Light scattering during infrared spectroscopic measurements of biomedical samples

A thesis submitted to the University of Manchester for the degree of Ph.D. in the Faculty of Engineering and Physical Sciences

2011

Paul Bassan

Contents

Word count: 26,064

List of figures

List of publications

Paul Bassan, Hugh Byrne, Joe Lee, Franck, Colin Clarke, Paul Dumas, Ehsan Gazi, Michael Brown, Noel Clarke and Peter Gardner. "Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells", *Analyst*, 2009, **134**, 1171-1175.

Paul Bassan, Hugh Byrne, Franck Bonnier, Joe Lee, Paul Dumas and Peter Gardner. "Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the 'dispersion artefact'", *Analyst*, 2009, **134**, 1586-1593.

Paul Bassan, Achim Kohler, Harald Martens, Joe Lee, Hugh Byrne, Paul Dumas, Ehsan Gazi, Michael Brown, Noel Clarke and Peter Gardner. "Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples", *Analyst*, 2010, **135**, 268-277. **[Front cover]**

Paul Bassan, Achim Kohler, Harald Martens, Joe Lee, Edward Jackson, Nicholas Lockyer, Paul Dumas, Michael Brown, Noel Clarke and Peter Gardner. "RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing". J. Biophotonics, 2010, **3**, 1-12.

Caryn Hughes, Matthew Liew, Ashwin Sachdeva, **Paul Bassan**, Paul Dumas, Claire Hart, Mick Brown, Noel Clarke and Peter Gardner. "SR-FTIR spectroscopy of renal epithelial carcinoma side populations cells displaying stem cell-like characteristics". *Analyst*, 2010, **135**, 3133-3141.

Kevin Flower, Intisar Khalifa, **Paul Bassan**, Damien Demoulin, Edward Jackson, Nicholas Lockyer, Alan McGown, Philip Miles, Lisa Vaccari and Peter Gardner. "Synchrotron FTIR analysis of drug response to different drugs?", *Analyst*, **136**, 498-507.

Paul Bassan, Ashwin Sachdeva, Achim Kohler, Caryn Hughes, Alex Henderson, Jonathan Boyle, Michael Brown, Noel Clarke and Peter Gardner. "Does the reference spectrum used influence the outcome of the Resonant Mie Scattering Extended Multiplicative Signal Correction (RMieS-EMSC) algorithm?" In preparation.

Paul Bassan and Peter Gardner. "Scattering in biomedical infrared spectroscopy", Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach. *RSC Publishing*, 2010, 260-275.

Light scattering during infrared spectroscopic measurements of biomedical samples

Paul Bassan, University of Manchester, 2011

Degree: PhD

Abstract

Infrared (IR) spectroscopy has shown potential to quickly and non-destructively measure the chemical signatures of biomedical samples such as single biological cells, and tissue from biopsy. The size of a single cell (diameter ~10 – 50 μm) are of a similar magnitude to the mid-IR wavelengths of light (~1 – 10 μm) giving rise to Mie-type scattering. The result of this scattering is that chemical information is significantly distorted in the IR spectrum.

Distortions in biomedical IR spectra are often observed as a broad oscillating baseline on which the absorbance spectrum is superimposed. A spectral feature commonly observed is the sharp decrease in intensity at approximately 1700 cm^{-1}, next to the Amide I band (~1655 cm^{-1}), which pre-2009 was called the 'dispersion artefact'.

The first contributing factor towards the 'dispersion artefact' investigated was the reflection signal arising from the air to sample interface entering the collection optics during transflection experiments. This was theoretically modelled, and then experimentally verified. It was shown that IR mapping could be done using reflection mode, yielding information from the optically dense nucleus which previously caused extinction of light in transmission mode.

The most important contribution to the spectral distortions was due to resonant Mie scattering (RMieS) which occurs when the scattering particle is strongly absorbing such as biomedical samples. RMieS was shown to explain both the baselines in IR spectra, and the 'dispersion artefact' and was validated using a model system of poly(methyl methacrylate) (PMMA) of varying sizes from 5 to 15 μm. Theoretical simulations and experimental data had an excellent match thus proving the theory proposed.

With an understanding of the physics/mathematics of the spectral distortions, a correction algorithm was written, the RMieS extended multiplicative signal correction (RMieS-EMSC). This algorithm modelled the measured spectrum as superposition of a first guess (the reference spectrum) which was of a similar biochemical composition to the pure absorbance spectrum of the sample, and a scattering curve. The scattering curve was estimated as the linear combination of a database of a large number of scattering curves covering a range of feasible physical parameters. Simulated and measured data verified that the RMieS-EMSC increased IR spectral quality.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this of any other university or other institute of learning.

Copyright statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and she has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

II. Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

III. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Vice-President and the Dean of the Faculty of Life Sciences.

1. Introduction

1.1. Infrared spectroscopy

Infrared (IR) spectroscopy is a well established analytical chemistry technique used to gain insight into the molecular composition of samples. The approach is to transmit or reflect IR light (mid-IR wavelengths range from 1 to 10 μm) and observe the transmission or reflection profile against a non-absorbing background. The interaction of the photons with the sample occurs when the electromagnetic (EM) field interacts with molecular electric dipoles created betweens atoms that are vibrating. The molecular electric dipole is the result of unequal electronegitivities of the constituent atoms. The frequency of this vibration is characteristic and unique for different functional groups relating to unique frequencies of light being absorbed – this forms the premise of IR spectroscopy. IR spectroscopy comes under vibrational spectroscopy, which is concerned with studying the vibrations in molecules and giving information in terms of frequencies at which the vibrations occur. The most convenient way to display the information is by plotting a physical phenomenon (e.g. absorbance for IR spectroscopy) as a function of wavenumber (reciprocal wavelength) – such plots can be called vibrational spectra. Vibrations can occur from two or more atoms coupled, such as the carbonyl (C=O) group in amino acids[1].

The vibration of different functional groups occur at specific and characteristic frequencies related to the masses of the atoms involved and bond strength (analogous to the spring force constant in Hooke's law). The characteristic vibrational frequencies are known for the majority of commonly occurring chemicals[1].

During the 1940's and onwards, a host of literature was published investigating changes on the fundamental vibrations of bonds such as the N-H stretch in proteins under perturbation by effects such as acidity of sample solution[2-3]. Other regions of the infrared spectrum, such

as the near infrared (NIR) were also used in applications such as investigating the orientation of the carbonyl group in protein crystals and fibres[4]. Fundamental questions such as the quantification of protein denaturation by observing hydrogen bonding changes in “native protein” (i.e. protein not in a higher order structure) and coiled proteins gained supporting evidence from IR spectroscopy[5].

Vibrational spectra from biological samples yield a chemical signature of the sample, and cannot provide information as detailed as for example mass spectrometry which can identify individual peptides/proteins[6]. The IR spectra measured contain convoluted peaks in the region of interest as many biological components such as carbohydrates and proteins contain vibrations in the 600 to 1800 cm^{-1} wavenumber range[7]. This limits the application of the technique for specific studies such as protein identification, however overall profiles of chemical compositions can be acquired rapidly to yield useful information. Systems much more complicated than pure proteins such as single biological cells, and tissues are being increasingly studied using vibrational spectroscopy even though specific chemical identification is not possible. Applications such as cancer diagnostics need not necessarily require specific and quantitative chemical analysis, if a “broad” picture of the present biochemistry can be quickly measured. This is discussed in greater detail in the following sections.

1.2. Biomedical studies using IR spectroscopy

Figure 1.1 shows an IR spectrum from prostate tissue on a MirrIR substrate (which is reflective to mid-IR light). This spectrum is typical of a biological sample, containing chemical features from the proteins, carbohydrates, lipids and nucleic acids. Biological samples measured using spectroscopy range from single human cells[8-14], viruses[15-21], bacterial strains[16, 20, 22-40], to tissue[41-65].

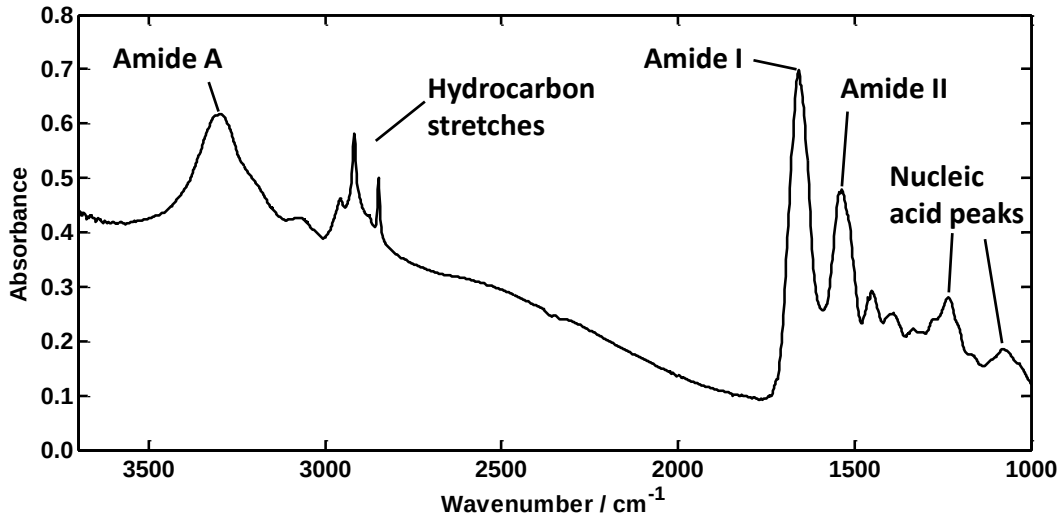

Figure 1.1 IR transflection spectrum from the stroma of prostate tissue.

The study of individual single biological cells has been an area of interest, for a number of different reasons. The change in chemistry induced within a single cell within a population from an external perturbation such as the addition of a cytotoxic agent has potential in applications such as drug screening. The particular change of chemistry within each cell and the population as a whole might give indications as to which drugs may be more suitable. Other applications of single cell study include cytology, such the study of oral cancers where single cells can be easily acquired from a mouth swab [7, 66-68].

In 1998, Jamin *et. al.* used a synchrotron (a high brightness that can be focused onto small areas such as 10 x 10 μm^2, used because equivalent brightness is not achievable using a

bench top source) coupled FTIR microscope to implement highly spatially resolved chemical imaging of single living cells[69]. This demonstrated that the distributions of important biochemical components of cells, such as the proteins and lipids could be observed using a non-destructive measurement technique. In Europe alone, there are over 10 infrared microspectroscopy beamlines at various Synchrotron installations showing the increased use and demand of the experimental setup. The advance in technology has allowed chemical mapping at a sub-nuclear level as demonstrated by Pijanka *et. al.* [12, 70-72].

The size of single human cells (10 – 50 µm in diameter) is similar to the wavelength of the IR light which gives rise to Mie-type scattering[73]. The effect of this scattering is that the IR spectrum is rendered unreliable as "features" in the spectrum are due to morphology rather than chemistry. In chapter 4 the physics of the scattering is stated in further detail.

There are numerous cases where conclusions related to biology are deduced from spectral information where the spectra are of questionable reliability. An example is a study by Holman *et. al.* on the IR spectroscopic signatures from cells during different phases of the cell cycle, and cell death[74].

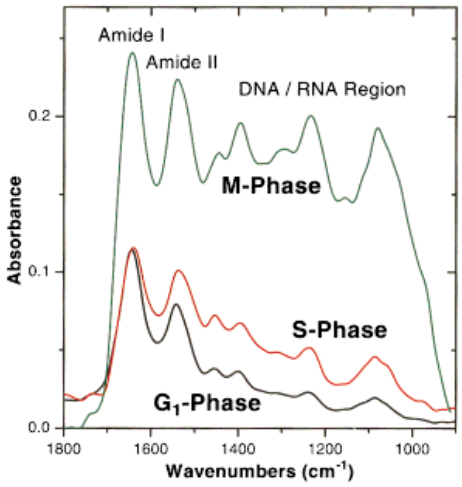

Figure 1.2 Three spectra from human lung fibroblasts, image taken from [74].

Figure 1.2 shows three spectra of single human lung fibroblast cells taken from reference[74]. These spectra have already had a linear baseline subtracted from 2000 to 650 cm^{-1} to deal with the scattering features present in the spectrum. These spectra are highly distorted, especially in the case of the M-phase spectrum where the amide I and amide II bands are in an unusual ratio which is almost always observed when strong Mie scattering is present. Holman *et. al.* concluded that these line shapes were typical of cells in the respective cell phases stated which at the time would have been a reasonable conclusion to have drawn, however in later years evidence that apparent changes in chemistry may be due to morphology rather than actual chemistry became apparent.

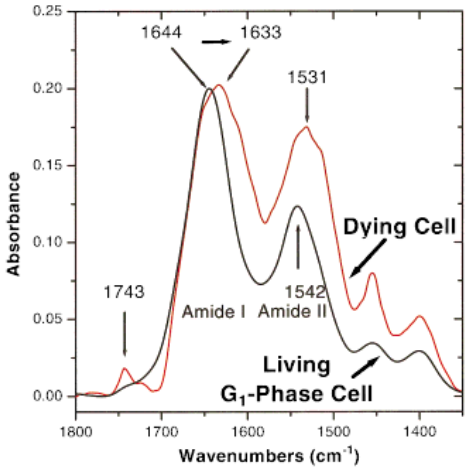

Figure 1.3 IR spectra of a living and dying cell, image taken from [74].

Figure 1.3 shows another figure from the same publication by Holman *et. al.* in which the difference between a living and a dying cell are compared. It is concluded that the dying cell exhibits a shift of the amide I band to a lower wavenumber, from 1644 cm^{-1} to 1633 cm^{-1}. A characteristic of a dying cell is that its shape becomes more rounded, which will inherently change its scattering characteristics. When Mie scattering occurs, the amide I band in the majority of cases is shifted to a lower wavenumber due to morphology only [73, 75].

Single cell IR spectroscopy is a field with great potential to quickly acquire a biochemical signature of cells within a population for numerous applications. The morphological characteristics and the wavelength of IR light make the acquisition of reliable spectra difficult. The usual solution for scattering related problems with sample for IR spectroscopy is to either dissolve the sample into liquid form, or grind the sample such that the size of the particles after grinding are sufficiently small not to cause scattering. Neither of these options are feasible for single cells, tissue samples or indeed other biological specimens because their IR spectrum would be affected. The secondary structure of proteins is greatly affected when in a liquid form, or when physically distorted by an operation such as grinding. These changes will be evident in the IR spectrum upon measurement, and are no longer representative of the original sample. Therefore it is imperative to understand the scattering physics occurring such that a solution can be pioneered to recover the true IR absorbance spectrum.

1.3. Spectral distortion

Infrared spectroscopy aims to measure the chemical signature of the sample irrelevant of its morphology, hence it is desired that the IR spectrum contain only chemical information. In practice, the “perfect” infrared spectroscopy experiment (see Figure 1.4) is impossible to achieve due to the nature of interaction of the electromagnetic field and the sample.

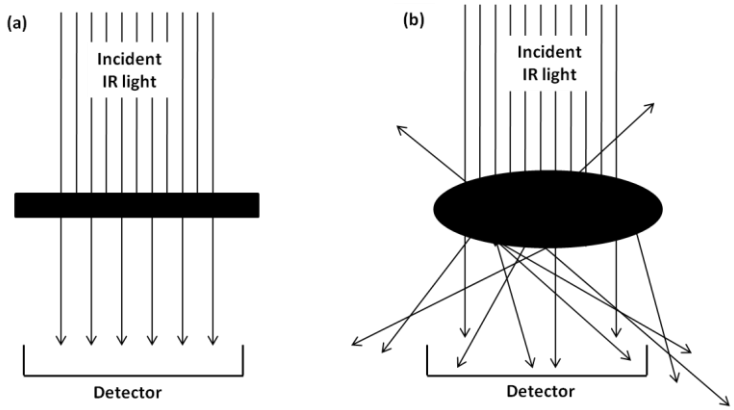

Figure 1.4 Infrared spectroscopic measurement of a flat and scattering sample.

In an ideal IR spectroscopic measurement, the light which does not reach the detector has been absorbed by the sample, which is approximately true for flat samples. Flat samples still result in some light loss due to reflections at interfaces, e.g. the air-sample interface, however the magnitude of this is usually rendered negligible compared to the absorbance of light by the sample. Biological samples such as single cells, give rise to scattering which results in the deviation of photon paths such that less light reaches the detector. When scattering occurs, the light that does not reach the detector is still interpreted as absorbance even though it was not absorbed by the sample. It is this problem that is the focus of this thesis, and background of progress in the literature is stated next.

1.3.1. Baselines

The most prominent feature in an IR spectrum due to light scattering is the baseline upon which the absorption bands are superimposed. Figure 1.5 shows a spectrum from a single prostate cancer cell[76] derived from a bone marrow metastases cell line, the so-called PC-3 cell line.

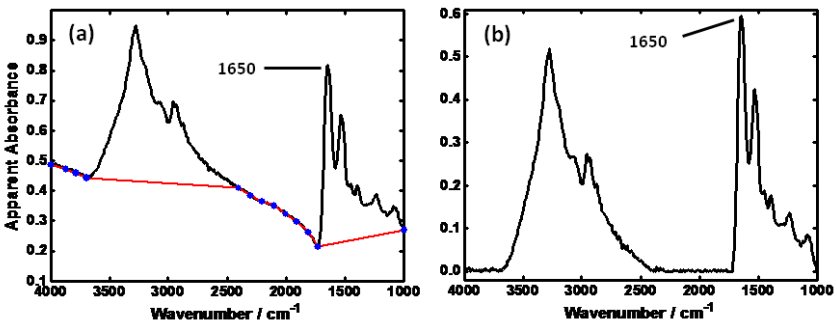

Figure 1.5 IR transmission spectrum of a single PC-3 cell.

Figure 1.5(a) shows the measured spectrum before any spectral pre-processing has been implemented. The wavenumber region of 1800 – 2500 cm^{-1} does not contain any vibrational absorption bands for biological spectra, hence the absorbance value should be zero. The apparent absorbance is greater than zero indicating that light has not been detected by the instrumentation, and this is the result of scattering. The broad baseline extends across the whole wavenumber range shown, where the blue points connected with the red line show an approximation of the curve.

Scattering results in additional features to the measured spectrum, beyond the broad baselines, the peak maxima positions are shifted either up or down in wavenumber value. At approximately, 1700 cm^{-1}, a sharp decrease in intensity is present which is unexpected when considering the general broad curve of the baseline, this effect has been name the 'dispersion artefact' and its origins not clearly understood.

Initial attempts at correcting spectra affected scattering involved placing a number of points along the baseline (Figure 1.5) and subtracting the resultant curve from joining the points together. This yields spectra which visually appear to be free from scattering effects, Figure 1.5(b), however still exhibit features of the scattering such as shifted peak maxima.

In 1991, Hartens *et. al.* published a method for estimating the "physical effects" present in spectra by assuming that the measured spectrum was a superposition of the absorption bands (analyte information) and a second order polynomial to account for the broad baselines[77]. This correction methodology worked well for near infrared spectra of turbid media, however could not account for all of the scattering variation in mid-IR spectra. The reason for this simply being that the scattering curves cannot be modelled as a second order polynomial[77].

In 2005, Mohlenhoff *et. al.*, attributed and approximated the broad baselines as Mie-type scattering of the infrared light from the nuclei of biological cells[73]. Figure 1.6 taken from this publication shows the measured spectrum from an oral mucosa cell (bottom) and the scattering curve calculated using a dielectric sphere of 4.2 µm diameter as an approximation (top trace).

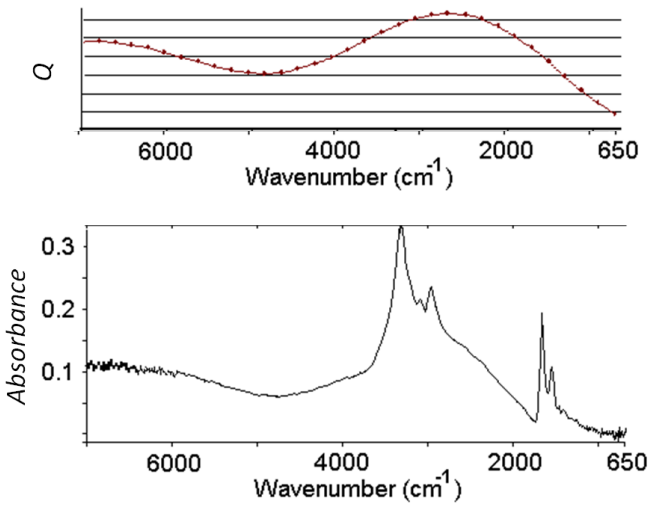

Figure 1.6 Bottom: Spectrum from an oral mucosa cell, and modelled scattering curve using the van de Hulst approximation. Figure reproduced from [73].

The scatter curve was calculated using the van de Hulst [78] approximation equation:

$$Q_{ext} = 2 - \frac{4}{\rho}\sin\rho + \frac{4}{\rho^2}(1 - \cos\rho) \quad (1)$$

$$\rho = 4\pi r(n-1)\lambda \quad (2)$$

where Q_{ext} = the Mie scattering extinction efficiency, r = radius of scattering particle, n = real part of the complex refractive index and λ = wavelength of light[78].

This formula is an approximation based on Mie theory[79], and describes the light loss as a function of wavelength for non-absorbing spherical particles illuminated by a parallel beam. In reality, an infrared spectroscopic measurement of a sample involves a strongly absorbing particle which is non-spherical and illuminated by a focused beam[80-82].

In 2008, Kohler *et. al.*, pioneered a method for the subtraction of Mie scattering baselines using a physical model to estimate the scattering contributions[10]. The method involved calculating 200 possible scattering curves, Q_{ext}, based on a particle radius, r, range of 2 to 40 µm, and a real refractive index, n, range of 1.1 to 1.5. With a database of a realistic range of scattering curves, the measured spectrum, Z_{Raw}, is estimated as a linear combination of the curves and a reference spectrum, Z_{Ref}. The reference spectrum is used to stabilise the estimation of the coefficients of each curve during a least squares fitting step[10].

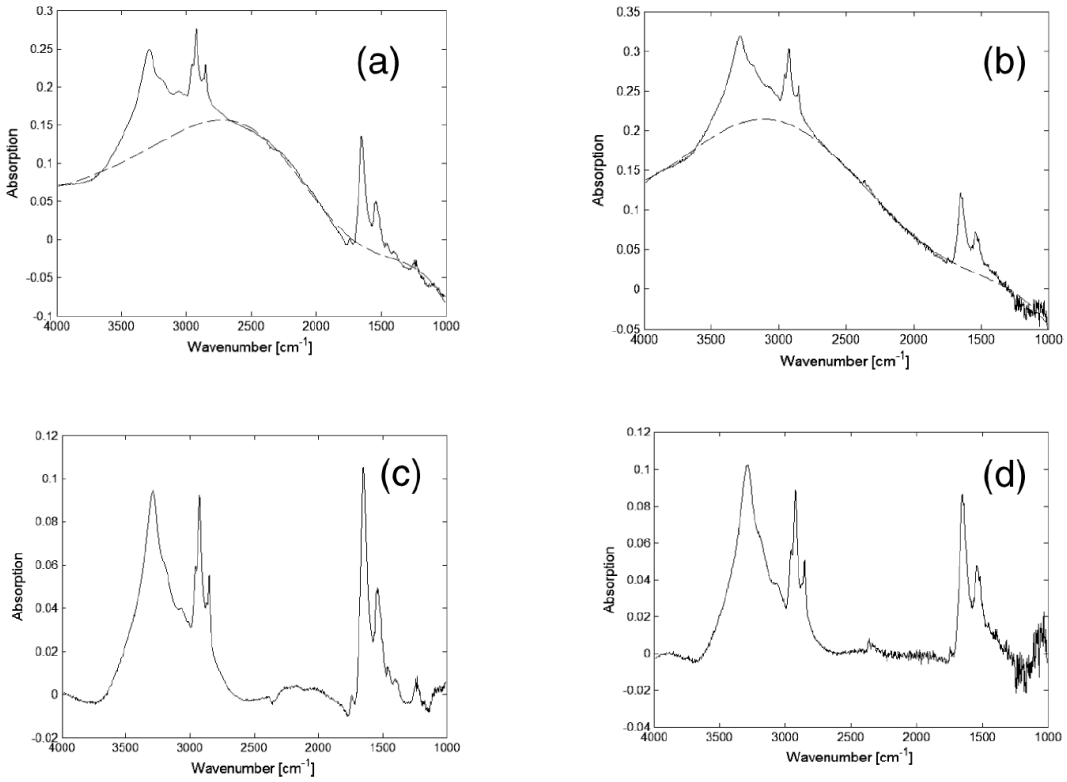

Figure 1.7 (a) and (b) are IR spectra from single lung cancer cells. (c) and (d) are the respective corrected spectra using the Kohler *et. al.* EMSC, figure reproduced from[10].

This Mie scattering correction algorithm removed the broad oscillation baselines successfully, however residual scattering effects remained as can be seen in Figure 1.7(c). The absorbance in the wavenumber range of 2000 to 2500 cm^{-1} shows a line shape indicative of light loss from scattering, and also at approximately 1800 cm^{-1}, there is a sharp decrease in intensity, the "dispersion artefact".

The inaccuracies in the correction algorithm were unknown, but assumed to be due to an oversimplified model of the physical phenomena.

