isothermal conditions.

Figure A.6 shows the variation of E_a as a function of α. E_a is observed to increase and decrease in the early stages of the dehydration ($0 \leq \alpha \leq 0.4$), which could suggest that the early dehydration stages involved different processes. At later stages, E_a is approximately constant and decreases at $0.4 \leq \alpha \leq 0.75$. The region where E_a is approximately constant can be assumed to follow a single-step reaction under isothermal heating condition.

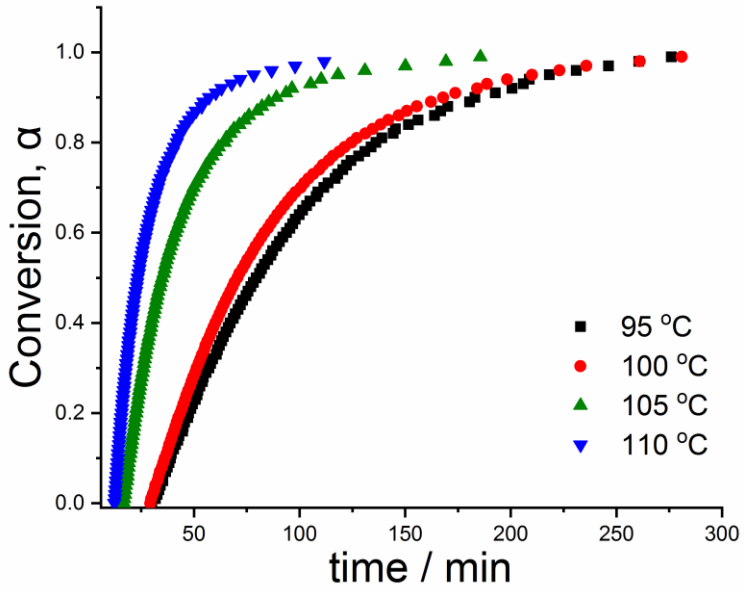

Figure A.5**:** Representative plot of conversion, α, against temperatures under different isothermal conditions.

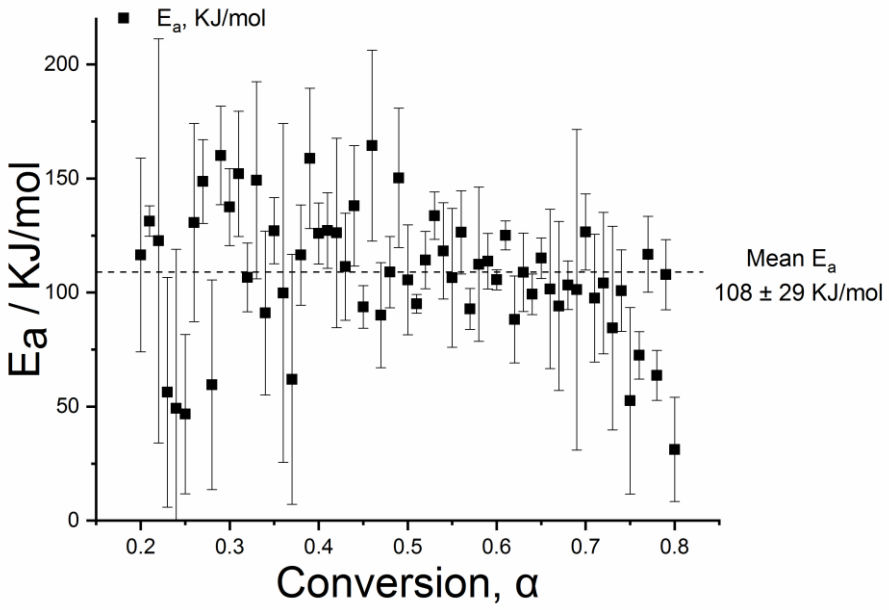

Figure A.6**:** Mean activation energy and standard deviation, E_a, plot against conversion, α (n = 3).