1.4. Aims

There are several aims for this thesis, the first is to understand the origin of all the spectral distortions that occur during IR spectroscopic measurement of a scattering sample. This understanding should involve a mathematical basis which will be used to start producing a solution to remove the scattering effects. With a mathematical description of the physical phenomena occurring, an algorithm will be constructed which removes the scattering effects and recovers the true absorption spectrum of the sample. In short, the major aim of this project is remove morphological effects from IR spectra and retain the pure chemical signal thus allowing the development of IR analysis of scattering samples such as single biological cells and tissue.

2. Methods

In this section, methods including instrumentation and computational resources are presented that were used during this project.

2.1. Experimental methods

2.1.1 Infrared (IR) spectroscopy

Infrared (IR) spectroscopy is concerned with the interaction of IR light with materials, namely the transmission and reflection properties. The transmission and reflection properties give an insight into the chemical composition of the sample as different molecular vibrations occur at different wavelengths of light. The basic premise of the technique is to take a background measurement of the IR light where when no sample is present, and then to either transmit or reflect the light through the sample. By taking an appropriate ratio, a transmission or reflection spectrum can be computed, which can then be transformed into an absorbance spectrum.

2.1.2 Molecular vibrations

When IR light is incident on a sample, absorption can occur due to transition of molecular vibrations to excited states, namely from ground state to the first excited state, see Figure 2.1. When absorption occurs, the intensity of the light is reduced, because the ‘cost’ of the transition to an excited vibrational state is photons of a specific frequency.

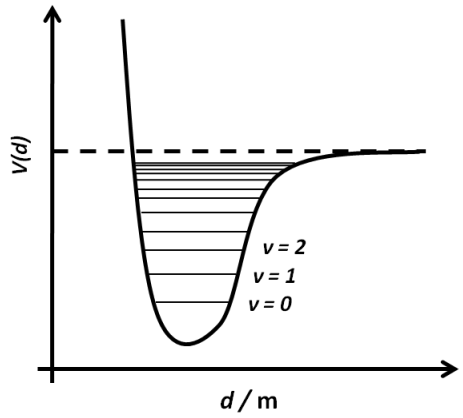

Figure 2.1 Potential energy for a diatomic as a function of displacement (d) during vibration for an anharmonic oscillator.

Coupled atoms in molecules vibrate in a conceptually similar manner as two masses joined by a spring. The frequency of vibration is specific to the atoms coupled, and the molecular environment in which they are contained. Factors such as the masses of atoms, bond strength force constants and nearby electromagnetic fields from surrounding and nearby atoms affect the frequency of vibration. When a vibration occurs between two atoms of differing electro-negativities a molecular electric dipole is created due to an uneven electron cloud distribution. During the vibration, the displacement of the atoms increases and decreases around an equilibrium position which results in a change in the electric dipole moment. This molecular electric dipole can interact with the EM field in the IR wavelength range such that energy from the field can be absorbed and promote the vibration to an excited state. IR spectroscopy is highly sensitive to vibrations which have a change in the electric dipole moment, less sensitive to vibrations which have little change in the electric dipole moment, and insensitive to no change in the electric dipole moment.

2.1.3. Fourier Transform Infrared (FTIR) method for spectroscopy

The preferred method to measure IR spectra today is by using Fourier Transform Infrared (FTIR) spectroscopy which has a number of advantages. The major advantage compared to dispersion instruments is that every wavelength of light can be measured at the same time, the so-called multiplex advantage. FTIR requires the use of a Michelson interferometer so that the EM-field through the sample can be measured as a time-domain signal, the interferogram.

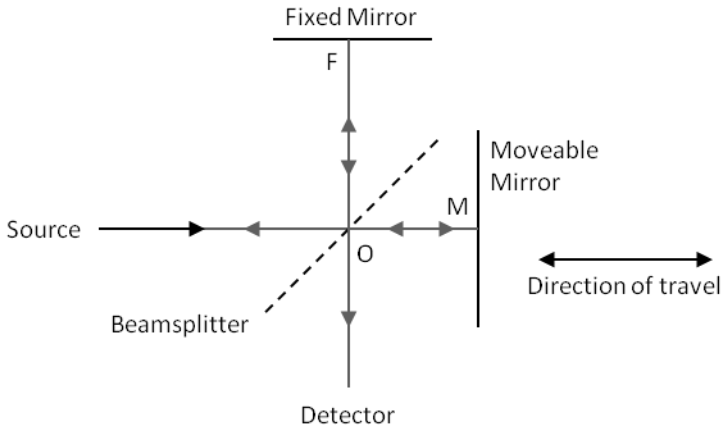

Figure 2.2 Michelson Interferometer use to measure an FTIR interferogram.

Figure 2.2 shows a Michelson interferometer which comprises of an incoming beam of IR light (the source) and two mirrors. One of these mirrors is fixed in position while the other moves back and forth, such that the beam leaving the beam-splitter towards the detector undergoes constructive and destructive interference of all the wavelengths at the same time. The position of the moving mirror is continuously recorded allowing intensity at the detector to be recorded as a function of mirror distance. The moving mirror is moving at a known speed, and this signal forms a time-domain signal.

An IR spectrum is most useful when displayed in the frequency domain, such that the x-axis of the spectrum is either in wavenumber (frequency) or wavelength. This is achieved by

performing a forwards discrete Fourier transform (DFT) on the time domain signal (the interferogram) using

$$F(\tilde{\nu}) = \int_{-\infty}^{\infty} f(t) e^{2\pi i \tilde{\nu} t} dt \qquad (3)$$

where *f(t)* is the intensity of the interferogram as a function of time.

An IR spectrum can be collected in a number of different geometries depending on whether a transmission or reflection spectrum is to be measured, this is discussed next.

2.1.2. Transmission mode FTIR

Transmission mode IR spectroscopy and specifically microspectroscopy is arguably the most established of the methods of collecting an IR spectrum. Most IR microscopes using an optical configuration similar to that shown in Figure 2.3 whereby a Cassegrain lens focuses the IR light onto the sample in a cone like fashion where it is transmitted through the sample. The sample needs to be supported by a window which ideally has 100% transmission for the IR light, a common material and the one used throughout this work is calcium fluoride (CaF_2) as it has near 100% transmission in the 1000 – 4000 cm^{-1} wavenumber range.

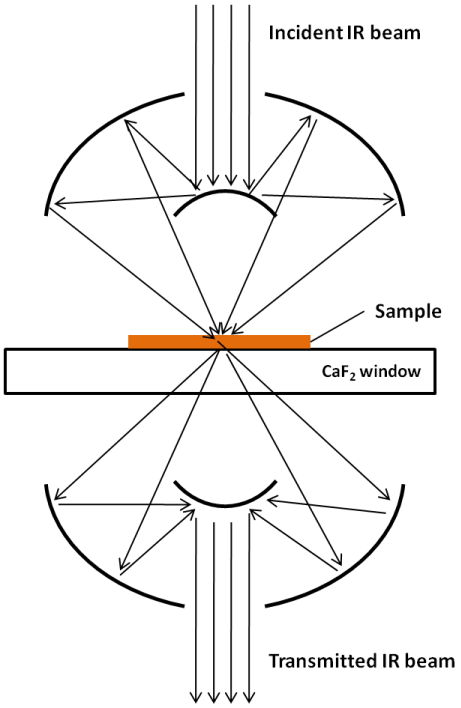

Figure 2.3 Schematic of transmission mode FTIR

Underneath the sample and window, the transmitted light is collected by a 'condenser lens' which is similar in construct to the Cassegrain lens, which collects the cone-like transmitted beam and focuses it into a point where it is then passed to the detector through a series of mirrors.

Transmission mode FTIR works well for samples that have optical density such that the absorbance values are within the Beer-Lambert regime (absorbance between 0.1 and 1.2 are typically quoted for this). Samples which are thick, and/or very strongly absorbing may result in total extinction of the IR light as it passes through the sample meaning that the detector will not count enough photons to give a meaningful spectrum. Samples which are very thick however, can still be measured using FTIR by reflecting light off the surface and observing the reflection spectrum.

2.1.3. Reflection mode FTIR

Reflection mode IR spectroscopy involves using just one Cassegrain placed above the sample, where half of the lens is used to focus light onto the surface of the sample, and the other half of the lens collects the reflected light and passes it to the detector, see Figure 2.4.

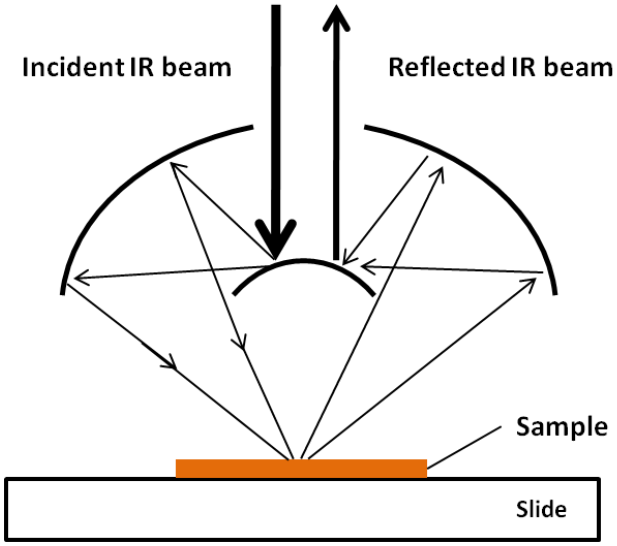

Figure 2.4 Schematic of reflection mode FTIR

The advantage of reflection mode is that sample thickness (assuming homogeneity is not an issue) is irrelevant, meaning that samples of several millimetres thickness can be measured. The reflection spectrum can be transformed into an absorbance spectrum by performing the Kramers-Kronig transform, see chapter 3.

A limitation of reflection mode FTIR is that the surface of the sample needs to be sufficiently polished to give rise to a strong signal, as reflection is an inherently weak phenomenon for biological samples. Surfaces giving rise to specular reflection rather than diffuse result in a much higher signal to noise ratio of the measured spectrum.

Other methods of collecting reflection involve the use of attenuated total-internal reflection (ATR) spectroscopy where a crystal with a high refractive index (such as diamond) is placed in contact with the sample surface. The angle of incidence of the incident light is such that it causes total-internal reflection. ATR instrumentation has been adapted to perform very high spatial resolution IR imaging[83-86].

The penetration depth of light when reflected from a surface is dependent on the wavelength and is approximately of the order of the wavelength itself. This means that the surface layer (typically several micrometers) is sampled which may not be a problem if the sample is chemically homogenous throughout, however this may not be the case for all samples.

2.1.4. Transflection mode FTIR

Transflection mode IR spectroscopy as the name suggests is a combination of transmission and reflection mode measurements. A sample is placed on a slide which is reflective to IR light, throughout this work MirrIR slides (Kevley Technologies, Chesterland, Ohio, USA) have been used due to their reflectivity to IR but high transmission of visible light. Transmission to visible light allows visible microscopy to be conducted which aids in sample location and slide positioning.

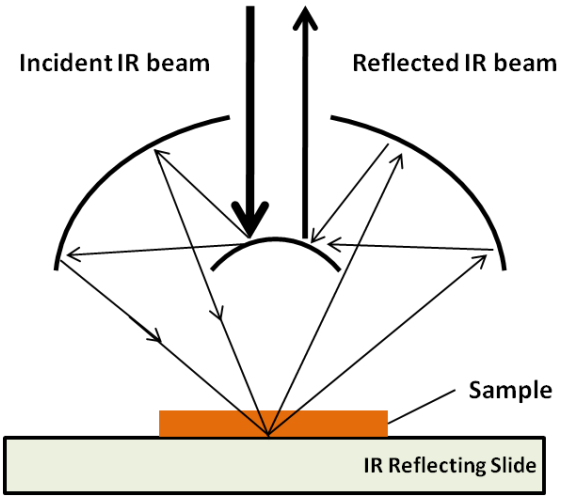

Figure 2.5 Schematic of transflection mode FTIR

Figure 2.5 shows a schematic of a transflection mode experiment, the incident beam transmits through the sample and is then reflected from the MirrIR surface before being transmitted back through the sample again to the collection optics. This essentially allows a transmission spectrum to be collected in a quasi-reflection geometry; the benefits include a double pass through the sample yielding a stronger absorbance signal.

MirrIR slides are often preferentially chosen over transmission substrates as they are considerably cheaper to work with, and are physically less brittle meaning that sample preparation is rendered somewhat easier. Optically speaking, fewer configurations are required as no condenser lens below the sample is required and as the Cassegrain lens distributes and collects the IR light.

2.1.5. Synchrotron coupled FTIR spectromicroscopy

Conventional bench-top FTIR spectrometers use a 'globar' as their source of IR light. This globar source is essentially a blackbody emitter made from a suitable material which can undergo many heating and cooling cycles, and does not degrade quickly at high

temperatures. The distribution of wavelengths emitted follow Plank's law of blackbody emission and the maintenance of a constant temperature is essential to maintain a stable wavelength distribution.

For the majority of IR experiments involving a sample size of larger than 20 x 20 μm^2, a globar source suffices to give adequate signal to noise. If however, smaller areas are to be interrogated for purposes such as sub-cellular IR imaging, then a globar source cannot provide a sufficiently brilliant light source[87]. Experiments involving an aperture of less than 10 x 10 μm^2 are often performed using a Synchrotron light source which provides a highly collimated light source that is several orders of magnitude more brilliant than globar sources[87].

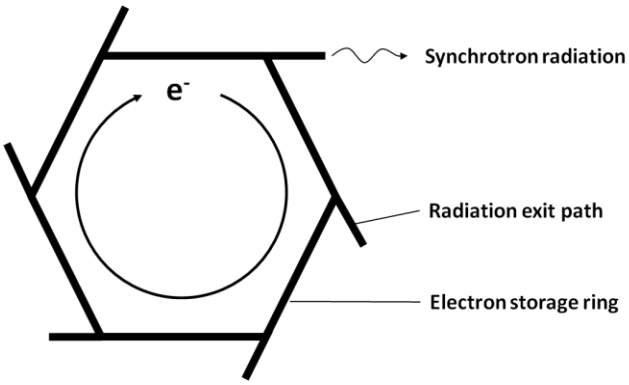

Figure 2.6 Schematic of a synchrotron storage ring, showing photon production at bending magnet.

Synchrotron coupled FTIR microscopy is essentially identical to bench-top measurements in which the same spectrometer and IR microscopes are used. The only difference is that the source of the IR light is produced externally by a Synchrotron ring, and inserted into the spectrometer using a set of external mirrors. A synchrotron is described most simply is a set of magnets with a central storage ring under vacuum in which electrons are accelerated to near the speed of light. A moving charge produces an EM field, the frequency of which is dependent on the movement of the charged particle. Figure 2.6 illustrates the concept

that when electrons change their path due to a magnetic field, synchrotron radiation is emitted in the original direction of the path. A range of photon energies are produced ranging IR to high energy x-rays, and are channelled using appropriate optical systems into various instruments.

2.2. Mathematical methods

2.2.1. Vectors and matrices

2.2.1.1. Vectors

Vectors are simply described as an array of numbers which can be denoted as a variable, such as: **x** = 1, 2, 3, 4, 5. Throughout this report, vectors will be denoted as bold lowercase letters. Vectors form a convenient way of manipulating lists of numbers, such as the wavenumber values of an IR spectrum. Mathematical descriptions are much simplified upon the use of vectors as they form concise variables to describe numerical entities.

Numbers within vectors can be denoted with a simple system such as x_i which represents the i^{th} value in vector ***x***.

2.2.1.2. Matrices

A matrix can be considered as a "block" of numbers in either a square or rectangular shape. Throughout this report, all matrices are row matrices meaning that each row is a vector. Matrices containing spectral information are constructed such that each row is the absorbance values of a single spectrum. The number of rows, *N*, corresponds to the number of spectra, and the number of columns, *K*, to the number of absorbance values, see Figure 2.7.

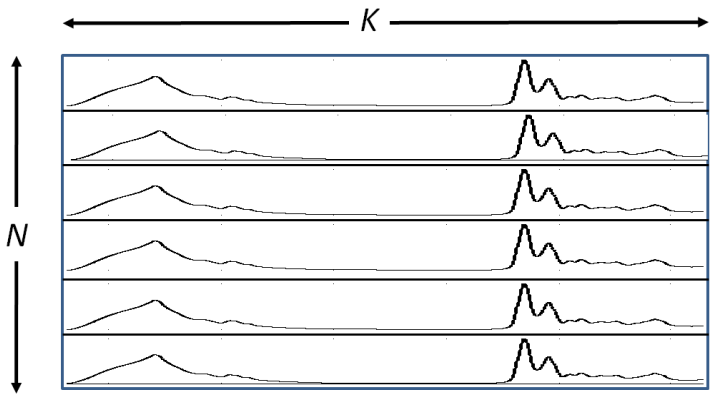

Figure 2.7 A spectral data matrix where each row corresponds to the absorbance values of each spectrum

Matrices will be denoted with a bold, non-italicised uppercase symbol such as **Y**. The operation of a matrix transpose rearranges a matrix such that its rows are organised as columns and vice versa. A transpose is written as a superscript "**T**" to the right of the relevant matrix, the transpose of matrix, **Y**, would simply be stated as $\mathbf{Y}^{\mathbf{T}}$.

2.2.2. Orthogonal vectors and the dot product

The dot product, *d*, of the two vectors **a** and **b**, of length *K* values, can calculated using,

$$d = \boldsymbol{a}.\boldsymbol{b} = \sum_{i}^{K} a_i b_i = \|\boldsymbol{a}\|\|\boldsymbol{b}\| \cos\theta \qquad (4)$$

Intuitively, this can be described as multiplying the elements of each vector and calculating the sum. Another definition of the dot product is product of the length of each vector multiplied by the cosine of the angle, ϑ, between them. Two vectors are considered to be orthogonal if their dot product is zero, e.g. the *x* and *y* axes on a graph which intersect at the origin only.

2.2.2. Principal component analysis (PCA)

Principal component analysis (PCA) is a powerful tool used for many applications, ranging from data analysis to data compression[88-90]. When data, such as vibrational spectroscopy spectra are to be analysed in an explorative manner (*i.e.* with no prior knowledge of the samples), PCA is often the first tool employed to gain insight into any patterns within the data. The patterns of interest in the bio-spectroscopy community are namely finding similarities and differences in spectra from two or more groups[91-99]. The two aspects of interest in this thesis are data exploration and data compression.

2.2.2.1. PCA for exploratory data analysis

The concept of PCA is best explained with an example, and so a dataset comprising of 25 spectra from two groups (totalling 50 spectra) will be used for illustration purposes. This data set was simulated such that there exists a clear difference between the two groups, the details of the construction of this data are stated in full in section 2.3.4.

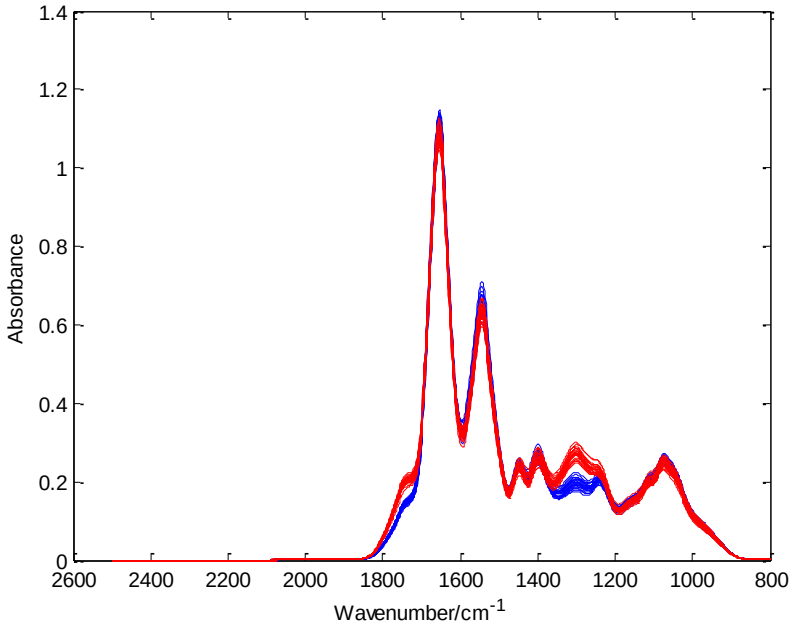

Figure 2.8 Simulated data comprising of two groups of 25 spectra.

Figure 2.8 shows the simulated data that are going to be used to illustrate the concepts of PCA. The data are visually similar to typical biomedical samples, exhibiting features from proteins, carbohydrates, lipids and nucleic acids. Due to the colour coding of the spectra it is immediately obvious that there are two groups of data, measured spectra from biological samples rarely contain this degree of variance, but PCA can be applied none the less. If a quantitative representation of the similarity / dissimilarity of the spectra are required, this cannot be easily obtained from visual assessment, and hence a numerical approach must be taken. In this example there are 50 spectra, each one of which is comprised of 851 absorbance values. These can be organised in a 50x851 matrix where each row corresponds to an individual spectrum, this matrix hereafter will be referred to as **Y**.

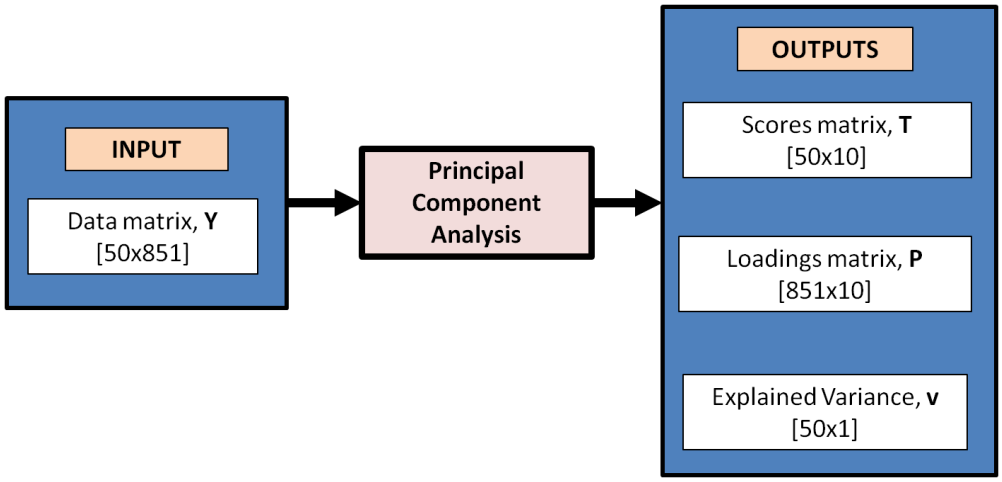

Figure 2.9 Flow chart illustrating input and outputs of principal component analysis (PCA), showing the sizes of the vectors and matrices involved.

Figure 2.9 shows a simple flow chart illustrating the input and outputs of PCA using our spectral data matrix example, **Y**, to show matrix dimensions of each variable. In this example, only 10 principal components (PCs) have been acquired as this is sufficient for the analysis to be conducted. The first output from PCA, is the scores, **T**, matrix and arguably the most informative for unsupervised analysis. The dimensions indicate that for each spectrum, there are 10 associated values, each relating to a principal component, PC1, PC2

... PC10. Figure 2.10 shows a scatter plot where the first and second columns of the scores matrix are plotted against one another.

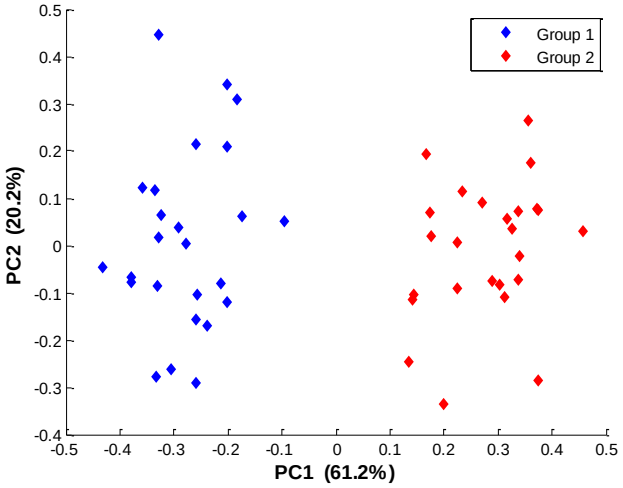

Figure 2.10 PCA scores plot for simulated data.

The scores plot represents each spectrum as a point in space, where spectra which are similar to one another are spaced closer together, and spectra which are dissimilar are spaced further apart. This visualisation technique allows for patterns of similarity in the data to be very quickly identified, and the process is unsupervised hence objective. Plots can also be created for any combination of the 10 PCs, and 3D-scatter plots can prove useful during pattern finding. The position of each point in the scores plot is defined by the mathematics behind principal component analysis.

2.2.2.2. The PCA algorithm

The mathematics behind PCA aim to decompose our data matrix, **Y**, into two simpler matrices which are the scores, **T**, and loadings, **P**. The first step of the algorithm, which is optional, but almost always performed is the mean centering of the data. This involves calculating the mean spectrum from our dataset of 50 spectra, and then subtracting this

from each individual spectrum. This results in a new matrix which now has a mean of zero, this data has been plotted in Figure 2.11.

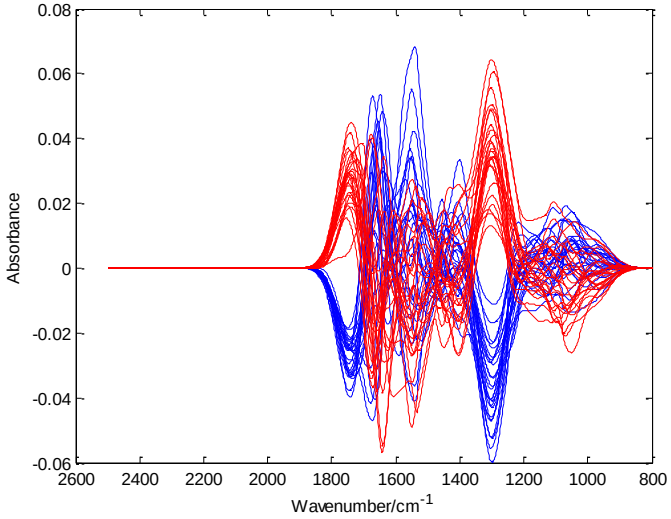

Figure 2.11 Mean centred data of the simulated data.

Figure 2.11 highlights the key differences in our data as differences in the absorbance the 1450 cm^{-1} and 1740 cm^{-1} peaks. The next operation PCA performs is to find a line of best fit through the data which minimises the sum of squares, which is the magnitude of the difference between the line of best fit and the data it is trying to describe. This line of best fit can be computed a number of different ways, giving the same answer, in this example the Non-Iterative Partial Least Squares (NIPALS) algorithm was employed to do this. The resultant best fit line is shown in Figure 2.12.

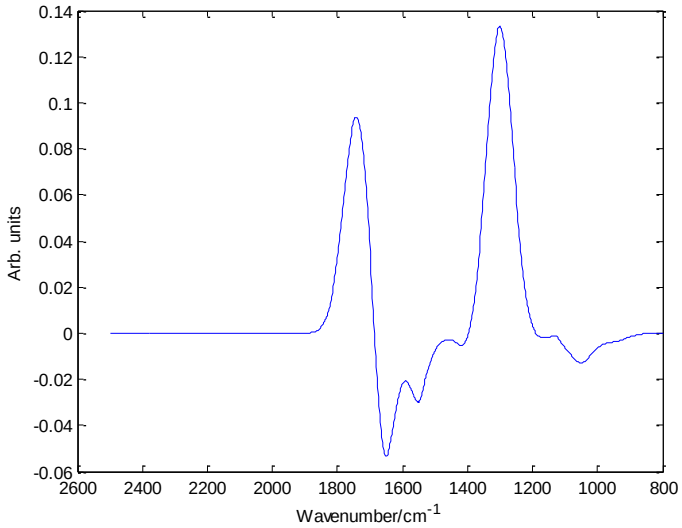

Figure 2.12 PC1 loadings curve for simulated data.

This line of best fit is called the loading curve for principal component 1, $\boldsymbol{p}_1$. The curve exhibits the obvious variance shown in the mean centred data, plus additional variance at approximately 1500 to 1630 cm^{-1}, which was not immediately obvious upon visual inspection. It is these subtle differences that a human observer cannot easily see, that make PCA a powerful tool. The next step is to calculate the "quantity" of this loading spectrum in each spectrum in our data matrix. This can be done a number of ways, either using a least squares fit, or by calculating the dot product of the loading and each spectrum. The number that is acquired from this for each spectrum is the score value for principal component 1, and can take a positive or negative value. Figure 2.12 shows that group 1 will have a negative score value as they are anti-correlated to the loading for PC 1. The scores plot in Figure 2.10 summaries all of these scores values concisely on a plot.

After the loading for PC 1 has been calculated, the matrix is then deflated which is done by subtracting the product of the score for each spectrum and the loading. This now yields a matrix which no longer contains any of the variance from PC 1, which allows PC 2 to be

calculated in the exact same manner as described above. This procedure can be repeated until the desired number of principal components are obtained.

The number of principal components "extractable" is limited to the dimensionality of the data, and typically varies between 4 and 20. Measured data contains noise which has by definition no underlying patterns, hence can give rise to an indefinite number of PCs. The aim of PCA is find as many PCs that can accurately describe all of the real chemical information in our dataset, and this is usually accomplished by PC 15.

The relation of the original data matrix (where no mean centring has been implemented), **Y**, the scores, **T**, and the loadings, **P** is:

$$\mathbf{Y}=\mathbf{T}\mathbf{P}^{\mathrm{T}} \tag{5}$$

PCA is often described as the mother of all chemometric techniques due to its simplicity and robustness at finding patterns in data with no prior knowledge of the data. The scores from PCA are routinely used for visual pattern finding, and then correlation to the loadings investigated to deduce what chemical features are contributing to the patterns in the data. The loadings from PCA exhibit useful properties which aid in data compression for a wide range of applications.

2.2.2.3. PCA for data compression

In the previous example, the data matrix, **Y**, contained 50 spectra which is relatively small compared to some data matrices. Matrices can range from hundreds to several thousands of rows and columns, e.g. IR images are routinely collected comprising of 128x128 = 16384 spectra.

It is often desirable to express large matrices in a simpler way and PCA offers this by decomposing a matrix into principle components. PCA identifies the most important features of the matrix and summarises the information into a much smaller number of vectors as done by Kohler *et. al.* in the Mie scattering EMSC[10] algorithm. In this work, the authors condensed 200 spectra of scattering curves into 6 curves which were able to almost fully reconstruct any of the original curves when added in the correct linear combination.

2.2.3. Linear regression

Linear regression, commonly performed using a least squares fitting approach is where a vector is modelled as a set other vectors. In the case of the Mie scattering EMSC by Kohler *et. al.*[10], the raw measured spectrum, Z_{Raw}, is modelled as the linear combination of a reference spectrum (such as a spectrum of Matrigel) and a number of 'descriptive vectors' which make up the baseline. Expressed algebraically, using vectors denoted as symbols with an arrow above them, the Mie scattering EMSC algorithm is

$$\vec{Z}_{Raw} = h\vec{Z}_{Ref} + c + m\vec{\nu} + \sum_{i=1}^{6} g_i\vec{p}_i + \vec{E} \qquad (6)$$

where h = multiplicative factor to describe effective optical path length of sample, c = constant offset baseline, m = gradient of sloping baseline, g_i = weighting i of loading vector p_i, and E = vector of un-modelled features, also known as the residual of the model.

In this particular EMSC model using 6 loading vectors to model the scattering curves, a total of 9 parameters have to calculated using a least squares algorithm. Computationally speaking, this calculation is trivial for modern computers as the number of parameters required to be calculated is small.

2.3. Computational methods

2.3.1. Programming language

All of the computations presented were implemented in a commercial programming language called Matlab 2010a (Mathworks, Natick, MA, USA). This package was chosen due to its accessibility to scientists with no prior programming experience and its superb precision at computing linear algebra functions. Also supported in this package are complicated features such as Artificial Neural Network (ANN) pattern recognition.

2.3.2. High throughput computing (HTC)

In chapter 6 the resonant Mie scattering correction algorithm is implemented for infrared images of prostate tissue which comprise of 128x128 pixels equating to 16384 spectra. The time required to correct one image is approximately 4 days using a quad core Intel Xeon processor which makes correcting large numbers of images unfeasible. To reduce computation time for these images, high throughput computing was employed using a pool of 1500 Linux machines via the Condor computing system developed at the University of Wisconsin in the USA.

The Condor philosophy is that spare computational resources can be put to use for scientific computing. Spare resources frequently arise at institutions when computers are not in use, e.g. during lunch times and overnight. This approach only works for jobs that can be run in parallel and are independent of each other. The spectra contained within an infrared image are independent of one another, and signal correction of these can be implemented in parallel.

To use Condor computing, an IR image initially has to be split into a number of jobs, typically around 1024, resulting in 16 spectra per job. These jobs are then submitted to a master node which is connected to 1500 worker nodes, see Figure 2.13.

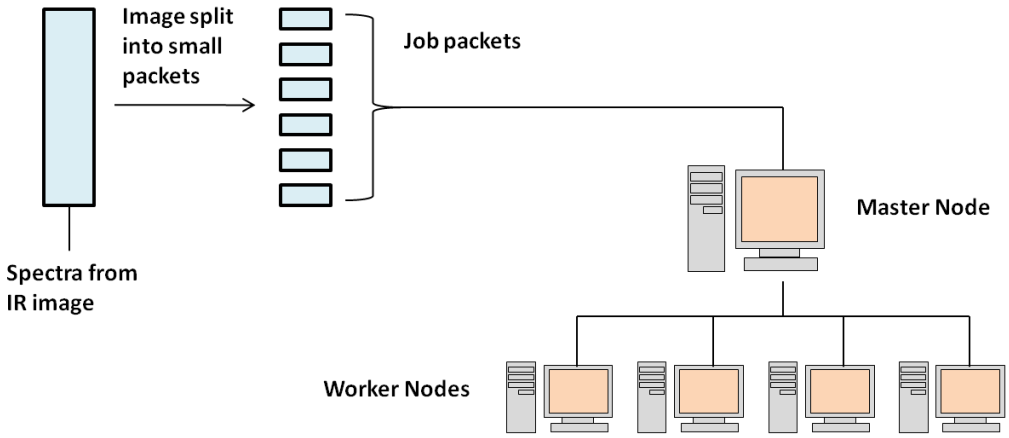

Figure 2.13 Condor high throughout computing.

Using the Condor system has reduced computation time for IR images from 4 days to 30 minutes meaning that large numbers of images can be collected and processed on a daily basis. It is acknowledged that the first step towards efficient computation is to rewrite the program in a faster language such as FORTRAN or C/C++, however in the absence of the required expertise and an opportunity to use Condor for zero cost, high throughput computing was chosen preferentially.

2.3.3. Artificial Neural Networks (ANNs)

The concept of pattern recognition and machine learning involves “teaching” a computer via a set of algorithms to recognise data types as belonging to a particular class. In the field of cancer diagnostics, a simple example would be training the machine to recognise what cancerous epithelial cells “look” like, and doing the same for healthy epithelial cells. Once the machine has learnt the patterns successfully, it can be subjected to new and unknown

spectra upon which it makes a decision from its learning process as to which class the spectra belong, i.e. cancerous or healthy in this case.

There are a host of machine learning algorithms available, for this project artificial neural networks (ANNs) have been chosen due to their proven success in the field of biomedical vibrational spectroscopy[100-103], and ease of implementation in the Matlab platform.

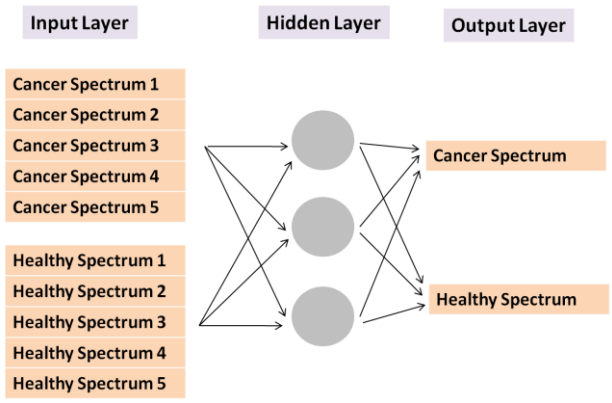

Figure 2.14 Schematic of a 3 layer artificial neural network.

In this project three layer ANNs are used for all pattern recognition and classification problems. The first layer comprises of the input spectra, in this case cancerous and healthy spectra, which then passed to a "hidden layer" where a number of neurons which are user defined perform a number of operations to "learn" that the input spectra belong to the memberships stated. These hidden neurons act in a manner which is meant to mimic the functioning of neurons within the human brain, the key property being that they can act in a "non-linear" way. The human brain seemly learns and solves problems in a manner which is different from conventional learning done by a computer with great success, and it is this property of ANNs that is being taken advantage of. In simple terms, the ANN is given some

data and told which class membership each belongs to, and is then asked to find some way to learn the features of cancerous and healthy tissue.

The ANN learning function used throughout this project is the conjugate gradient back-propagation method as this was found to have the highest classification accuracy of those available in the Matlab platform.

2.3.4 Simulated spectra

Later in this thesis a set of spectra are required to test the scattering correction presented. Measured spectra from biological samples would seem ideal for testing purposes, however as the pure absorbance spectrum is unknown there is nothing to compare results to, hence the need for some known standard.

An absorption band in an infrared spectrum can be approximated as a Gaussian lineshape, and a spectrum approximated as merely a summation of Gaussians centred at various peak positions. Stated below is the equation of a Gaussian:

$$f(\tilde{\nu}) = A.\exp\left(-\frac{(\tilde{\nu}-\tilde{\nu}_0)^2}{2c^2}\right) \quad (7)$$

Where $\tilde{\nu}_0$ = peak maximum position, A = amplitude of Gaussian and c = width parameter.

The spectrum, **S**, is the sum of a number of Gaussians with various peak positions, heights and width, this can be expressed as:

$$S(\tilde{\nu}) = \sum_i A_i \exp\left(-\frac{(\tilde{\nu}-\tilde{\nu}_{0i})^2}{2c_i^2}\right) \quad (8)$$

Where $S(\tilde{\nu})$ = absorbance at wavenumber $\tilde{\nu}$, A_i = amplitude of peak i, $\tilde{\nu}_{0i}$ = peak position of peak i, and c_i = width parameter of peak i.

The starting point to construct spectra similar in appearance to those of biomedical samples was the use of a spectrum of Matrigel which is an artificial extracellular matrix comprising of proteins, lipids, carbohydrates and other growth factors, its IR spectrum is shown in Figure 2.15.

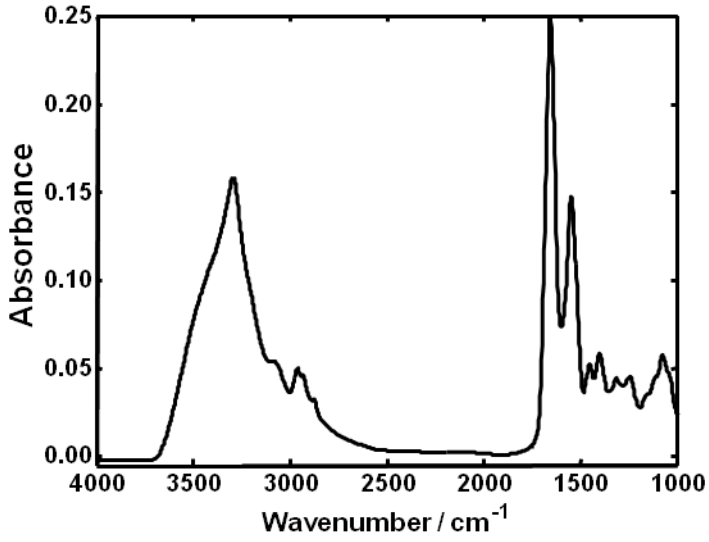

Figure 2.15 IR absorbance spectrum of a thin film of Matrigel measured in transmission mode.

The Matrigel spectrum was used as a guideline for creating a database of peak positions, heights and widths. An additional peak was also added at 1740 cm^{-1} which is invariably present in biological spectra, assigned to the C=O bond in lipids. The wavenumber range of the spectra simulated was 800 to 2500 cm^{-1} at a resolution of 2 cm^{-1}, as this is typically the region of interest for biological IR data.

Two data sets (group 1 and group 2) were created comprising of 25 spectra all based on the Matrigel peak database. A random number generator was used to vary the positions (± 1 cm^{-1}), heights (± 20%) and widths (± 2.5%) of peaks within each spectrum. The second data set was subject to the same random variation as the first but was intentionally given a

higher absorbance by 0.1 at the 1300 cm^{-1} and 1740 cm^{-1} peaks so that the two groups would appear different when analysed using PCA, see Figure 2.10).

2.4 Summary

This chapter has stated the methods & resources used to investigate the scattering of infrared light during FTIR measurements of biological samples. The instrumentation used to measure FTIR data has been discussed, followed by a description of elementary mathematics used to describe the scattering of the light. Finally, computational resources have been discussed that were used as the platform to create a correction algorithm for the scattering.

3. Reflection contributions to spectral distortions

In this chapter the contribution first investigated towards spectral distortions is presented, and is related to the reflection of IR light from the sample surface. A brief outline of the physics of reflection is stated followed by the experiments conducted and results deduced.

3.1. Reflection

The percentage of reflection of light from a surface is defined by the real part of the complex refractive index (hereafter called the real refractive index), n, the wavelength of light and the angle of incidence, ϑ_i. The phenomenon of refraction occurs due to the slowing down of light in a medium, as can commonly be seen by observing a straw in a glass of water which seemingly appears to be bent in shape upon entry into the medium. The real refractive index is defined as the speed of light in a vacuum divided by its speed in the material[104].

The real refractive index of air can be approximated to a value of 1, its actual value is negligibly larger than 1, which simplifies some calculations. The reflection and transmission factors of light going from one medium to another, such as air to water can be calculated using the Fresnel equations[104].

3.1.1 Fresnel equations

An incident beam of light approaching a surface at angle ϑ_i, is partially reflected at the surface at angle, ϑ_r, and partially transmitted into the sample at angle ϑ_t, as Figure 3.1 illustrates. The real refractive indices of air and the medium are denoted n_1 and n_2 respectively[104].

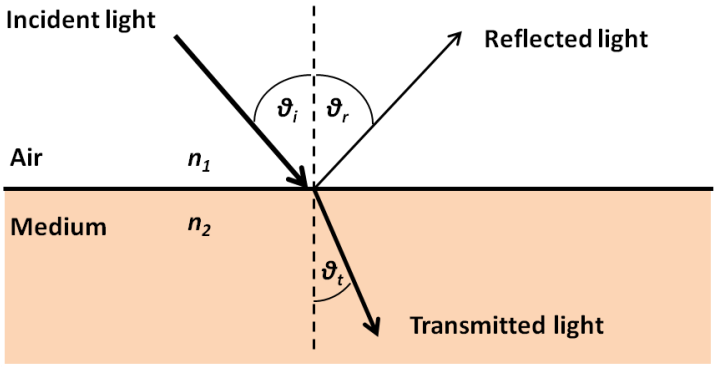

Figure 3.1 Reflection and transmission of light at a surface.

The reflectivity for s-polarised (R_s) and p-polarised (R_p) can be calculated using the following relations:

$$R_s = \left[\frac{\sin(\theta_t - \theta_i)}{\sin(\theta_t + \theta_i)}\right]^2 = \left[\frac{n_1(\cos\theta_i) - n_2(\cos\theta_t)}{n_1(\cos\theta_i) + n_2(\cos\theta_t)}\right]^2 \quad (9)$$

$$R_p = \left[\frac{\tan(\theta_t - \theta_i)}{\tan(\theta_t + \theta_i)}\right]^2 = \left[\frac{n_1(\cos\theta_t) - n_2(\cos\theta_i)}{n_1(\cos\theta_t) + n_2(\cos\theta_i)}\right]^2 \quad (10)$$

When photons are at near normal incidence, the polarisation becomes negligible and the equation can be simplified to give a general term for the reflection, *R*. In the interest of simplification, it will be assumed that light from the microscope optics is incident upon the sample at normal incidence. This is not the case in reality and there is a large range of angles of the light as it is focused on the sample in a cone-like manner, however near normal incidence approximation suffices in due computations.

$$R = \left[\frac{n_1 - n_2}{n_1 + n_2}\right]^2 \quad (11)$$

In theory, the factor of reflected and transmitted photons should sum to equal one, ignoring diffraction and absorption processes. This allows an expression for transmission to be calculated:

$$T = 1 - R = \frac{4n_1 n_2}{(n_1 + n_2)^2} \quad (12)$$

Equations (11) and (12) allow the reflection and transmission factors for an incident beam of photons to be calculated knowing only the real refractive indices of the two concerned materials at a near normal incidence. This will aid in the understanding of line shape distortions caused in IR spectra explained later.

3.1.1. Imaginary refractive index

The imaginary refractive index, *k*, commonly referred to as the absorptive index governs a material's absorption properties as a function of wavelength. In IR spectroscopy, absorption bands are present as Voigt profiles (convolution of a Lorentzian and Gaussian line shapes) with peak maximum at a specific wavenumber. The position and relative heights of these peaks is determined by the imaginary refractive index, if one were to compute an infrared spectrum with corresponding data of the *k* spectrum, the two would appear very similar in shape.

The relation of *k* with the absorption coefficient, α is

$$\alpha = 4\pi\tilde{\nu}k \quad (13)$$

The absorption spectrum measured is simply this absorption coefficient at each wavenumber, multiplied by the effective optical path length traversed by the IR light.

$$A = \alpha l \tag{14}$$

where *l* = the path length.

3.1.2. Relation of the real and imaginary refractive index

The real and imaginary refractive index are not independent variables, they coexist and are related to one another as graphically depicted in Figure 3.2.

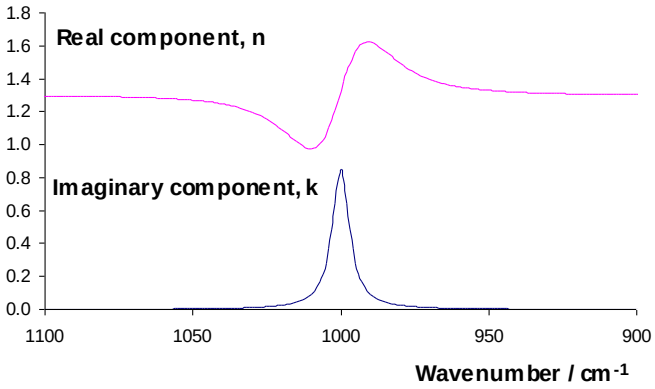

Figure 3.2 Plot showing the variation of the *n* and *k* spectra for a Lorentzian band shape.

In Figure 3.2 the *k* spectrum was computed using the equation of a Lorentzian line shape to create a theoretical peak for illustration purposes. The *n* spectrum of *k* is shown by the red trace, and was computed using the Kramers-Kronig transformation[1]:

$$n(\tilde{\nu}) = n_{\infty} + \frac{2}{\pi}\,\mathbf{P}\int_{0}^{\infty} \frac{sk(s)}{s^2 - \tilde{\nu}^2}\,ds \tag{15}$$

where n_∞ is the real refractive index at infinity, **P** is the Cauchy principal of improper integrals, required in this case due to the singularity created due to a division by zero. *s* is the subject of integration.

An inverse transform also exists where *k* can be computed from *n*:

$$k(\tilde{\nu}) = \frac{2\tilde{\nu}}{\pi} + \mathbf{P}\int_0^\infty \frac{n(s)}{s^2 - \tilde{\nu}^2} ds \tag{16}$$

The Kramers-Kronig relations provide a powerful tool for allowing the computation of *n* and *k* given only the other. The shape of a the real refractive index spectrum in Figure 3.2 looks similar to the shape of a first derivative of an absorption band, however the values are not the same. The significance of *n* is that it defines the reflection properties of a material but it is not constant across a spectrum; at an absorption band, it undergoes a sharp change in value (a phenomenon called anomalous dispersion).

3.2 Reflection contributions

During an IR spectroscopy measurement, the incident light will be reflected according to the mathematics of the Fresnel equations. Transmission and transflection experiments involve illumination of the sample from an upwards direction thus resulting in a reflected signal towards to optics of the beam origin, see Figure 3.3.

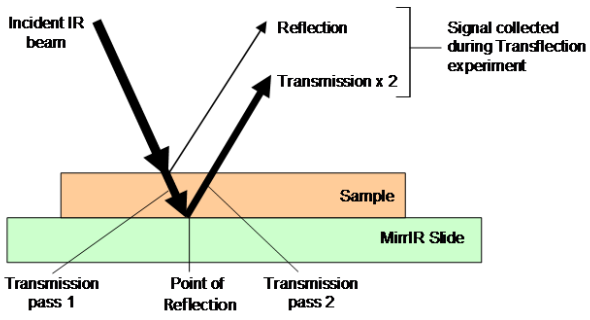

Figure 3.3 Schematic showing the signals involved during a transflection IR experiment[105].

Figure 3.3 shows a schematic of a transflection experiment including the paths and origins of the transmission and reflection signals. In a transflection experiment, both the reflected, *R*, and transmitted signal, *T*, are collected[105]. Hereafter, this transflection signal will be denoted as I_{Tf}.

$$I_{Tf} = 2T + R \tag{17}$$

This equation suggests that the transmission signal of samples with low optical density will be dominant making the effect of the reflection signal negligible. However when the transmission signal becomes weaker, tending towards zero the reflection signal will have a greater influence in the transflection signal. The transmission signal will vary with optical density of the sample, however the reflection signal arises from the surface of the sample and is ever present irrelevant of the sample thickness / optical density.

To demonstrate the line shapes possible from different weightings of the transmission and reflection signals in a transflection signal, a Lorentzian peak was computed, and a Kramers-Kronig transformation computed to acquire a theoretical reflection signal, as Figure 3.4 shows.

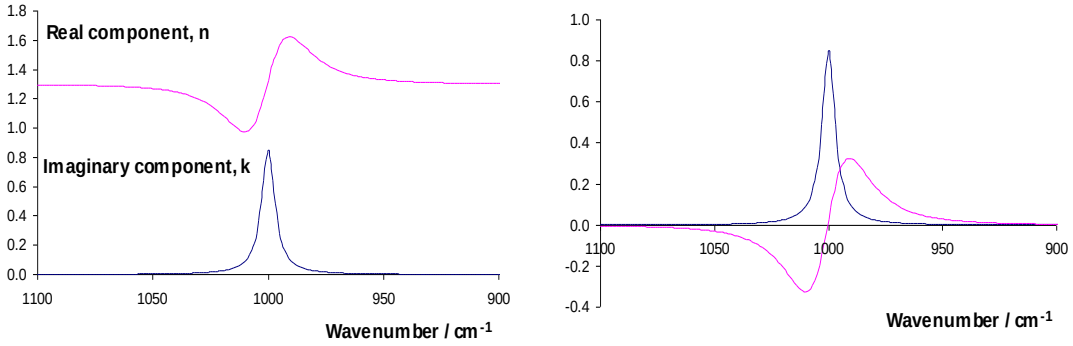

Figure 3.4 A Lorentzian peak and corresponding reflection spectrum.