A.4 Vibrational mode assignments

The calculated vibrational modes were visualized using the MOLDRAW 2.0 (version H1) software

Table A.1: LFR Raman mode assignment for anhydrous piroxicam form I

Wavenumber (cm^{-1})	Relative intensity	Vibrational mode	Principal axis
16	393	Torsion	abc
29	238	Torsion	abc
42	190	Torsion	ab
47	163	Torsion	ab
61	159	Translational	ab
66	137	Torsion	b
71	114	Torsion	abc
76	121	Torsion	ac
86	84.6	Torsion	a
95	259	Torsion	abc
108	115	Torsion + Methyl group stretching	abc
116	8.12	Torsion + Methyl group stretching	abc
134	74.9	Torsion	ac
143	78.0	Torsion + Methyl group rotation	abc
154	76.6	Torsion + Methyl group stretching	abc
163	41.3	Torsion + Methyl group stretching	abc
171	89.1	Torsion + Methyl group rotation	abc
182	17.4	Complex torsion	abc
212	8.63	Torsion + Methyl group rotation	abc
252	22.1	Torsion + Methyl group rotation	abc
277	27.5	Torsion + Methyl group rotation	abc

Table A.2: LFR Raman mode assignment for piroxicam monohydrate

Wavenumber (cm^{-1})	Relative intensity	Vibrational mode	Principal axis
31	0.56	Torsion	ac
44	2.16	Torsion	abc
53	1.50	Torsion	abc
59	2.10	Torsion	abc
71	6.54	Torsion	abc
76	3.65	Torsion	a
84	0.97	Torsion	abc
88	7.87	Torsion	abc
92	2.81	Torsion	ca
103	17.3	Torsion + methyl group rotation	abc
103	5.14	Torsion	abc
121	3.87	Torsion + methyl group rotation	abc
128	5.69	Torsion + methyl group stretching	abc
132	2.61	Torsion	abc
144	7.63	Torsion	ab
161	3.45	Torsion	a
168	1.82	Torsion (H-bond stretching) + methyl group rotation	abc
173	2.57	Torsion	abc
178		Torsion (H-bond stretching)	abc
188	5.32	Torsion	abc
191	1.29	Torsion	ac
193	7.49	Torsion	ca
227	3.82	Torsion	a
242	5.01	Torsion	bc
247	7.63	Torsion	abc
251	3.05	Torsion	abc
267	3.35	Torsion	ac
275	11.1	Torsion	c
297	4.84	Torsion	abc

Table A.3: LFR Raman mode assignment for anhydrous theophylline form II

Wavenumber (cm^{-1})	Relative intensity	Vibrational mode	Principal axis
22	206	Torsion	abc
28	1.18	Translational	ac
34	54.6	Complex torsion	ab
44	223	Torsion	abc
52	18.9	Torsion	ca
52	40.0	Torsion	ac
65	144	Complex torsion	ab
82	1000	Torsion	ac
101	264	Torsion	a
122	84.0	Torsion + Methyl group rotation	abc
131	196	Torsion	a
144	95.0	Torsion	ab
150	23.4	Torsion	ab
165	1.11	Torsion + Methyl group rotation	abc
175	89.3	Torsion + Methyl group rotation	abc
193	38.3	Torsion + Methyl group rotation	abc
219	25.1	Torsion + Methyl group rotation	ab
288	0.28	Torsion	a

Table A.4: LFR Raman mode assignment for theophylline monohydrate

Wavenumber (cm^{-1})	Relative intensity	Vibrational mode	Principal axis
4	1000	Torsion	a
30	11.9	Torsion	ab
53	2.47	Torsion	abc
57	3.8	Complex torsion	abc
66	16.2	Torsion	ab
76	28.2	Torsion	ab
78	25.4	Torsion	ca
105	86.6	Torsion	abc
117	14.7	Torsion + methyl group rotation	abc
129	8.94	Torsion + methyl group rotation	abc
139	6.66	Torsion	abc
148	1.1	Torsion + methyl group stretching	abc
150	1.36	Torsion + methyl group stretching	abc
158	2.31	Torsion	ab
171	2.42	Torsion	ab
181	10.5	Torsion	abc
190	2.29	Torsion	abc
195	0.97	Torsion + methyl group rotation	abc
222	3.12	Torsion	ab
224	6.60	Torsion	ab
291	0.07	Torsion	abc
298	1.07	Torsion	abc