By adding together these theoretical transmission and reflection profiles in different weightings, a series of interesting line shapes can be observed. Although somewhat numerically arbitrary, the line shapes produced serve a useful purpose in that they reveal a contribution to the 'dispersion artefact' line shape[105].

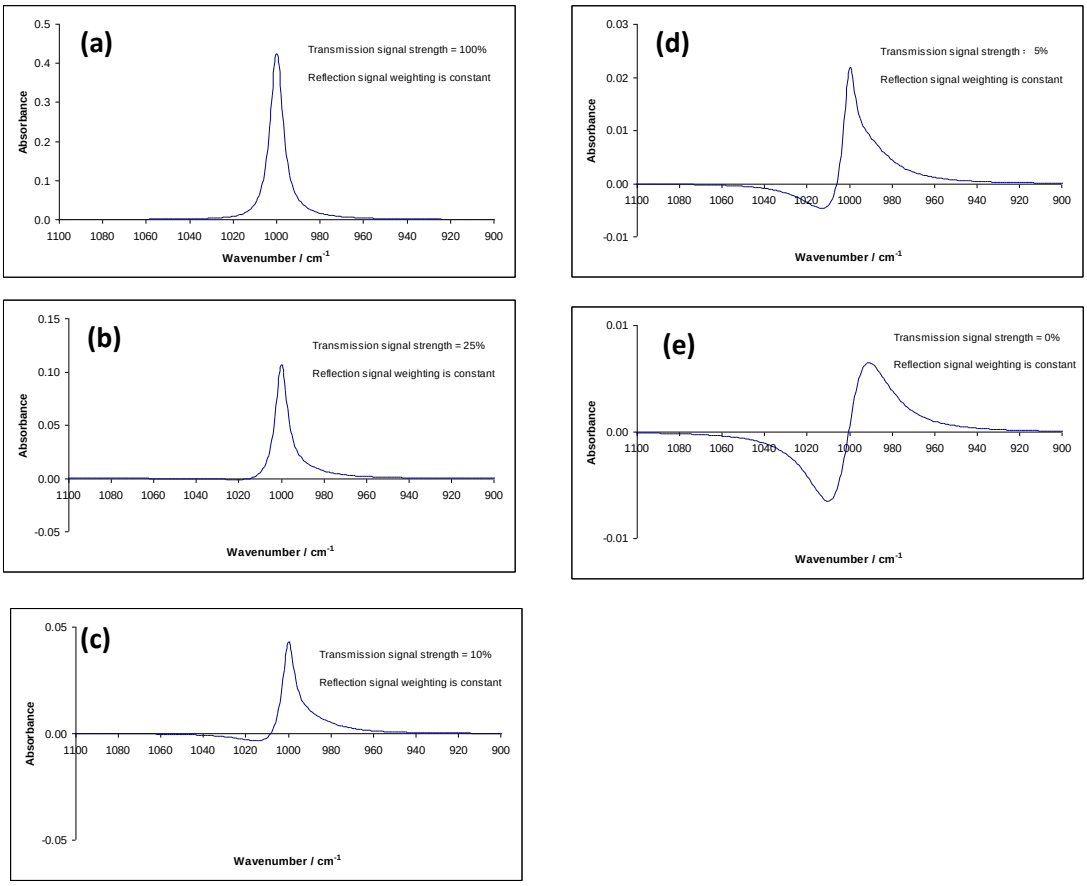

Figure 3.5 Resultant line shapes from different weightings of reflection and transmission signals for a theoretical peak.

Figure 3.5 shows the resultant line shapes produced when a transmission and reflection signal are added in different weightings. The magnitude of the reflection signal is kept constant in each plot as its magnitude in experiments should remain the same regardless of the transmission component. The magnitude of the transmission component is reduced from 100% to 0% to simulate the effect of an increasing optical density / sample thickness, finally reaching extinction at 0%. It can be seen that when the transmission signal is 100%,

the reflection signal is not observable in the spectrum. As the transmission signal is reduced, the reflection signal becomes increasingly evident, and the derivative-like line shape of the 'dispersion artefact' becomes ever more prominent until a pure reflection signal is observed[105].

3.3. Measurement of reflection contributions

The contribution of the reflection signal during transflection experiments was confirmed by using a sample comprising of prostate cancer cells (PC-3 cell line) cultured onto a transmission window, calcium Fluoride in this case. Using a high brightness source at the SOLEIL synchrotron, a number of experiments were conducted to measure the reflection contributions during IR spectroscopic measurements.

In the previous section it was stated that during a transflection experiment, the contribution of the transmission signal is of a greater magnitude than that of the surface reflection. By measuring a cell on a transmission window, the transmission component can be assumed to be of zero magnitude during a transflection experiment, meaning that the only signal measured was that of the surface reflection. Using an aperture size of 10 x 10 μm^2 and spacing of 10 µm, a line map was taken across a PC-3 cell to measure the reflection contributions from different areas of sample, including cytoplasm and the nucleus.

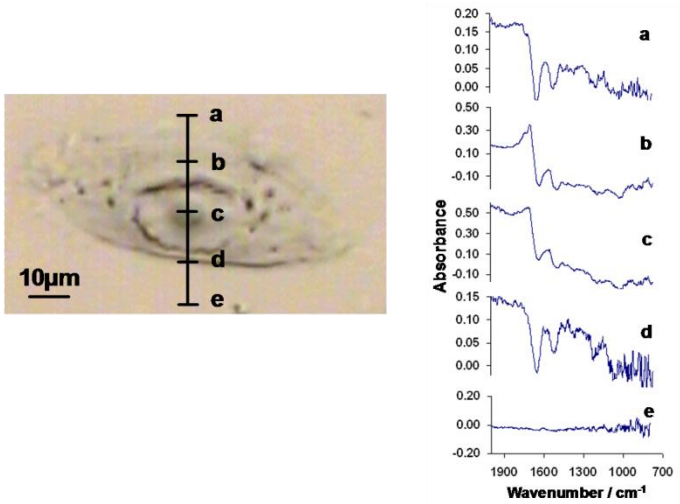

Figure 3.6 Left: Optical image of PC-3 cell on a CaF_2 substrate. Right: Spectra of the cell at the points indicated in (a) expressed as absorbance[105].

Figure 3.6 shows a PC-3 cell on a CaF_2 slide with reflection spectra from points along a line map. It is immediately evident that even though this cell is placed on a transmission slide, measurement in transflection mode results in a signal containing biochemical information. These spectra are surface reflection spectra and can be clearly seen as the transmission signal is passing through the sample and not being collected by the collection options. The characteristic derivative-like line shapes are immediately obvious, especially at the position of the amide I band which is usually at approximately 1655 cm^{-1}. This line shape is similar to the 'dispersion artefact' frequently observed during transflection experiments of single cells, and tissue.

Spectrum (a) closely resembles an inverted IR biomedical sample with the amide I and II bands in approximately the correct relative magnitudes. Spectrum (c), positioned at the nucleus looks significantly more distorted due to the combination of a reflection signal and Mie scattering effects. The nucleus had previously been considered as a "black hole" which was too dense to pass light through making spectroscopy difficult [73]. This experiment has shown that using a refection based methodology, a signal can be acquired. The penetration

depth of the electromagnetic field incident upon a surface is approximately that of the wavelength of the field, meaning that at a wavenumber of 1240 cm^{-1}, corresponding to DNA bands, a depth of ~8 µm is penetrated.

The reflection, rather than being an unwanted contamination of IR experiments could be used to calculate the absorption spectrum using the Kramers-Kronig transform, which would be particularly convenient for thick samples where light cannot be passed through.

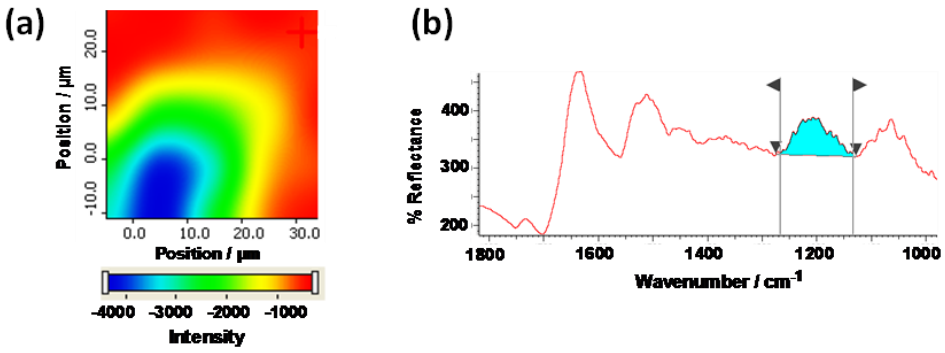

Figure 3.7 (a) Map of a PC-3 cell on CaF_2 based on the integrated band at 1240 cm^{-1} shown in (b). Note that blue indicates highest reflectance, red the lowest[105].

Figure 3.7 shows an IR mapping image using a 10 x 10 μm^2 aperture, and a 10 µm step size using a synchrotron source. A "chemical map" of the 1240 cm^{-1} peak attributed to DNA phosphate groups has been produced. This image shows the nucleus contains a greater signal from these DNA bands than the cytoplasm as would be expected. The colour scheme is negative due to the processing of the reflection spectra not being in conventional absorption; blue represents greater intensity and red lower.

3.4. Conclusion

In this chapter a theoretical simulation of the reflection contributions in transflection experiments has been demonstrated, showing that derivative-like line shapes can occur

when the transmission signal becomes weak. The transmission signal during a transflection experiment can become weak when the sample thickness is great, and when the optical density of the material is high resulting in the extinction of the light.

The reflection contributions were validated by measuring the transflection signal of a single biological cell on a transmission window, where the transmission component collected was zero. This confirmed that the collection optics do indeed collect the reflection signal, and that this is a contribution factor to the so called 'dispersion artefact'[105].

During a transmission experiment, the reflection signal is not collected by the optics, however a loss of light will have occurred before the light propagates through the sample. The 'dispersion artefact' is often present in transmission experiments when the reflection signal could not be playing a major role. This can be explained due to scattering and is the subject of the following chapters.

4. Resonant Mie Scattering (RMieS)

In this chapter an introduction into light scattering and namely, Mie scattering is presented. Experiments to observe the effects of Mie scattering during infrared spectroscopic measurements are presented using a model system of known size polymer micro-spheres. Experimental results are compared with theoretical simulations to show high agreement[106].

4.1. Mie scattering

In 1908 Gustav Mie wrote his paper on the subject of light scattering from turbid media, such as milk which is an emulsion of water, lipids and proteins[79]. When light is incident upon a scattering particle, its path can be deviated which has major implications for IR spectroscopy. The usual aim of an IR spectroscopic measurement is to measure the absorption properties of the sample relative to a background of zero absorption. In an ideal experiment, any IR light that does not reach the detector has been absorbed by the sample. However due to the nature of biomedical samples, namely single cells and tissue the morphology is far from the ideal “flat surface”, see Figure 1.4.

Figure 1.4 shows an ideal IR measurement where light not reaching the detector was absorbed by the sample, thus meaning that the computed absorbance spectrum (negative base 10 logarithm of the transmission profile) is representative of only sample chemistry. Figure 1.4(b) illustrates the concept of scattering when the IR light is incident on a particle, e.g. a single cell resulting in deviation of photon paths outside of the detector. These photons are not detected by the detector but are still considered to have been absorbed upon the computation of the absorbance profile, but this is not the case as they were simply lost to the system. This phenomenon results in additional features in the

absorbance spectrum that are not related to the chemistry of the sample, but rather the morphology.

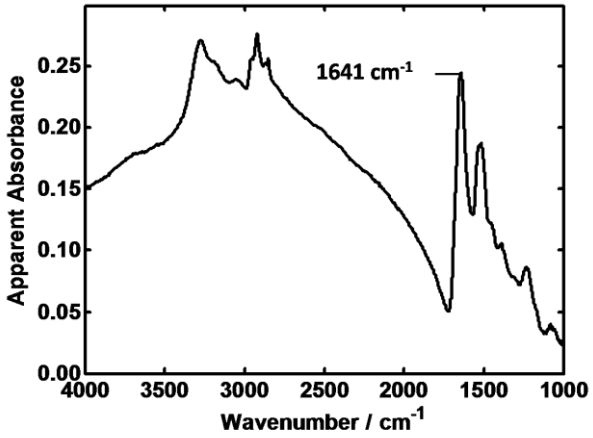

Figure 4.1 IR transmission spectrum of a single PC-3 cell.

Figure 4.1 shows a spectrum of a single PC-3 cell which exhibits the characteristic effects of scattering. These effects include the broad baseline upon which the spectral features are superimposed, and the shifting of peaks, illustrated in this case by the amide I peak at 1641 cm^{-1}, which is shifted to a lower wavenumber from its characteristic frequency of 1655 cm^{-1}. When spectra exhibit scattering features, the true biochemistry of the sample is distorted, and further analysis becomes difficult.

The mathematics governing the quantity of photons lost to the system at each wavelength due to the scattering particle are vastly complicated, and hence a number of initial simplifications must be made in order to understand the underlying physics. In 1957, van de Hulst published an approximation equation for the efficiency with which each wavelength of light is scattered (the Mie scattering efficiency, *Q*) given the real refractive index and radius of the particle (assumed to be a dielectric sphere in this case).

$$Q = 2 - \frac{4}{\rho}\sin\rho + \frac{4}{\rho^2}(1 - \cos\rho) \quad (18)$$

where

$$\rho = 2\pi d(n-1)\tilde{\nu} \quad (19)$$

Where *d* = diameter of the scattering particle, and *n* = the ratio of the real refractive indices of the particle to the surrounding medium. In the case of IR measurements of biological samples, the surrounding medium has a real refractive index of approximately 1 meaning that *n* can simply be assumed to be the particle refractive index. These equations are the starting point to aid in understanding the fundamental principles behind the scattering effects in IR spectra of single cells and tissue.

4.2 Synchrotron FTIR measurements of isolated poly(methyl methacrylate) (PMMA) microspheres

As a model to gain insight into the characteristics of IR light incident upon a scattering particle a simple model involving microspheres of PMMA of various sizes were employed. Due to the homogenous chemistry, and spherical shape, a much simpler model is obtained compared to a single biological cell which has inhomogeneous chemistry and non-spherical shape making the underlying mathematics more complicated[106].

4.2.1 Sample preparation

The PMMA spheres in suspension (Microspheres-Nanospheres, Cold Springs, NY, USA) were obtained in three different sizes, 5.5, 10.8, 15.7 μm in diameter. The spheres were re-suspended in deionised water twice to reduce any chance of salt formation after deposition. For each sphere size, 5 μL of suspension was separately re-suspended in 100 μL of deionised water to reduce the density of the spheres on the slide thus increasing the possibility of depositing isolated spheres. 20 μL of each sphere/water suspension was deposited onto 0.5 mm thick CaF_2 slides (Photox Optical Systems, Sheffield, UK) and allowed to air-dry for several hours. The PMMA reference material was supplied as a powder by Polymer Laboratories UK. A solution of PMMA in toluene was prepared and similarly deposited on to a CaF_2 slide and allowed to dry forming a thin PMMA film. The film thickness was not specifically determined but was sufficient to give a maximum absorbance value close to 1. The film was visibly transparent with no evidence of physical inhomogeneity.

Spectra of the isolated PMMA spheres were measured using a synchrotron source at the SOLEIL SMIS beamline[107]. A synchrotron source was required due to small sample sizes, in particular the 5.5 μm spheres which would be difficult to measure using a conventional bench top source. The spectra were recorded at 4 cm^{-1} resolution, and the number of co-added scans were selected depending on the sample experiment performed. The FTIR microscope used was a Nicolet Continuum XL coupled to a Nicolet Nexus spectrometer.

4.2.2. Infrared spectra of PMMA

The IR absorbance spectrum (measured in transmission mode) of a thin PMMA film on a CaF_2 window is shown in Figure 4.2(a). The spectrum shows that a number of strong absorption bands on a flat baseline and is consistent with published spectra of PMMA, see

table 1 [106]. The most prominent bands observed are the CH_2/CH_3 stretching modes at 2991 cm^{-1}, 2950 cm^{-1} and 2843 cm^{-1}, C=O (1731 cm^{-1}), the asymmetric C-C-O stretch (1276 cm^{-1} and 1240 cm^{-1}), and the asymmetric C-O-C stretch (1193 cm^{-1} and 1150 cm^{-1}). All the observed bands are within 5 cm^{-1} of published reference data, and given possible differences in the exact molecular weight of the polymer and tacticity, this variation is to be expected. The large isolated carbonyl (C=O) band at 1731 cm^{-1} serves as a crude model of the amide I band present in cells and is uncomplicated by the presence of an amide II band[106].

Figure 4.2b(i) shows an optical image of 5.5 µm PMMA spheres on a CaF_2 window. On this area of the surface, the spheres are reasonably close packed. Figure 4.2b(ii) shows the infrared spectrum of a single sphere incorporated into a close packed layer of spheres imaged with a 6 x 6 μm^2 aperture. In this figure, and subsequent figures, the axis is labelled 'apparent absorbance' presenting $\log_{10}(I_0 / \langle I \rangle)$ where <I> is the average transmitted intensity reaching the detector which includes light scattering from and going round the sphere. The spectrum shows a slope in the baseline with some low amplitude oscillations similar to those of the PMMA film on CaF_2, but the shape of the carbonyl band has become noticeably asymmetric and the band position has shifted to a lower wavenumber, to 1725 cm^{-1} (~- 6 cm^{-1}). One can notice the appearance of noise in spectrum below ~1100 cm^{-1}, which is due to diffraction limitation when using a small aperture.

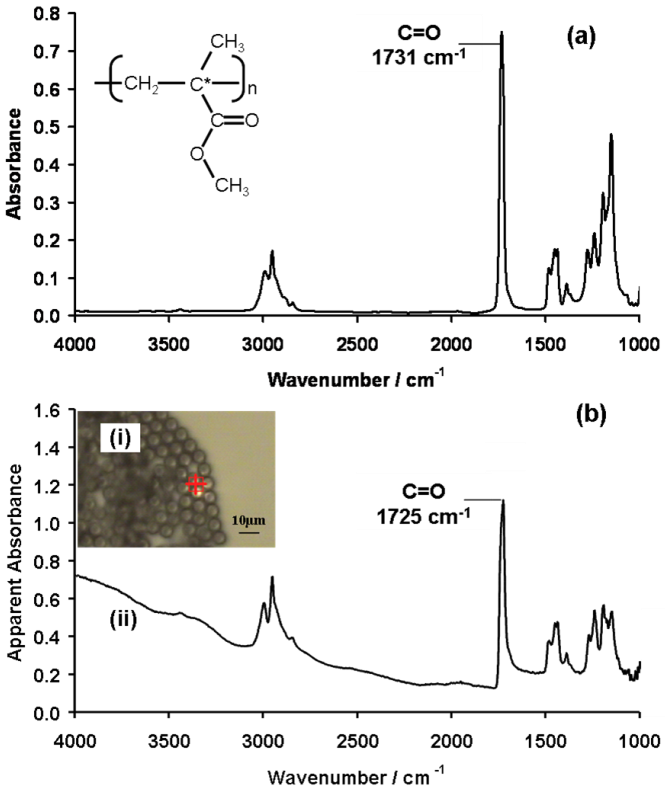

Figure 4.2 (a) The infrared transmission spectrum of a thin film of PMMA deposition on CaF_2; (b)(i) an optical image of 5.5 µm diameter PMMA microspheres deposited on CaF_2, and (b)(ii) the infrared transmission spectrum taken from a region where the PMMA spheres are close packed as indicated by the red box in (i).

Table 1. Assignments of vibrational bands observed in spectra of PMMA in a KBr Pellet, a PMMA film on a CaF_2 plate, a close packed (CP) layer of 5.5 µm PMMA spheres, (I) 5.5 µm, 10.8 µm and 15.7 µm PMMA spheres on CaF_2. All numbers are wavenumber (reciprocal wavelength) values and have units of cm^{-1}.

Vibration	KBr pellet	Thin film on CaF_2	5.5 µm spheres (CP)	5.5 µm spheres (I)	10.8 µm spheres (I)	15.7 µm spheres (I)
ν(CH) of O–CH_3	2996	2991	2995	3007	2988	2998
ν(CH) of O–CH_3	2950	2951	2950	2959	2944	2954
ν(CH) of O–CH_3	2844	2843	2842	2846	2841	2844
ν(C=O)	1732	1731	1725	1707	1767, 1729	1754, 1708
$\delta(CH_2)$	1485	1480	1482	1479	1505	1491
δ_a(CH) of α-CH_3	1448	1449	1448	—	1468	1453
δ_s(CH) of OCH_3	1436	1435	1435	1431	—	—
δ_s(CH) of α-CH_3	1387, 1367	1386,	1387	1386	1406, 1351	1390
ν_s(C–C–O)	1273, 1242	1276, 1240	1270, 1240	1267, 1233	1262, 1224	1289, 1257
ν_a(C–O–C)	1193, 1150	1193, 1150	1191, 1147	1184, 1136	1183	1212
ν(C–C)	1063	1068	—	—	1085, 1048	—

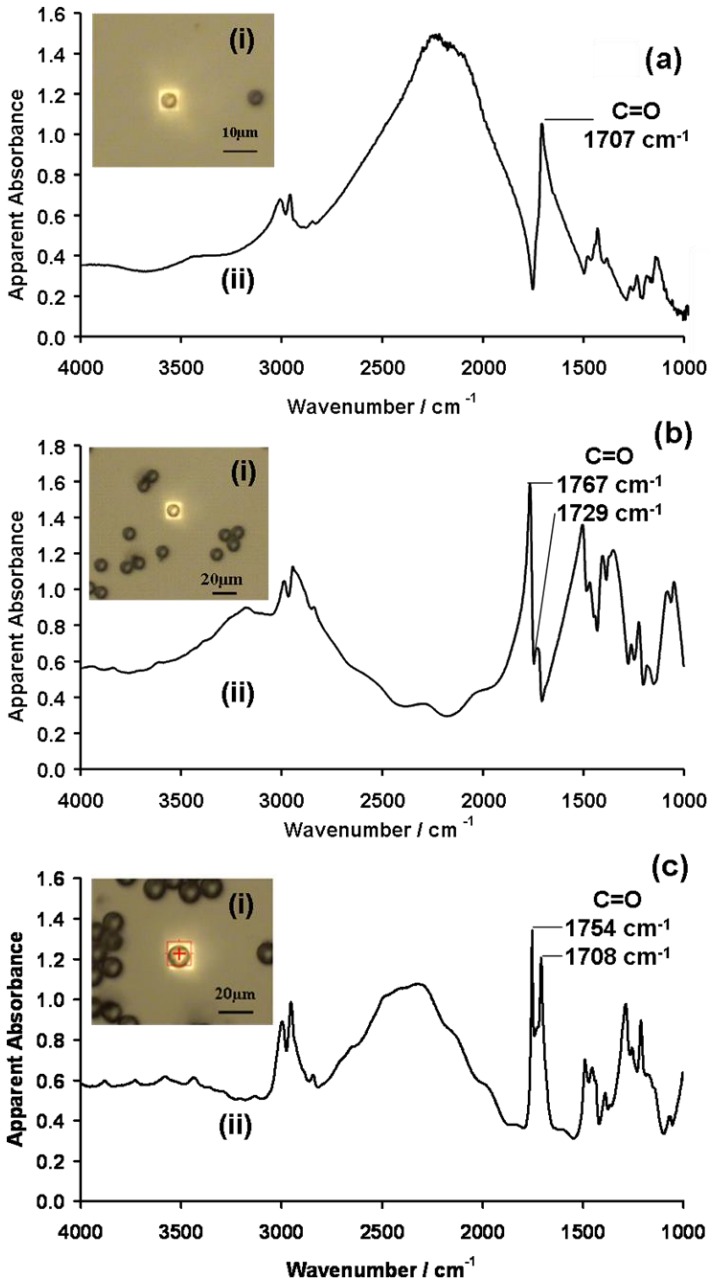

Figure 4.3 The optical images (i) and the infrared transmission spectra of (ii) isolated (a) 5.5 µm, (b) 10.8 µm, (c) 15.7 µm diameter PMMA microspheres deposited on CaF_2.

Figure 4.3a(i) shows an optical image of an isolated 5.5 µm PMMA sphere on the sample, illuminated by light from the 6 x 6 μm^2 aperture. Isolated spheres were selected since in single cell cytology it is isolated cells rather than clumps of cells that are preferentially analysed. Figure 4.3a(ii) shows the infrared spectrum obtained from the sphere and repeat

measurements on several spheres show that it is representative of isolated single spheres of this size.

It is immediately obvious that despite having the same chemical composition, the spectrum does not resemble the absorbance spectrum of the PMMA film shown in Figure 4.2(a), or indeed the absorbance spectrum of the same spheres in a closely packed layer, Figure 4.3b(ii). The spectrum of the isolated sphere clearly shows strong Mie scattering, indicated by the pronounced oscillation of the baseline. In addition, the presence of the very strong dispersion artefact is clearly indicated by the first derivative-like line shape in the region of the carbonyl stretching mode which gives rise to a positive peak at 1707 cm^{-1} (instead of 1731 cm^{-1}). Note that the derivative-like line shape of the dispersion artefact has shifted the position of the positive maximum of this band by 24 cm^{-1} and has significantly reduced the apparent relative peak intensity. It is also evident that the dispersion artefact is not limited to the carbonyl band but is influencing, albeit to a lesser extent, weaker bands, perhaps also due to the presence of neighbouring peaks, does not mean that bands are unaffected. There are some changes in intensity in the CH stretching region and, as indicated in table 1, band positions have moved significantly. In contrast to the carbonyl band, however, all of the prominent CH stretching bands have moved to a *higher* wavenumber. This is an important point which is discussed later.

Figure 4.3b(i) shows the optical image of an isolated 10.8 µm diameter PMMA sphere illuminated by a 12 x 12 $µm^2$ aperture. Figure 4.3b(ii) shows the IR spectrum of the sphere and again several measurements show that it is typical of isolated spheres of that size. The spectrum is strikingly different from either the PMMA film or the 5.5 µm diameter isolated sphere. The strong oscillation in the baseline is still evident but the amplitude and period of oscillation have changed. In addition to the broad oscillations due to Mie scattering, a higher frequency oscillation in the baseline is observed which is more likely to be due to an

interference fringing effect and will not be considered further. Importantly, the dispersion artefact, previously present by a negative excursion on the high frequency side of the carbonyl band in Figure 4.3a(ii), has become a positive excursion giving rise to an apparent peak at 1729 cm^{-1}. In addition to the carbonyl band there are significant changes in intensity and position of the lower frequency bands as can be seen in table 1.

Figure 4.3c(i) shows the optical image of an isolated 15.7 μm diameter PMMA sphere illuminated by a 16 x 16 μm^2 aperture. Figure 4.3c(ii) shows the IR spectrum of the sphere and yet again it is strikingly different from the spectra of the other spheres and the PMMA film. Although the CH stretching region looks relatively undistorted, the carbonyl band is now split into two strong bands of similar intensity at 1754 cm^{-1} and 1708 cm^{-1}, as shown in table 1.

It is clear from Figure 4.2 and Figure 4.3 that despite having identical chemical composition, the PMMA spheres give rise to IR spectra that are dramatically different with apparent intensity, position, and numbers of peaks all varying. This variation is due exclusively to the different physical size of the spheres. Both strong Mie scattering and strong dispersion artefacts are evident in all the spectra of the isolated spheres.

4.3. Computational modelling of scattering extinction

If the same pure absorption spectrum is to be recovered from the above spectra, it is essential to understand the phenomenon that is causing the spectral distortions. From eqn. 18, the scattering efficiency *Q*, is dependent on both wavenumber and sphere diameter, and this leads to a pseudo-periodic oscillation of *Q* for a given refractive index of the sphere and a given sphere diameter. Knowing the size of the sphere, the scattering can be modelled by comparison with the experimental data.

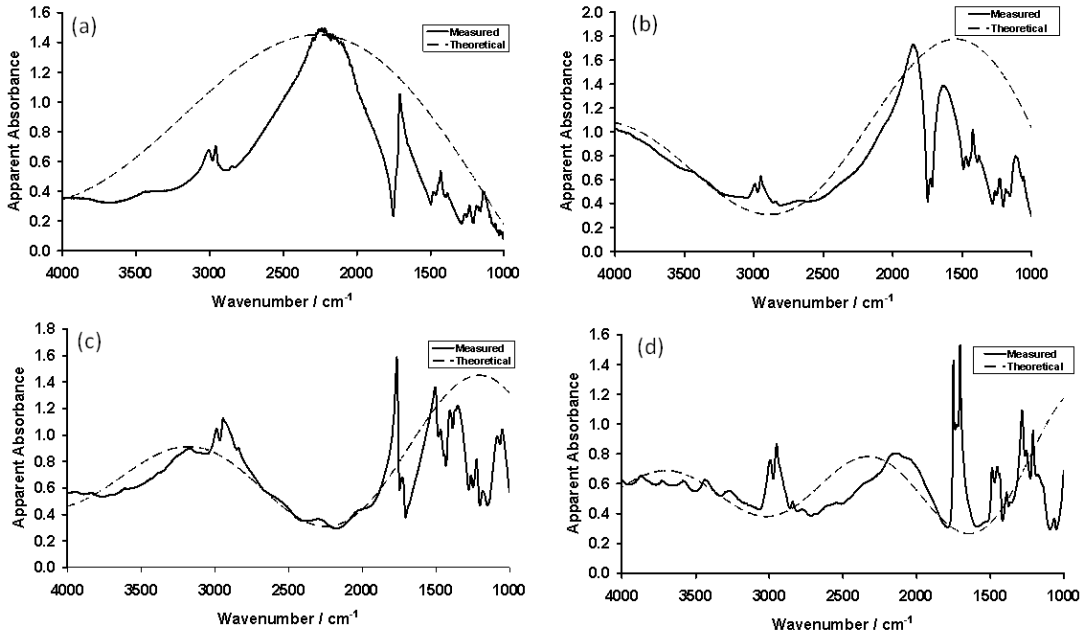

Figure 4.4 Infrared spectra of isolated PMMA spheres fitted with a single Mie scattering curve calculated using the van de Hulst equations: (a) 5.5 µm, *n* = 1.26, (b) 7.0 µm, *n* = 1.3, (c) 10.8 µm, *n* = 1.25, (d) 15.7 µm, *n* = 1.24.

Figure 4.4 shows spectra from four different size spheres with best fit scattering efficiency *Q* curve on the same plot. A 7 µm diameter PMMA sphere was found amongst the distribution of the 5.5 µm spheres and an IR spectrum was measured to enable modelling of another sphere size in addition to the previous three sizes. *Q* is a function of wavenumber, sphere diameter (*d*) and real refractive index ratio (n_1/n_2). When fitting the data, the value of *d* for each sphere, which is known, is kept constant, as is n_2, the refractive index of air taken as 1. The parameter n_1 is calculated using

$$n_1 = a + b(n(\tilde{\nu}) - n_\infty) \tag{20}$$

Where *a* and *b* are constants that are varied to fit the data. Note that ideally, n_1 should be the same for all the spheres, but due to the non-ideal experimental geometry, adjustment of these parameters is necessary.

The experiment is not perfect since the approximation equation for Mie scattering is for a sphere illuminated evenly by a parallel beam, whereas in this situation the sphere is on a surface and is illuminated by a beam focused using a high numerical aperture. However, the trend in the fitted data is clear to see. As the size of the sphere gets larger, the periodicity of *Q* decreases.

Although the broad oscillations in the baselines are essentially accounted for, the dispersion artefacts are still present and it is very difficult to reconcile the spectra as being derived from the same chemical compound even if the spectra were corrected for Mie scattering.

The problem is that Mie scattering is strongly dependent on the refractive index, *n*. If the dielectric sphere is a non-absorbing medium, as described in the original Mie theory, it is reasonable to assume a fixed constant value of *n* as a function of wavenumber. However, at an absorption resonance the real and imaginary parts of the refractive index are linked by the Kramers-Kronig relation.

Given the wavenumber dependent variation of the real refractive index *n* of a Material in the region of an absorption band, the scattering efficiency *Q* will change rapidly also. Figure 4.5(a) shows the variation of the scattering efficiency curves as calculated from eqn 18, over a small range of *n* (0.01 increment, for a 5.5 µm diameter sphere). Figure 4.5(b) shows the region 1820-1640 cm^{-1} expanded for clarity. Superimposed upon these curves is an indication of how *Q* would vary due to the change of value of *n* at an absorption band, indicated by the black dotted line.

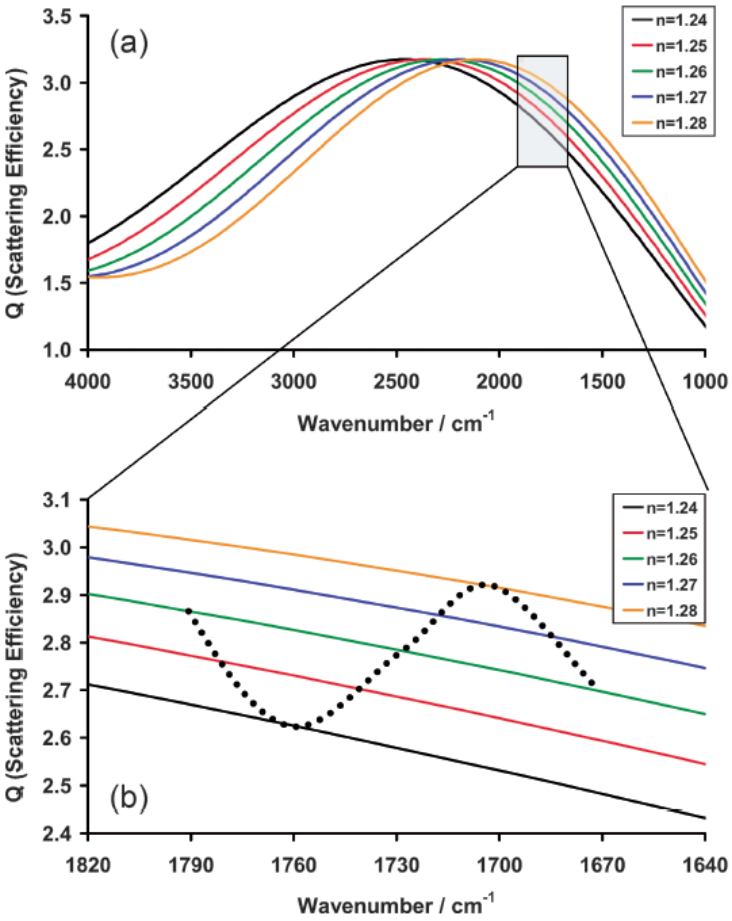

Figure 4.5 (a) Plot of the scattering efficiency, *Q*, as a function of wavenumber for a 5.5 µm diameter PMMA sphere in the region of 4000 - 1000 cm^{-1} for five fixed values of *n*. (b) Expanded view of the curves between 1820 and 1640 cm^{-1}. The filled dots on the line show qualitatively the change in *n* at an absorption band, centred at 1730 cm^{-1}. Due to the order of the scattering curves this would result in derivative-like line shapes as observed for the spectrum of an isolated 5.5 µm diameter sphere.

As can be seen, on the high wavenumber side of an absorption band at 1730 cm^{-1} there is a pronounced reduction in the scattering efficiency followed by a slightly less pronounced increase after 1730 cm^{-1}, similar to the behaviour observed experimentally. It would appear therefore that the dispersion artefact is simply due to the rapidly changing scattering efficiency at the absorption band, *i.e.* it is due to *resonant* Mie scattering.

Figure 4.6(a) shows a series of Mie scattering curves for a 10.8 µm diameter PMMA sphere. Due to the increase in size of the sphere, the period of oscillation in the scattering

efficiency is now shorter. As a result, the gradients of the curves at 1730 cm^{-1} have changed direction compared with those of the 5.5 µm diameter sphere. Figure 4.6(b) shows the expanded region around the absorption band. A superposition of the variation in *n* at the absorption band now gives rise to a positive increase in *Q* on the high wavenumber side of 1730 cm^{-1} with a reduction on the low wavenumber side. This qualitatively accounts for the behaviour observed in Figure 4.6(b). The equivalent curves for the 15.7 µm diameter sphere, not shown, indicate that the carbonyl band occurs at the turning points of the curves (where the sign of the gradient changes) thus is it easy to envisage how a single band could be split as the sequence of the scattering curve reverses at the turning point.

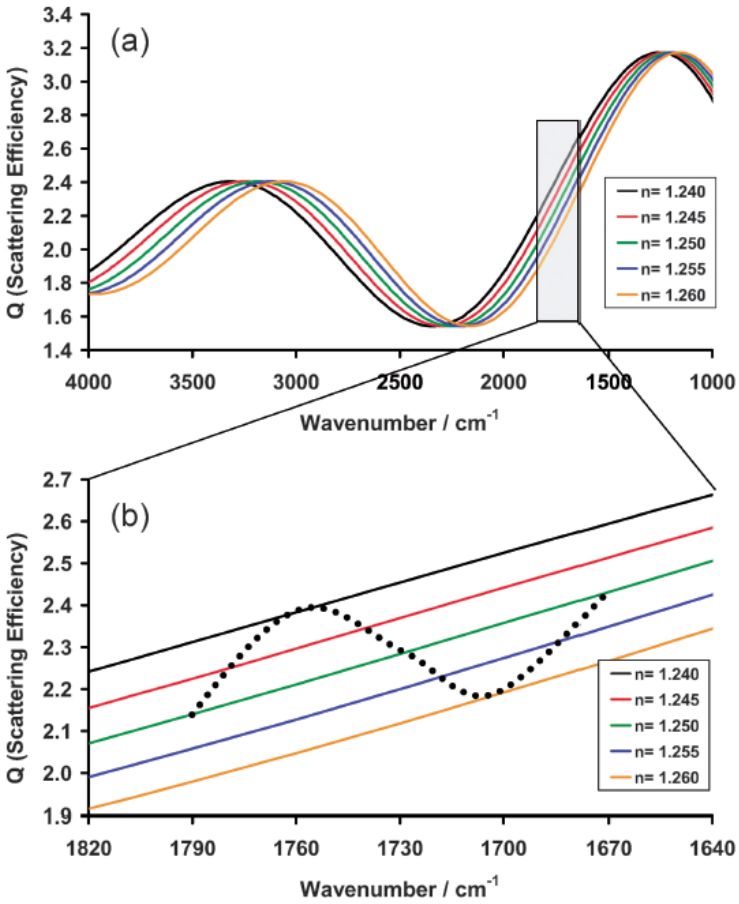

Figure 4.6 (a) Plot of the scattering efficiency, *Q*, as a function of wavenumber for a 10.8 µm PMMA sphere in the region 4000 - 1000 cm^{-1} for five fixed values of *n*. (b) Expanded view of the curves between 1820 and 1640 cm^{-1}. The filled dots on the line show qualitatively the change in *n* at an absorption band centred at 1730 cm^{-1}. Note that because the slope of the scattering curves is positive rather than negative as for the 5.5 µm diameter sphere, the order of the curves is reversed. This would again result in a derivative-like line shape

but in this case there is an increase in *Q* on the high wavenumber side of the 1730 cm^{-1} band just as observed for the spectrum of an isolated 10.8 µm diameter sphere.

The idea of the resonant Mie scattering causing the dispersion artefact can be tested in more depth. From the spectrum of the PMMA film shown in Figure 4.2(a), it is possible to apply the Kramers-Kronig transformation and thus obtain the variation in the real refractive index, n_{PMMA} as a function of wavenumber as shown in Figure 4.7.

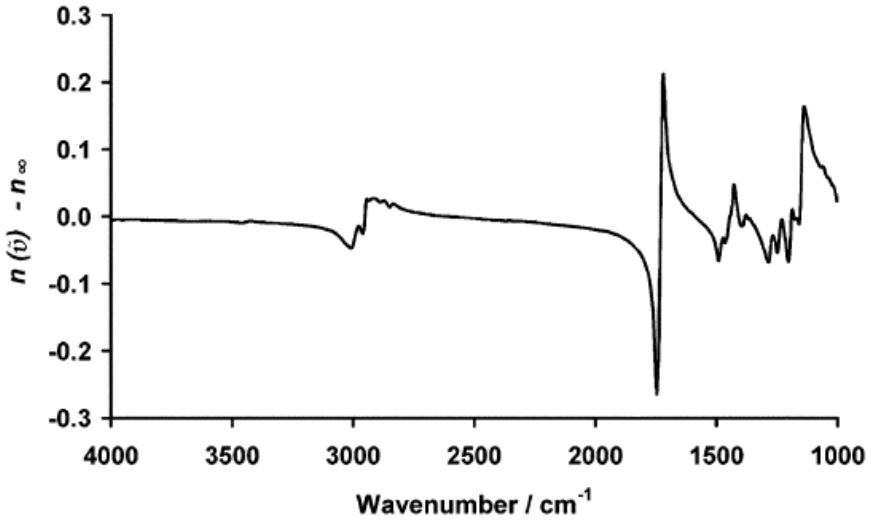

Figure 4.7 The variation of n_{PMMA} - n_∞ as a function of wavenumber calculated using the Kramers-Kronig transformation of the spectrum of PMMA.

This variable value of n_{PMMA} can be fed into the approximate Mie scattering eqn. 18 to produce a scattering efficiency curve that simulates resonant Mie scattering across the entire measured spectral range. This was done by finding the best nominal fit value of $n_\infty + mn_{PMMA}$. The results of such simulations for each of the four sphere sizes are shown in Figure 4.8(a-d). The top curve in each case is the simulated scattering efficiency and the bottom curve is the actual measured infrared spectrum. As can be seen there is remarkably good agreement between the theoretical and experimental curves. It should be noted,

however, that in order to get the best fit to the experimental data it was not possible to use either the same fixed value of n_∞ or a fixed fraction m of n_{PMMA} for all sphere sizes.

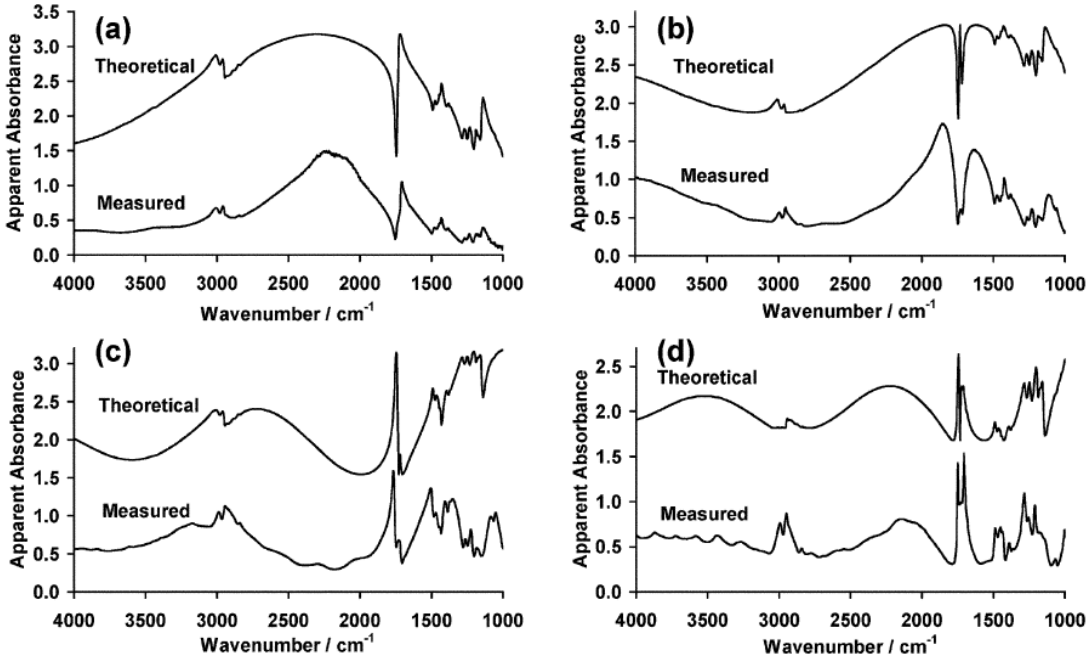

Figure 4.8 Theoretical resonant Mie scattering curves (upper trace, offset for clarity) and experimental spectra (lower trace) of 5.5, 7.0, 10.8 and 15.7 m diameter PMMA spheres. The refractive index values used for the simulated data are: 5.5 µm, n = 1.26 + 0.4 x n_{PMMA}; 7.0 µm, n = 1.28 + 0.6 x n_{PMMA}; 10.8 µm, n = 1.29 + 0.4 x n_{PMMA}; 5.5 µm, n = 1.26 + 0.4 x n_{PMMA}.

4.4 Discussion

It is clear from the IR spectra presented in this chapter that the size, morphology and packing of chemically identical particles dramatically affect the spectra. This phenomenon is frequently observed in the spectra of physically inhomogeneous materials where the length-scale of the inhomogeneities are of the order of the wavelength of the source radiation. In the field of bio-spectroscopy, this will apply equally to biological cells in the same way as for PMMA spheres, and thus it is essential to separate these physical effects from the pure chemical effects that one ultimately needs to recover. Clearly, the confluence of cells is an issue since close packed cells (highly confluent) are likely to give rise to a lower degree of scattering compared with an isolated cell and the difference is

quite dramatic. Previously, Mie scattering and the dispersion artefact have been treated as two separate phenomena. Methodologies have been developed to correct for the former but the latter has been largely ignored. The fact that both the oscillations in the baseline and the dispersion artefact increased dramatically for the isolated 5.5 μm sphere compared with close packed spheres of the same size strongly suggest the theory of a common origin.

Modelling the Mie scattering curves of varying refractive index explains the experimental data. In theory, given the constant chemical composition of the PMMA it should have been possible to obtain these fits using exactly the same values of n_∞ and m but this was not the case and slight adjustments to the parameters had to be made. This may be due to the fact that the approximate Mie scattering equation assumes a non-absorbing dielectric sphere uniformly illuminated with a parallel beam, none of which strictly applies in this case. However, the excellent qualitative agreement between the model curves and the experimental curves demonstrate the principal that resonant Mie scattering is the main cause of the dispersion artefact. Importantly, this resonant Mie scattering can give rise to either a 'positive' or 'negative' dispersion artefact which also accounts for the observation that band positions can be red- or blue-shifted depending on whether the absorption band coincides with a negative or positive gradient of the Mie scattering efficiency curve. This explains why in the spectrum of the 5.5 μm diameter sphere the CH stretching vibrations are blue-shifted whereas the carbonyl band red-shifted compared with the thin film spectrum. This is significant since it distinguishes the effect from the reflective dispersion artefact that may be present in transflection spectra but will follow the Kramers-Kronig dispersion and will not 'invert' as observed here[108].

In IR spectra of single human biological cells, a decrease in the absorption intensity is generally seen on the high wavenumber side of the amide I band. Given the data presented in this paper it can be postulated that it is the smaller components of the cell, *i.e.* nuclei (5

– 6 μm diameter) rather than whole cells (15 – 25 μm diameter), that are responsible for the resonant Mie scattering, in agreement with recent experimental data from isolated cell nuclei[12]. This does not rule out the possibility that, depending on morphology, scattering from the edge of cells may make a significant contribution to the 'dispersion artefact' but as yet this is still being modelled and will be the subject of a future publication.

In general, arrays of biological cells and indeed single cells are far more complex than the PMMA microspheres studied here, in that they are not rigid microspheres of uniform size and density. Both Rayleigh and Mie scattering may contribute to the spectra and thus the system is far more complex. Nevertheless, the simple model system demonstrates that consideration of the resonance effects of the real component of the refractive index and its contribution to the scattering process can account for the spectral distortion of the absorption band in physically inhomogeneous materials such as single cells and collections of cells[106].

5. Signal correction for RMieS

In the previous two chapters the theory behind the distortions in infrared spectra of highly scattering spectra was stated and verified experimentally. In this chapter, the newly understood theory has been used to begin to implement a correction algorithm to remove the scattering effects and recover the pure absorption spectrum of the sample.

5.1. Mie scattering EMSC

As stated earlier, Kohler *et. al.*[10] published an algorithm for correcting non-resonant Mie scattering using a model based system where the raw measured spectrum (Z_{Raw}) is modelled as the linear combination of a first guess (reference) spectrum (Z_{Ref}), a constant offset baseline, a sloping baseline and a number of scattering curves, Q. Stated as simply as possible, this model assumes that Z_{Raw} is a superposition of the pure absorbance spectrum (Z_{Pure}) plus a scatter spectrum ($Z_{Scatter}$), expressed mathematically as (variables accented by arrows denote vectors):

$$\vec{Z}_{Raw} = \vec{Z}_{Pure} + \vec{Z}_{Scatter} \tag{21}$$

The $Z_{Scatter}$ spectrum comprises of the constant and sloping baselines, as well as the variations due to Mie scattering. The mathematics describing the loss of light due to Mie scattering describe situations involving spherical particles, however, biological samples are non-spherical, having considerably more complicated shapes. It is therefore assumed that by summing together a number of scattering curves of different diameter spherical particles (Z_{Sc1}, Z_{Sc2} ... Z_{Scn}), the scattering curve of a non-spherical shape can be reasonably well approximated:

$$\vec{Z}_{Scatter} = c + m\vec{\nu} + \vec{Z}_{Sc_1} + \vec{Z}_{Sc_2} + \ldots \vec{Z}_{Sc_n} \tag{22}$$

$$\vec{Z}_{Scatter} = c + m\vec{\tilde{\nu}} + \sum_{n} \vec{Z}_{Sc_n} \quad (23)$$

The algorithm by Kohler *et. al.*[10] was for the case of non-resonant Mie scattering which successfully removed the broad oscillating baselines from spectra, but often scattering effects remained such as the 'dispersion artefact'. In earlier chapters, it was shown that the 'dispersion artefact' and baselines derived from the same phenomenon of resonant Mie scattering (RMieS). With an understanding of the theory, a preliminary correction algorithm is presented.

5.2 Resonant Mie Scattering EMSC (RMieS-EMSC)

Samples measured using IR spectroscopy are strong absorbers, meaning that they have a non-constant real refractive index. It is this changing of the refractive index at an absorption band coupled with a non-flat sample surface which give rise to resonant Mie scattering. For the purposes of creating a preliminary correction algorithm for RMieS, a number of assumptions have been made. It is assumed that the measured spectrum is the superposition of the pure absorbance spectrum of interest and a resonant Mie scattering curve. This assumption is not the correct mathematical handling of absorbance spectrum and the scatter spectrum, however, in the interest of producing a preliminary correction algorithm it has been used.

Kohler *et. al.*[10] constructed 200 Mie scattering efficiency, *Q*, curves covering a range of possible average refractive index and particle diameter permutations. To correct for RMieS, a different set of *Q* curves need to be constructed taking into consideration the changing refractive index of the samples. The starting point to create a database of *Q* curves of an absorbing biomedical sample, a spectrum of Matrigel has been used, and a Kramers-Kronig transformation applied to calculate the n_{KK} spectrum, see Figure 5.1.

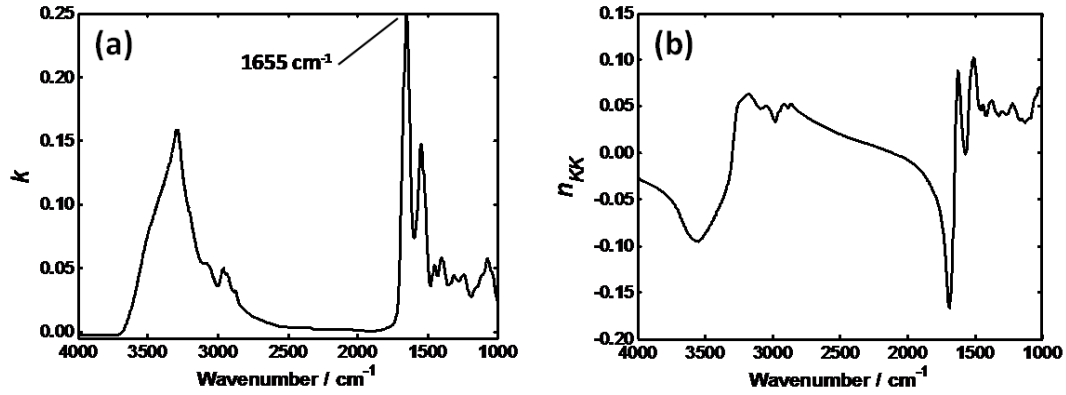

Figure 5.1 (a) Infrared transmission spectrum of Matrigel normalised to maximum absorbance of 0.25. (b) Kramers-Kronig transform of Matrigel from (a).

The exact magnitude of the *k* spectrum is unknown, and in Figure 5.1(a), our reference spectrum, Z_{Ref}, is an absorbance spectrum of Matrigel that has arbitrarily been normalised to maximum absorbance of 0.25. The *k* spectrum is essentially proportional to the absorbance spectrum:

$$Z_{\mathrm{Ref}} \propto k_{Z_{\mathrm{Ref}}} \tag{24}$$

For further convenience, a slight modification to the Kramers-Kronig has been chosen, the 2/π factor before the integral has been omitted and the additive average real refractive index also omitted. Instead of looking at an exact relation (using an equals sign), instead the proportional relation is of interest, the new equation giving n_{KK} is

$$n_{KK}(\tilde{\nu}) \propto \mathbf{P}\int_0^{\infty} \frac{sk(s)}{s^2-\tilde{\nu}^2}\,ds \tag{25}$$

The average refractive index of each sample is again unknown as is the imaginary refractive index. The n_{KK} spectrum is arbitrarily normalised so that its minimum value is -1, the reason for this is explained later. To construct a refractive index for insertion into eqn. (18), two

terms, *a* and *b* need to be defined as the average refractive index and an amplification factor for n_{KK} respectively[75].

$$n = a + bn_{KK} \tag{26}$$

The parameter *b* is required as the Z_{Ref} used is not the correct input for the Kramers-Kronig transform. It is, however, directly proportional and so a scaling parameter can be used to compensate. The refractive index cannot go below a value of 1, and this ensured by carefully controlling the value of *b*, if an average refractive index *a* = 1.3 is used (typical for biological samples), then *b* can range from 0 to 0.3, resulting in the minimum value being 1. The particle diameter *d* is the last parameter which needs to be varied to cover many scattering possibilities giving a total of three parameters: *a*, *b* and *d*.

For each parameter, 10 equidistant values were used between the ranges:

a: 1.1 to 1.5

b: 0 to (a-1)

d: 2 to 20 µm

This results in 1000 permutations thus giving a database containing many scattering curves of a range of feasible physical parameters that biological samples can take. The same approach as Kohler *et. al.*[10], this database of 1000 scattering curves is compressed using principal component analysis (PCA), resulting in a small number of curves (7 loadings spectra in this case) that can describe 99.9% of the original database. The total number of 'descriptive vectors' in the linear EMSC model is now ten, consisting of seven loadings, the reference spectrum, plus the constant and sloping baselines. All the model parameters are estimated simultaneously by multiple linear regression, solved by least squares

estimation[75]. As in the previous Mie scattering-EMSC, the parameter estimation is stable due to the orthogonality of the loadings[10].

The remainder of the algorithm is exactly the same as the previously published Mie scattering-EMSC except that seven loadings are now used from the matrix of 10^3 RMieS *Q* curves:

$$\vec{Z}_{Raw} = c + m\vec{\nu} + h\vec{Z}_{Ref} + \sum_{i=1}^{7} g_i\vec{p}_i + \vec{E} \qquad (27)$$

Once the 10 parameters have been computed using a least squares algorithm, an estimation of the scattering spectrum is acquired. The measured spectrum, Z_{Raw}, minus the estimated scatter spectrum should theoretically yield an estimated corrected spectrum, Z_{Corr}.

$$\vec{Z}_{Corr} = \frac{\vec{Z}_{Raw} - \left(c + m\vec{\nu} + \sum_{i=1}^{7} g_i\vec{p}_i\right)}{h} \qquad (28)$$

5.3. Testing the RMieS-EMSC

5.3.1 Creation of simulated RMieS affected spectra

Earlier it was stated that our model for spectra affected by RMieS was the linear addition of the pure absorbance spectrum and the RMieS spectrum, *Q*. Using the 50 simulated spectra described earlier in section 2.3.4, a set of spectra affected by RMieS was be constructed. This data set was constructed by superimposing each spectrum onto a unique RMieS curve comprising of curves from scattering particles of 10 different sizes, to simulate the multiple scattering centres in single cells and tissue such as nucleoli and mitochondrion.

The first step was to take each spectrum and compute its n_{KK} spectrum using the Kramers-Kronig transform, then using a random number generator to create an average refractive

index between 1.3 and 1.4 was calculated and added to n_{KK}. For each spectrum 10 diameters for the theoretical scattering centres were calculated between the range of 2 µm to 6 µm.

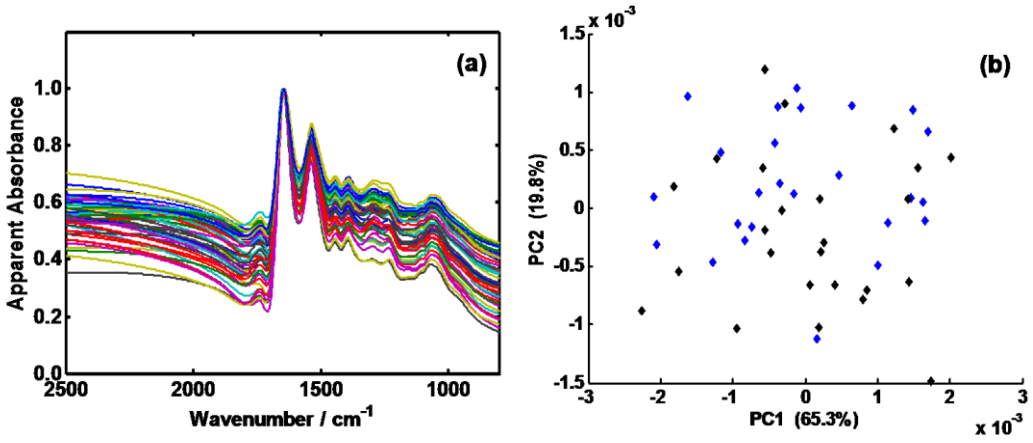

Figure 5.2 (a) The 50 simulated 'pure absorbance' after the superposition of 10 unique artificial Mie scattering curves. (b) PCA scores plot for the total data set of the 2nd derivative and normalisation of spectra from (a).

The first and most fundamental check was to see if the RMieS-EMSC algorithm could correct each spectrum given the perfect conditions to do so, *i.e.* the pure absorbance spectrum used as the reference spectrum for the calculation of the refractive index. If the mathematics of the algorithm are correct, then a subtraction of the corrected spectra from the original pure absorbance spectra should be zero as the algorithm should remove the scattering perfectly.

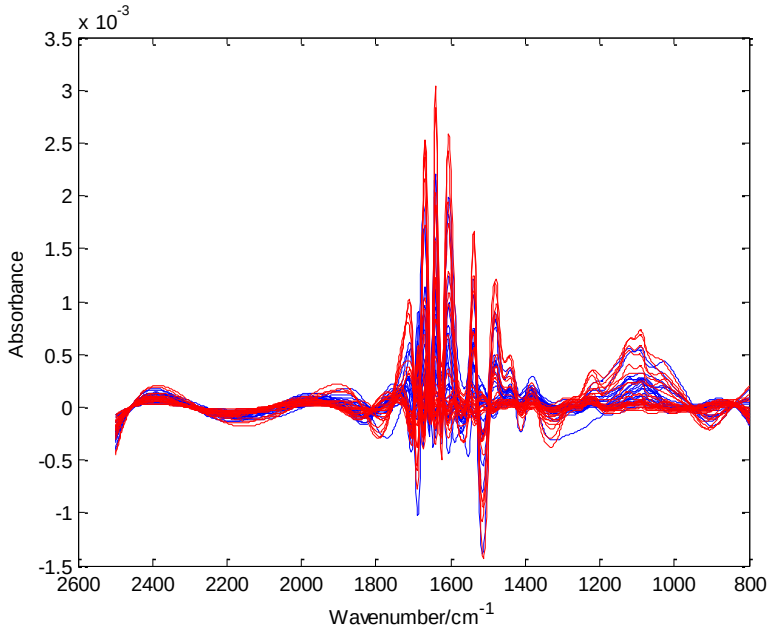

Figure 5.3 Difference spectra: Pure absorbance spectra - corrected spectra.

Figure 5.3 shows the difference spectrum between the pure absorbance spectra and the corrected spectra. The scaling of the absorbance axis shows that the magnitude of the differences is 10^{-3}, which is three orders of magnitude smaller than the pure absorbance spectra. This remaining difference is negligible and can be considered to be essentially zero. The origin of this difference can be explained by considering the fact that 1000 scattering curves were calculated and compressed using PCA. The exact scattering curve that was added to each spectrum was not covered exactly by the 1000 curves computed, however the approximation has given excellent results.

5.3.2. Using a non ideal reference spectrum

To simulate the case of a non ideal reference spectrum, the mean spectrum of the data set was used as the reference. This experiment simulates a real life case more effectively, as an

arbitrary spectrum such as Matrigel would be used when correcting single cell data and the outcome on the corrected spectra from this need to be understood.

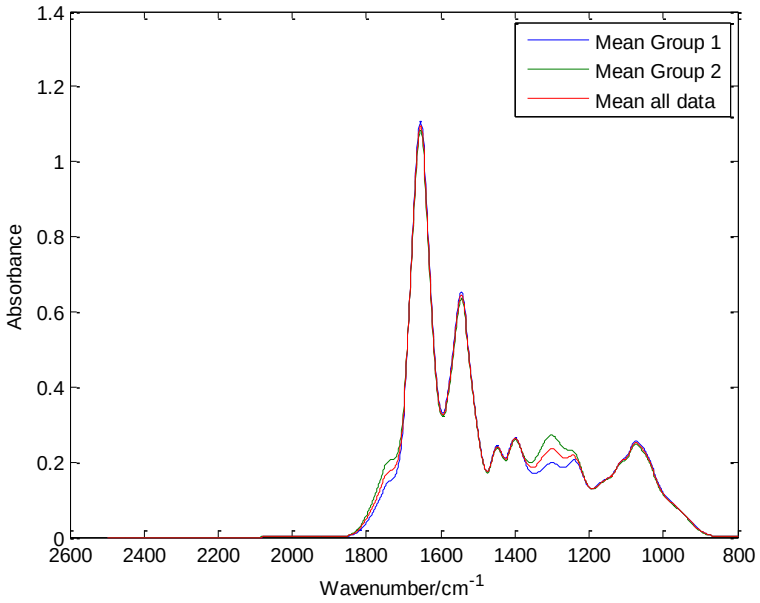

Figure 5.4 Blue trace is the mean spectrum for group 1, green trace is the mean spectrum group 2, and the red spectrum is mean spectrum for the whole data set.

Figure 5.4 shows that the mean spectrum of the data (red trace) will not act as an ideal spectrum for any of the spectra to be corrected as would be the case in real life. This will inevitably have effects on the quality of the corrected spectra. Below is a figure showing the difference between the pure absorbance and corrected spectrum for each sample, group 1 is blue and group 2 is red.

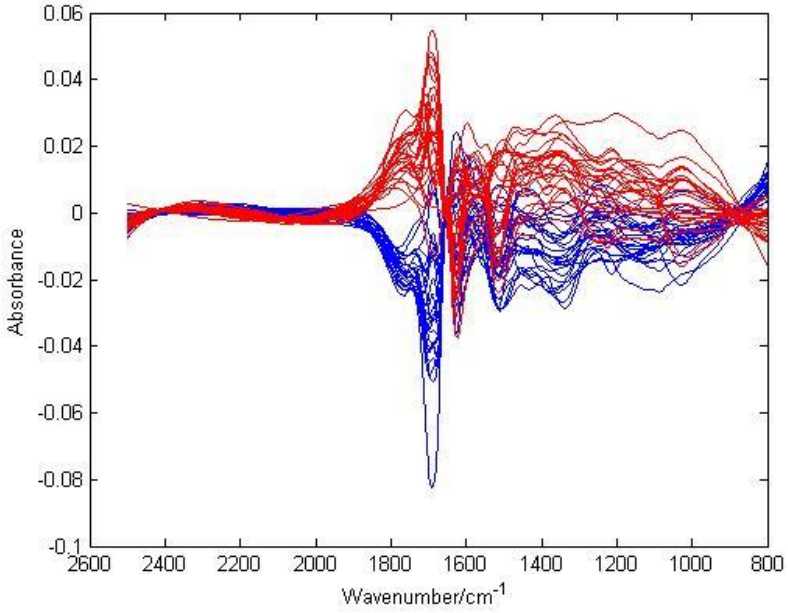

Figure 5.5 Difference between the pure and corrected spectra using a non-ideal reference spectrum.

Figure 5.5 shows that the difference between the pure and corrected spectra is more significant than before as indicated by the absorbance values. It is simple to understand how the previous algorithm corrected the spectra almost perfectly as the pure absorbance spectra were used as the reference and mathematically speaking, the conditions were optimum. In this case, the reference spectrum was not ideal hence there is a difference between the pure and corrected spectra.

This problem will occur when correcting spectra measured from single cells and tissue using an arbitrary reference spectrum such as Matrigel. What is important at this stage is to determine the extent to which the spectra have been distorted, and this will be done by using PCA to analyse the corrected data.

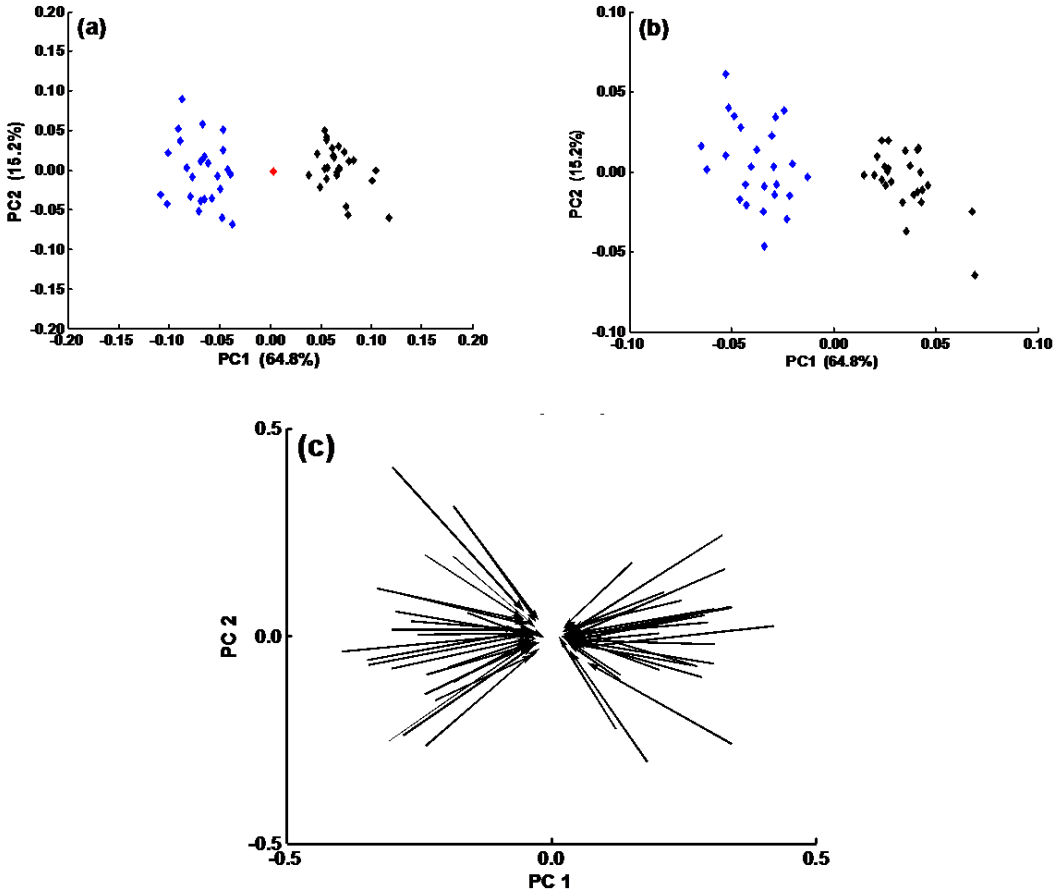

Figure 5.6 (a) PCA scores plot of the corrected spectra. (b) Scores plot of the non-ideal corrected data projected onto the loadings from the pure absorbance spectra PCA. (c) A plot showing the shift of the non-ideal reference corrected spectra from their correct pure absorbance positions.

Figure 5.6(a) shows the PCA scores plot of the pure absorbance spectra, ideally, our corrected spectra should have the same scores plot after PCA indicating that the spectra are identical and the correction accurate. To gain an insight into the effects of the distortion, the corrected spectra have been projected into the subspace produced from the PCA of the pure spectra. This essentially allows a visualisation of the scores plot of the corrected spectra to be produced but from the same mathematical 'point of view' as the pure spectra see Figure 5.6(b). This allows the position of the point in each plot to be compared to check similarity. Figure 5.6(a) shows a red point which is the mean spectrum of the pure absorbance spectra data set – by definition the mean of the data should be at

origin of the plot as it exhibits zero variance from the mean of the data set. Figure 5.6(b) shows that each spectrum has moved closer to the origin as indicated by the axis scaling and limits. This suggests that the corrected spectra bear similarity to the mean spectrum which was also the reference spectrum used[75].

Figure 5.6(c) shows arrows indicating the movement of each spectrum in the subspace visually illustrating the change in the spectra towards to the reference spectrum. An important feature of this movement is that there are still two clear groups of data, and each spectrum is approximately in the same position relative to the others. This result would be acceptable if the desired result was to check if there existed two groups of data within a dataset. This could be used successfully for classification exercises where each spectrum needs to be given a discrete class, i.e. cancer or non-cancer. If the corrected spectrum exhibits its own true chemistry sufficiently, then this may well serve an acceptable method for signal correction.

If possible, it is always desirable to recover the true pure absorbance signal representing the chemistry of the sample, and to this end, an iterative method of improving the spectral quality is stated in the next section.

5.3.3. Iterative RMieS-EMSC

In order to improve the spectral correction process even further, the reference spectra going into the RMieS-EMSC model may be optimised for a given purpose. This is achieved here by iteratively improving the reference spectrum, by letting the original, non-ideal reference spectrum (the mean of the input spectra) be replaced by the corrected spectrum from the previous iteration for each corresponding raw spectrum again using the new reference. This iterative approach is depicted schematically in Figure 5.7.

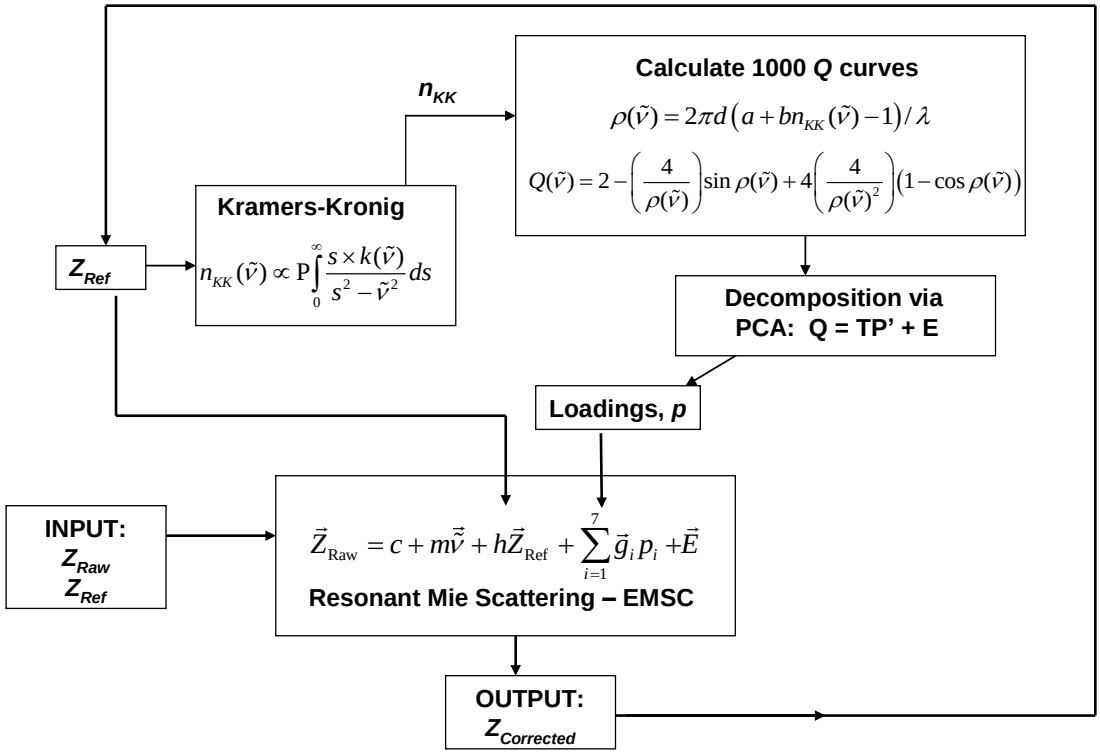

Figure 5.7 Flow chart illustrating the iterative procedure implemented to use the corrected spectrum as the new reference spectrum and running the algorithm once more.

Figure 5.8 shows the effect of increasing iterations on the accuracy of the correction, calculated by the projecting the corrected spectra onto the loadings from the pure absorbance spectra as stated earlier. Using this iterative approach, the corrected spectra move towards their pure absorbance spectra in score space with increasing numbers of iterations, indicating an improvement in the quality of the correction.

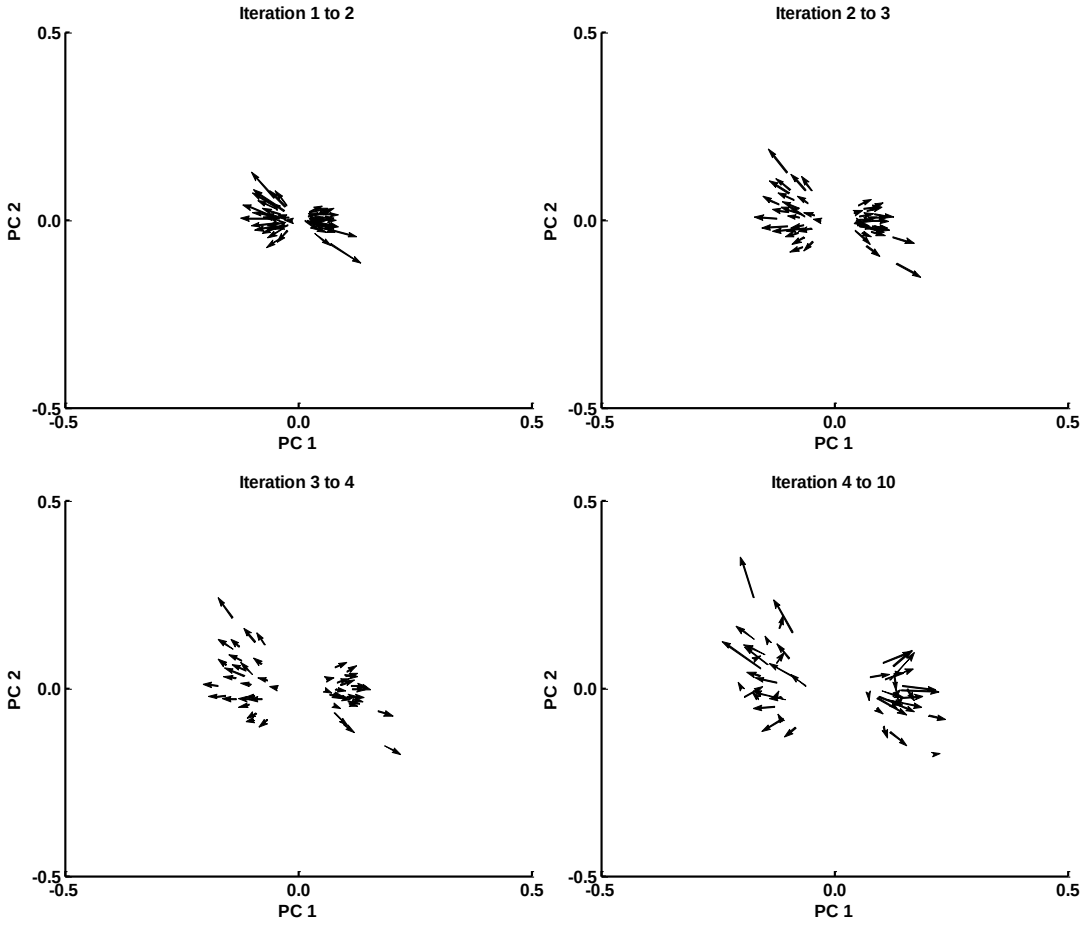

Figure 5.8 Scores plot showing the scores shift of the iteratively corrected spectra from iteration 1 to 2, 2 to 3, 3 to 4 and 4 to 10. Arrows show that each spectrum is moving towards its true absorbance spectrum position. All spectra were projected onto the loadings from the pure absorbance spectra.

For this particular data, convergence is reached after 8 iterations before the corrected spectra have stopped moving towards their pure absorbance positions; however, there is significant improvement upon the first iteration. This is illustrated further in Figure 5.9 which shows the sum of Pythagorean distances of the PCA scores plot of the corrected spectra, projected onto the original pure absorbance spectra subspace, as a function of iteration number[75].

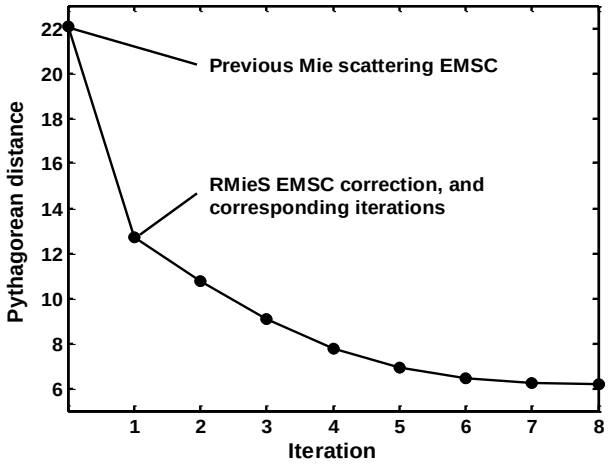

Figure 5.9 Plot of sum the Pythagorean distances of PCA scores away from the score positions for the pure absorbance spectrum (measured on a common subspace) against iteration number. The first point on the plot is for the previous Mie scattering-EMSC.

For comparison, the data point for the previous Mie Scattering-EMSC is also shown. As can be seen, using the RMieS-EMSC algorithm produced a significant improvement which continues with each iteration, in this case up to 8 iterations. The number of iterations required will of course depend on the data set. Although the sum of Pythagorean distances give a measure of how well the correction algorithm works, it is useful to consider other qualitative measures. Figure 5.10(a) shows the Amide I band for the original uncorrected simulated scatter data. The original position of the peak was set to 1655 ± 1 cm^{-1} indicated by the leftmost vertical line. The actual peak positions of the simulated scattering data range from 1635.9 to 1647.0 cm^{-1} with a mean of 1642.3 cm^{-1}. Thus it is clear that the significant shift in peak wavenumber is induced by the RMieS. The existing Mie scattering-EMSC correction significantly improves the overall data and brings down the sum of the Pythagorean distances (from the pure absorbance spectra) from a value of 95 to 22, but little impact on the Amide I peak position as can be seen in Figure 5.10(b).

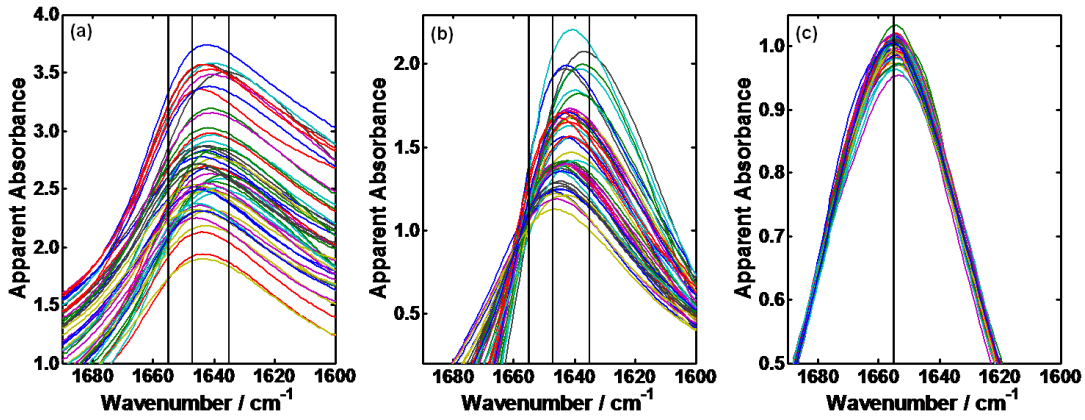

Figure 5.10 The Amide I band shown for (a) the uncorrected, (b) previous Mie scattering-EMSC and (c) RMieS-EMSC corrected spectra.

The peak position of the previous Mie scattering EMSC corrected spectra range from 1636.8 to 1647.1 cm^{-1} with a mean of 1642.7 cm^{-1}. It is only when the RMieS-EMSC correction is performed that a band position close to the correct wavenumber value is obtained. Figure 5.10(c) shows that the Amide I bands are now closely aligned, ranging from 1653.0 to 1656.4 cm^{-1} with a mean of 1654.5 cm^{-1}.

5.4 Conclusion

In this chapter, a correction algorithm to remove the distortion caused by resonant Mie scattering (RMieS) has been presented and tested. It was shown that using the true absorbance spectrum as the reference spectrum for each correction, every scattered spectrum can be corrected essentially perfectly. This result is non-trivial as it illustrates the concept of compressing 1000 Mie scattering curves into a small number of principal component loading spectra and using these in a least squares fitting algorithm to estimate the scattering contributions with no prior information about size and refractive index[75].

The second and more interesting test was the correction of the spectra using a non-ideal reference spectrum which was non-ideally suited to any spectrum as would be the case in real life as the true spectrum is unknown. This method yielded corrected spectra that still separated the test data set into two groups as they should when analysed with PCA.

Using the iterative correction process whereby the corrected spectrum becomes the new reference spectrum it has been shown that the new spectrum resembles its true pure absorbance spectrum even further. The limitation of this method is that convergence is reached before each spectrum is corrected perfectly; however each corrected spectrum is a significantly better representation of its true spectrum compared with that before the correction[75].

6. Validation of the RMieS-EMSC

In chapter 5 the resonant Mie scattering EMSC algorithm was presented and tested using simulated data. Two key questions surround the outcome of the algorithm, namely the effect of the choice of reference / first guess spectrum used, and the number of iterations required to be computed. A schematic of the algorithm showing the user defined inputs is shown in Figure 6.1.

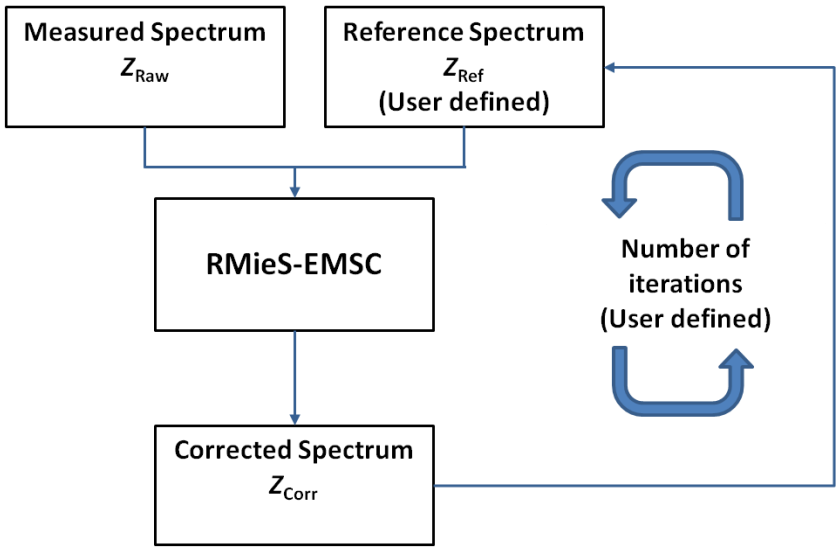

Figure 6.1 A simplified schematic diagram of the correction procedure. The user defined inputs are the choice of reference spectrum and the number of iterations used.

Results in chapter 5 showed that after one iteration of the RMieS-EMSC algorithm, corrected spectra contain some degree of similarity to the reference spectrum, however with increasing iterations, this influence is increasingly diluted. This chapter aims to investigate the effect of using an arbitrary reference spectrum, and the number of iterations required before the corrected spectra can be used in further data analysis.

6.1 Simulated data & classification

6.1.1. Simulation of data

A dataset was simulated to contain four different groups of data, consisting of 100 spectra, making the total data set of 400 spectra. The data were constructed in a similar manner to that documented in section 2.3.4, and Figure 6.2(a) and (b) show the pure absorbance spectra and the corresponding PCA scores plot respectively.

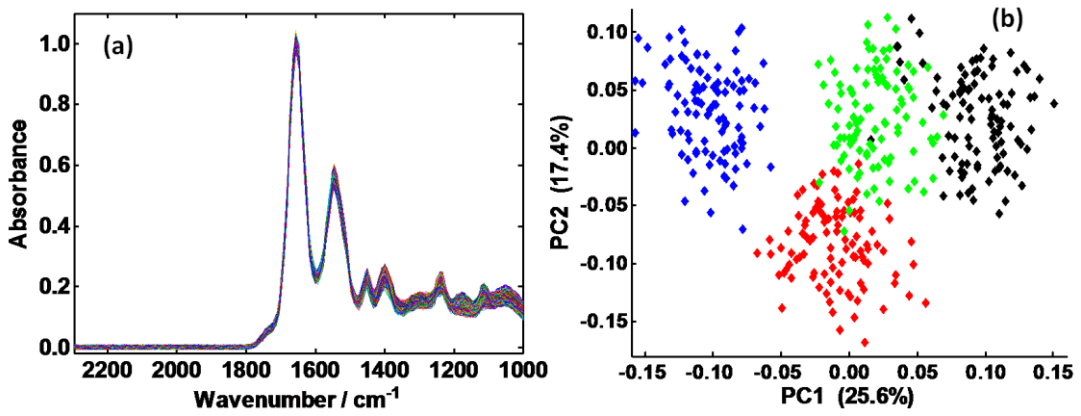

Figure 6.2 Simulated data of 4 groups containing 100 spectra each: (a) The absorbance spectra; (b) The PCA scores plot for the data.

Using these pure absorbance spectra and the same method outlined in section 5.3.1, resonant Mie scattering curves were created for each spectrum and added to the pure absorbance spectra to create simulated scattered spectra. The resultant spectra are shown in Figure 6.2(a), and the corresponding PCA scores plot shown in Figure 6.2(b).

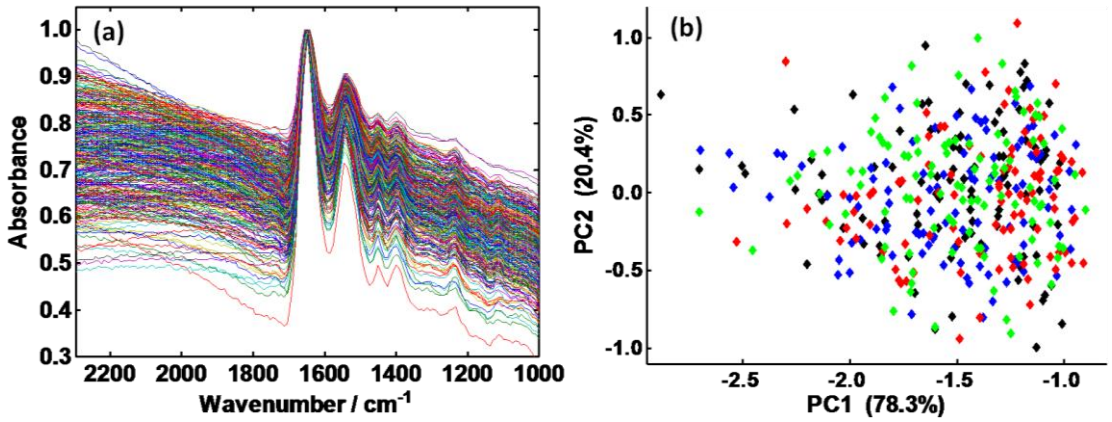

Figure 6.3 (a) Simulated scattered spectra based on Figure 6.2. (b) PCA scores plot of spectra from (a).

It is clear that after the additon of the RMieS curves, the spectra no longer seperate into four clusters as was previously seen. These spectra were corrected using the RMieS-EMSC algorithm, using a spectrum of Matrigel as the reference / first guess spectrum. A total of 25 iterations were computed, and the corrected spectra after each iteration were saved for later analysis. At this point, there are two ideas that can be investigated:

1. How well do the spectra cluster into their 4 groups after each iteration.
2. Using an artificial neural network (ANN) pattern recognition method, how well do the spectra classify into their correct group using half of the data to train, and the other half as a blind test.

6.1.2. Clustering accuracy of simulated data

Hierarchical cluster analysis (HCA) was used to investigate the ability to objectively cluster the data into four groups with zero prior knowledge of the data itself. The raw scattered data before any correction gave a 25% accuracy when objectively clustered using HCA, as is

to be expected as there is a 1 in 4 chance of correct assignment as no patterns can be found in the data.

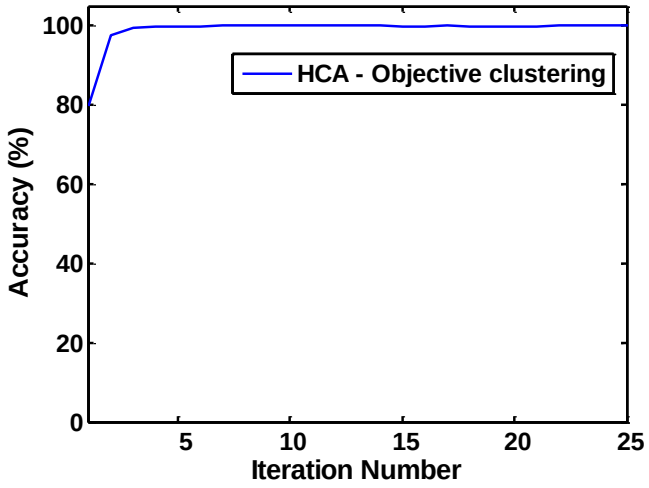

Figure 6.4 Plot showing HCA classification accuracy against RMieS-EMSC iteration number.

Figure 6.4 shows the percentage accuracy with which each spectrum was assigned to the correct class at different numbers of iterations of the RMieS-EMSC. After just one iteration, the classification accuracy increases from 25% for uncorrected spectra, to 80%. Subsequent iterations give further improvement on the classification accuracy, with 100% accuracy having been reached after 5 iterations.

6.1.3. ANN classification accuracy

An ANN model was built using the pure absorbance spectra in Figure 6.2. The RMieS-EMSC corrected data for each iteration was then subjected to the ANN model and the group classification results noted. This model was trained using the pure absorbance spectra, and subjecting the corrected spectra to the model after each iteration gives an indication into how accurate the RMieS-EMSC algorithm has corrected the data, *i.e.* do the corrected data resemble the pure absorbance spectra with which the model has been built.

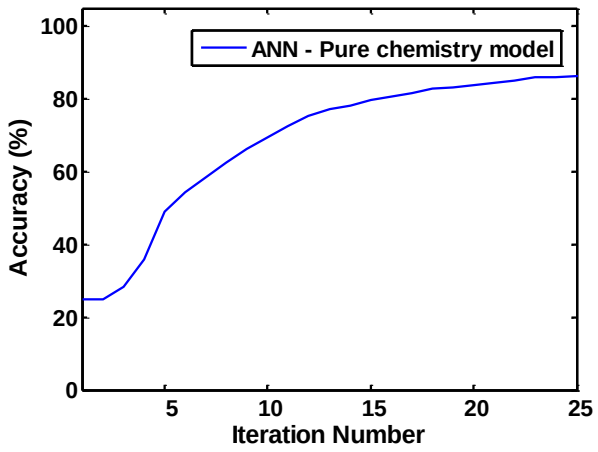

Figure 6.5 Classification % accuracy of an ANN model trained using the pure absorbance spectra subjected to corrected data from various iterations.

Figure 6.5 shows the classification accuracy of the RMieS-EMSC corrected data after each iteration, up to 25 iterations. The results show that the classification accuracy increases as the number of iterations increases with a plateau at approximately 25 iterations reaching 86% accuracy. This result shows that the corrected data represent the chemistry of the pure absorbance spectra more accurately with increasing iterations, however the plateau at 25 iterations show that spectral improvement stops before 100% accuracy. The significance of this result is discussed in greater detail later.

The situation where a classification model is built e.g. for tissue imaging where different tissue components needs to be classified, it is unlikely that training will be done using scatter-free pure absorbance spectra. The spectra used for training will most likely have to corrected for the effects of scattering themselves. To test the effects on classification of using data that has also been corrected, the data for each 25 iterations above has been used again. As a secondary test, after each iteration, half of the spectra in each group (50 spectra) were used to construct a four class ANN pattern recognition model (200 training

spectra in total). The remaining half of the spectra were then used as a blind test and subjected to the ANN model. The correct class assignments are again known, allowing the accuracy of the classification to be checked with each iteration number.

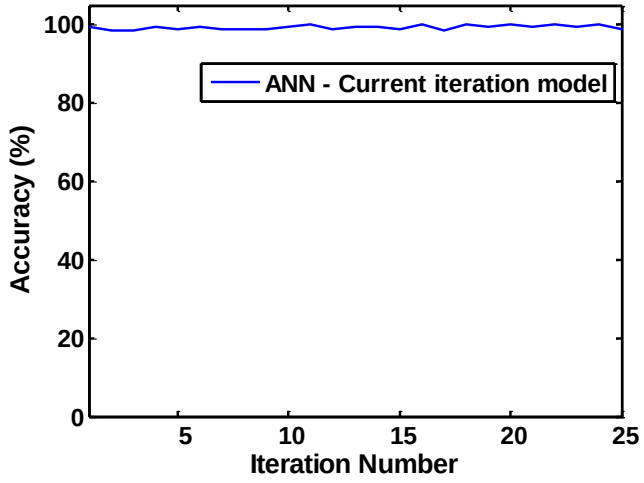

Figure 6.6 Classification % accuracy for model built using spectra from same iteration as those blind tested.

Figure 6.6 shows the result of the classification when the training spectra come from the same iteration number. The classification accuracy is close to 100% after just 1 iteration which is considerably better than the previous ANN model built using the pure absorbance spectra. This suggests that in each iteration, the influence of the reference spectrum on the data is of a 'similar effect' in all the spectra explaining the very high classification accuracy. This is an encouraging result showing that if real measured datasets are corrected using the same reference spectrum, and the same number of iterations as standard, then the effects of the RMieS-EMSC algorithm are more beneficial to the spectral quality than harmful. In the next section, further evidence of this idea is demonstrated using real measured data from the IR imaging of prostate tissue.

6.2. FTIR imaging of prostate tissue

FTIR imaging has shown potential in rapidly acquiring maps of the distributions of functional groups (sometimes called chemi-maps) within biopsies for diagnostic purposes. It has been shown that the major components of tissue have different FTIR spectral signatures which can be used to build predictive models such that new biopsy tissue can be imaged and classified to produce false colour images[109]. Using a Varian 670-FTIR spectrometer coupled to a Varian 620-FTIR imaging microscope (Varian Inc, - now Agilent Technologies, Santa Clara, CA) equipped with a 128×128 pixel liquid nitrogen-cooled Mercury-Cadmium-Telluride (MCT) focal planar array detector, FTIR images of prostate tissue were measured. Spectra were collected in the 850-4000 cm^{-1} range, at a resolution of 4 cm^{-1}, co-adding 64 scans for sample spectra, and 128 scans for the background spectra.

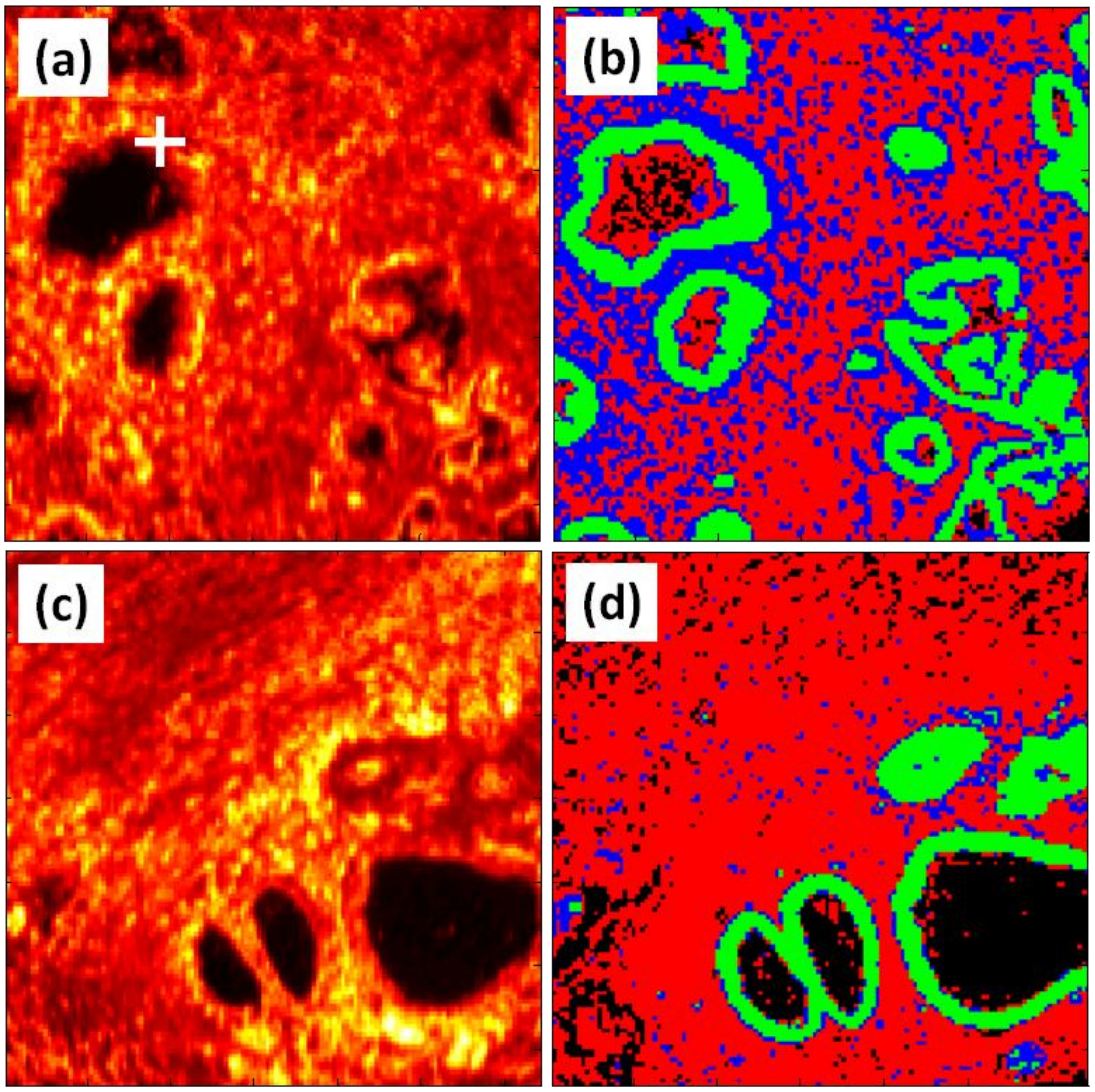

Figure 6.7 (a) and (c) are heatmap representations of the total absorbance from the FTIR images of prostate tissue from patient 1 and 2 respectively. (b) and (d) are serial sections which have been stained using the antibody anti-pancytokeratin; images have been thresholded so that green is epithelium, red is stroma, and blue is unclassed.

Figure 6.7(a) and (c) show the total absorbance intensity of FTIR images of benign prostate tissue from patient 1 and patient 2 (the colour scale is heat map style). Figure 6.7(b) and (d) show fluorescence images of the corresponding serial sections of tissue which have been stained using the antibody anti-pancytokeratin. This stain is fluorescence tagged and binds specifically to epithelial cells in the prostate tissue. By applying a thresholding method where bright pixels are epithelial (coloured as green) and dark pixels are stroma (red pixels)

a simpler representation of the prostate tissue can be acquired. The black areas in the thresholded fluorescence images indicate regions where there was essentially zero fluorescence meaning that no tissue was present due to lumen of a gland. The translation to the green and red colour system for the pixels gives as unequivocal and objective assignment of membership to either epithelial or stroma class.

Previously, it has been assumed that infrared spectra from microtomed tissue sections do not suffer from RMieS type distortion. It is clear however, that this is not the case. Figure 6.8 shows the raw spectra from the point marked with a white cross in Figure 6.7(a).

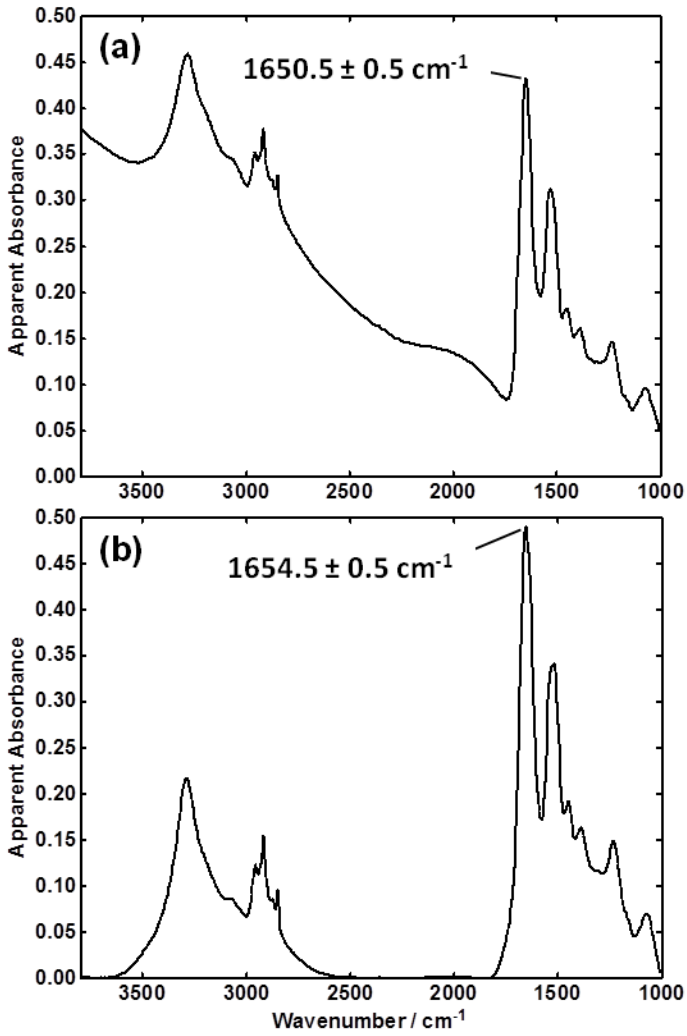

Figure 6.8 (a) Spectrum taken from edge of a gland in prostate tissue from area marked with white cross in Figure 6.7. (b) Corrected spectrum using the RMieS-EMSC.

Figure 6.8(a) shows a spectrum from the edge of a gland within prostate tissue, it is on the interface between the tissue and the slide on which the tissue is supported (spectrum taken from area marked with a white cross in Figure 6.7(a)). The morphology of the tissue at this point is not a flat surface, but rather an irregular curved surface which causes resonant Mie scattering. Figure 6.8(b) shows the same spectrum after the RMieS-EMSC algorithm has been applied using 20 iterations.

To investigate the influence of different reference spectra when using the RMieS-EMSC, the spectrum from Figure 6.8(a) was corrected using three different reference spectra which are shown in Figure 6.9.

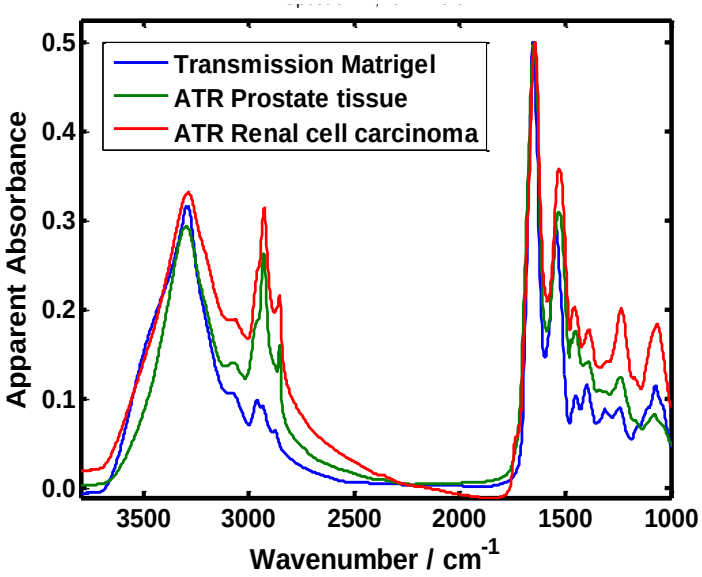

Figure 6.9 Three different reference spectra used as inputs for the RMieS-EMSC.

These three reference spectra were chosen as they are measured from different biological samples, and using different measurement techniques such as ATR. The spectrum from Figure 6.8(a) was corrected using each of these spectra in the RMieS-EMSC algorithm for a

total of 2000 iterations so that at each iteration, the resultant corrected spectra could be examined, Figure 6.10 shows the results.

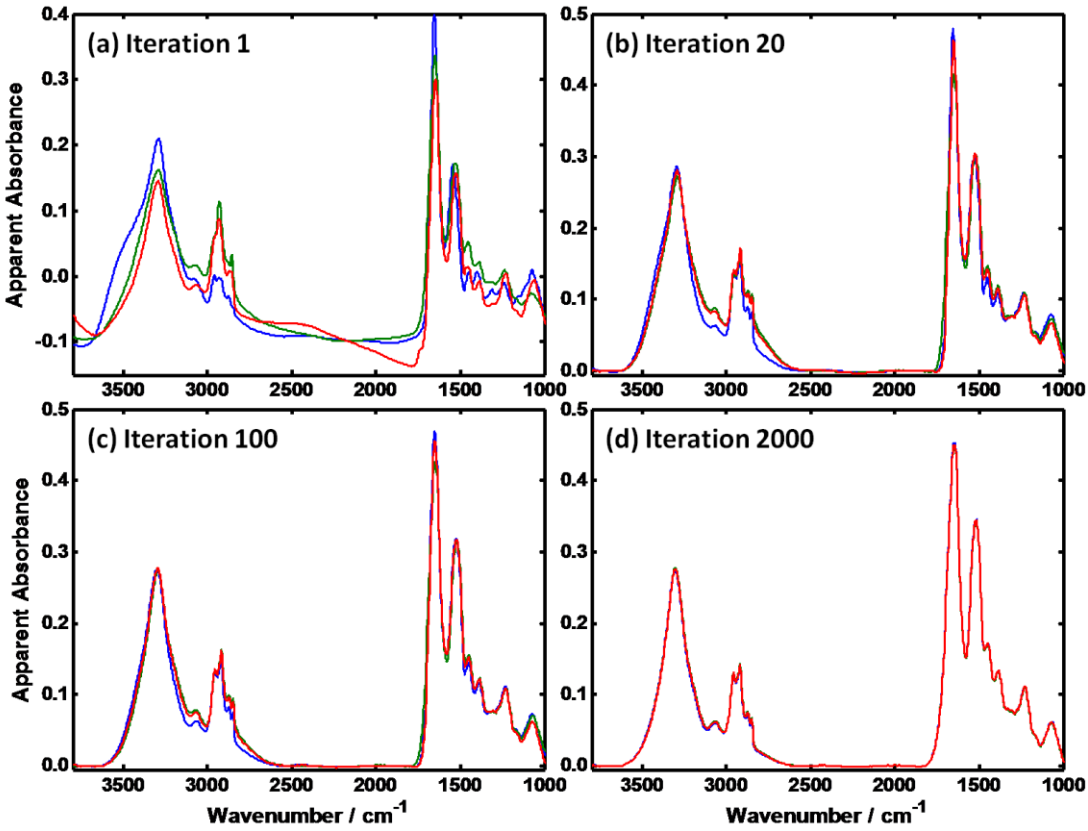

Figure 6.10 Corrected spectra of spectrum Figure 6.8(a) corrected using the 3 different reference spectra in Figure 6.9, for (a) 1; (b) 20; (c) 100; (d) 2000 iterations of the RMieS-EMSC.

Figure 6.10(a) shows that after one iteration of the RMieS-EMSC, the corrected spectra using the three different reference spectra give different looking spectra which still contain residual scattering features. After 20 iterations it can be seen that all three corrected spectra are now more similar to each other, with some differences still present in the 2800 to 3200 cm^{-1} wavenumber region. Further iterations produce increasingly similar corrected spectra showing that the corrected spectra are converging towards a common corrected spectrum. After 2000 iterations, the three corrected spectra are essentially identical showing that if enough iterations are used, then convergence can be reached. Although

2000 iterations is not computationally feasible, this exercise has shown that using three very different reference spectra, the same corrected spectrum can be obtained.

6.2.1. Classification of FTIR images from prostate tissue

Figure 6.7 showed the FTIR images from two different patients, both images contain regions of epithelial cells, and stroma (everything except the epithelium). The thresholded fluorescence images give us the correct assignment of the pixels from the infrared image which can be used to make a simple database of the two classes: epithelial and stroma spectra. Both images have been corrected using the RMieS-EMSC algorithm with Matrigel being the reference spectrum, and 30 iterations (results from all iterations were saved). This database can be used to build an ANN model for a two class problem, in this case the spectra from patient 1 have been used. Using just patient 1 to build the model allows patient 2 to be used as a completely blind test which can be subjected to the ANN model. The ANN model receives each of the 16384 spectra from the image of patient 2 individually then assigns them to either epithelium (green), stroma (red) or unclassed (blue) class membership. Figure 6.11 shows the classification results of the FTIR image from patient 2 for a number of different iterations.

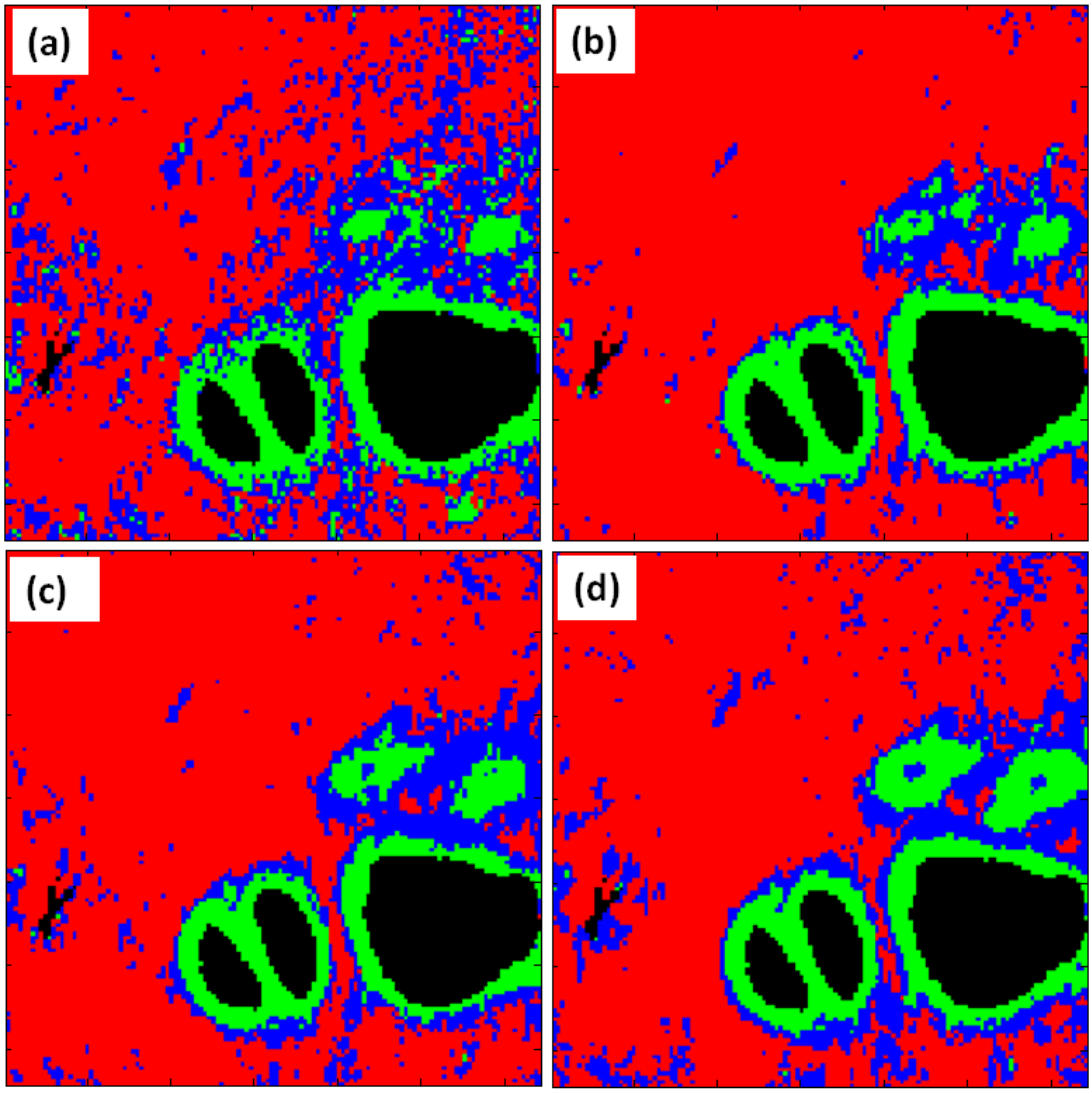

Figure 6.11 The classification of patient 2 using patient 1 as the training data after (a) 5, (b) 10, (c) 20 and (d) 30 iterations of the RMieS-EMSC algorithm. Training was done using spectra from the same iteration number.

Figure 6.11(a) shows the classification results of the FTIR image of patient 2 after 5 iterations, using an ANN model built from the FTIR image from patient 1 after 5 iterations. (b), (c), and (d) show the equivalent results for iterations 10, 20 and 30 respectively, it should be noted that the FTIR image for patient 1 that was used for model building had the same number of iterations each time for consistency.

The results show qualitatively that the classification accuracy improves with increasing iterations as the FTIR spectra are converging towards their pure absorbance spectrum. The

result of iteration 5 shows that the larger glands are still clearly defined and have been classified with a high accuracy.

6.3 Conclusion

In this chapter, spectra corrected using the RMieS-EMSC algorithm have been tested to see if they are of sufficient quality for real life applications. The major application investigated here is that of classification where an unknown spectrum is assigned to a class based on a computer pattern recognition model (artificial neural networks in this case). Using both simulated data and real measured data from prostate tissue it was shown that even though the corrected spectra are not 100% accurate in terms of representing the pure absorbance spectrum, they are sufficiently accurate that they can be used for classification purposes.

The results from the simulated data showed that if training data which has been corrected using the same reference spectrum and number of iterations as the blind data to be classified, then very high accuracy can be obtained even when the corrected spectra are not identical to their pure absorbance spectra. This shows that the pure absorption signal has been sufficiently recovered that the model can recognise which class each spectrum belongs to.

The results from the prostate tissue images show that using the RMieS-EMSC algorithm produces corrected spectra which can be used for building classification models, and are able to classify completely blind data with high accuracy. This data was not simulated, but real measured data providing further evidence that spectra may not need to be corrected 100%, but sufficiently so that the pure absorbance signal is greater than the distortion caused by the initial scattering.

7. Conclusion and future prospects

In this chapter, concluding remarks are presented for the results of the experiments and computations presented in this project. Following this, potential progression for the work in the future is presented.

7.1. Spectral distortion

In chapter 3, the reflection contributions to spectral distortions were theorised and validated with experiments. It was shown that in transflection experiments, if the transmission signal is weak compared to the magnitude of the reflection signal, then the measured spectrum will appear to be distorted. Instead of regular absorption band shapes, derivative-like shapes become prominent. These reflection contributions were not the only cause of spectral distortions as chapter 4 presented.

Chapter 4 showed that the broad oscillations in spectra and the 'dispersion artefact' were from the same phenomenon of resonant Mie scattering. Strongly absorbing samples such as those routinely measured in the field of biomedical spectroscopy have a changing refractive index at absorption bands. This change of refractive index produced the derivative-like line shapes of the 'dispersion artefact' and hence this effect was subsequently named resonant Mie scattering (RMieS). Experiments involving a model system of PMMA micro-spheres showed that theoretical calculations of scattering curves and measured data had excellent agreement thus validating the theory.

The mathematics used to describe the scattering curves was an adaptation of van de Hulst's approximation[78]. Although giving a good agreement with measured data, this approximation equation has limited applicability as it approximates the scattering efficiency for spherical dielectric particles which are illuminated by a homogenous parallel

beam of light. The configuration for FTIR spectroscopic measurements of biomedical samples is in general a focused illumination onto non-spherical particles which are strongly absorbing. At present there is no simple analytical mathematical description for this optical configuration meaning that approximations had to be made.

Recent work from the lab of Prof. Bhargava has started to give rigorous mathematical description of mid-IR spectroscopy experiments describing the propagation of the EM field through the microscope and optics[80-81]. Transflection and transmission experiments are described as well as the influence of substrates. This rigorous mathematical description will pave the way towards being able to describe the EM field incidence upon the scattering sample which will be a step towards modelling the physics of the scattering more accurately instead assuming a parallel beam.

7.2 The RMieS-EMSC algorithm

The RMieS-EMSC algorithm presented in this project has been shown to work with high accuracy assuming that the measured spectrum is comprised of the superposition of the pure absorbance spectrum and an RMieS curve. The accuracy of the algorithm was demonstrated with simulated data where the known results were available for comparison against the corrected data. The assumption of a superposition relationship between the pure absorbance spectrum and the scattering curve is a simplification of the real situation, and it is one of the first aspects of the correction algorithm that could be optimised. The formulation of an analytical description of measured spectrum in terms of the pure absorbance spectrum and the scattering is far from trivial as the mathematics is complicated.

The influence of the reference spectrum and the number of iterations used in the RMieS-EMSC was investigated and results showed that the influence of the reference spectrum decreased with increasing numbers of iterations. Further testing using real data from FTIR images from prostate tissue demonstrated that by using the RMieS-EMSC, the chemical signal of the corrected spectra classified using an artificial neural network model into the correct tissue component classes. This demonstrates that even if the pure absorbance spectrum is not recovered 100%, the corrected spectra may be sufficiently accurate for classification applications such as disease diagnostics from biopsies.

7.3 Future work

7.3.1. Theory

Many parts of this project can be optimised, starting with the correct mathematical description of the light loss due to scattering from biological samples illuminated by a focused beam. Using this theory, the measured spectrum's composition needs to be understood in terms of its pure absorbance spectrum and the scattering contribution. If this can be described as some form of linear addition of two components, then the RMieS-EMSC algorithm can be easily developed to incorporate this new theory. If the combination of the pure absorbance spectrum and the scattering contribution is non-linear, then alternative methods for signal correction will need to be pioneered.

The effects of diffraction have also not been considered in this thesis, which have effects on FTIR spectra from single point and imaging data. As an example, during an FTIR imaging measurement using a focal plane array detector, the transmission profile of an element may not be representative of the absorbance properties of the corresponding spatial area of the sample. Light that was not "transmitted" may have traversed a different path due to

scattering and distorted the spectra in adjacent elements. The converse is also possible, that light measured by an element may be due to scattered from a distance area of the sample. The RMieS-EMSC does not take this into consideration when correcting FTIR spectra as many additional input parameters would need to be inserted into the algorithm. The effects of diffraction are wavelength dependant introducing another level of complexity in conjunction to the scattering which is also wavelength dependant.

As mentioned earlier, encouraging work by the Bhargava group is starting to give a mathematically rigorous analytical description of the propagation of the EM field through infrared microscopes.

7.3.2. Experimental

In this project, the final outcome has been the production of a signal correction algorithm, the RMieS-EMSC, to mathematically / computationally deal with the problem of the spectra distortion. If an experimental solution could be found, then this would negate the need for complicated computational algorithms. One such experimental idea is the use of an integrating sphere which in principal could collect all light and return it to the detector. If all transmitted and scattered light could be returned to the detector, then the effects of scattering in the IR spectrum would no longer be noticeable. This idea however is technically difficult, as there is no simple optical setup with which one can perform light microscopy for finding single cells on a slide while placing the entire sample inside an integrating sphere.

7.4 Impact on infrared spectroscopy

The field of biomedical infrared spectroscopy faced a challenge when trying to measure spectra from single cells and tissue due spectral distortion. Single cell data especially was strongly affected, with spectra often rendered completely unreliable due to unexplained line shapes such as the 'dispersion artefact'. This project has demonstrated with theoretical arguments supported by experimental data that these previously unexplained line shapes are due to reflection contributions and resonant Mie scattering occurring during measurements. With an understanding of the theory, the first algorithm to correct for resonant Mie scattering was presented with an encouraging reception from peers within the field. Even in imaging of tissue data, previously spectra from the edges of glands were filtered out due to the 'anomalous' nature of the data[109]. The edges of glands are thought to be of interest in biology as this is where interesting biochemical changes may occur. The RMieS-EMSC offers the opportunity to include these spectra from the edges of tissues and areas of non-flat morphology.

8. References

1. Griffiths, P. R.; de Haseth, J. A., Fourier-Transform Infrared Spectrometry. *Science* **2007,** 222, (4621).
2. Buswell, A. M.; Downing, J. R.; Rodebush, W. H., Infrared Absorption Studies. XI. NH-N and NH-O Bonds. *J Am Chem Soc* **1940,** 62, (10), 2759-2765.
3. Sutherland, G. B. B. M., Application of Infrared Spectroscopy to Biological Problems. *Rev Mod Phys* **1959,** 31, (1), 118-122.
4. Bath, J. D.; Ellis, J. W., Some Features and Implications of the Near Infrared Absorption Spectra of Various Proteins: Gelatin, Silk Fibroin, and Zinc Insulinate. *The Journal of Physical Chemistry* **1941,** 45, (2), 204-209.
5. Uzman, L. L.; Blout, E. R., Infra-Red Spectra of Films of Native and Denatured Pepsin. *Nature* **1950,** 166, (4229), 862-863.
6. Choudhary, C.; Mann, M., Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Bio* **2010,** 11, (6), 427-439.
7. Diem, M.; Romeo, M.; Boydston-White, S.; Miljkovic, M.; Matthaus, C., A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004). *Analyst* **2004,** 129, (10), 880-885.
8. Huang, W. E.; Griffiths, R. I.; Thompson, I. P.; Bailey, M. J.; Whiteley, A. S., Raman microscopic analysis of single microbial cells. *Anal Chem* **2004,** 76, (15), 4452-4458.
9. Hughes, C.; Liew, M.; Sachdeva, A.; Bassan, P.; Dumas, P.; Hart, C. A.; Brown, M. D.; Clarke, N. W.; Gardner, P., SR-FTIR spectroscopy of renal epithelial carcinoma side population cells displaying stem cell-like characteristics. *Analyst* **2010,** 135, (12), 3133-3141.
10. Kohler, A.; Sule-Suso, J.; Sockalingum, G. D.; Tobin, M.; Bahrami, F.; Yang, Y.; Pijanka, J.; Dumas, P.; Cotte, M.; van Pittius, D. G.; Parkes, G.; Martens, H., Estimating and correcting Mie scattering in synchrotron-based microscopic Fourier transform infrared spectra by extended multiplicative signal correction. *Appl Spectrosc* **2008,** 62, (3), 259-266.
11. Maquelin, K.; Kirschner, C.; Choo-Smith, L. P.; van den Braak, N.; Endtz, H. P.; Naumann, D.; Puppels, G. J., Identification of medically relevant microorganisms by vibrational spectroscopy. *J Microbiol Meth* **2002,** 51, (3), 255-271.
12. Pijanka, J. K.; Kohler, A.; Yang, Y.; Dumas, P.; Chio-Srichan, S.; Manfait, M.; Sockalingum, G. D.; Sule-Suso, J., Spectroscopic signatures of single, isolated cancer cell nuclei using synchrotron infrared microscopy. *Analyst* **2009,** 134, (6), 1176-1181.
13. Romeo, M.; Diem, M., Correction of dispersive line shape artifact observed in diffuse reflection infrared spectroscopy and absorption/reflection (transflection) infrared micro-spectroscopy. *Vib Spectrosc* **2005,** 38, (1-2), 129-132.
14. Romeo, M. J.; Diem, M., Infrared spectral imaging of lymph nodes: Strategies for analysis and artifact reduction. *Vib Spectrosc* **2005,** 38, (1-2), 115-119.
15. Ackermann, K. R.; Koster, J.; Schlucker, S., Polarized Raman microspectroscopy on intact human hair. *J Biophotonics* **2008,** 1, (5), 419-424.
16. Agh-Atabay, N. M.; Dulger, B.; Gucin, F., Structural characterization and antimicrobial activity of 1,3-bis(2-benzimidazyl)-2-thiapropane ligand and its Pd(II) and Zn(II) halide complexes. *Eur J Med Chem* **2005,** 40, (11), 1096-1102.
17. Beer, M., Infrared Studies of the Structure of Protein in Tobacco Mosaic Virus. *Biochim Biophys Acta* **1958,** 29, (2), 423-423.
18. Benedict, A. A., The Study of Virus Preparations by Infrared Spectroscopy. *Annals of the New York Academy of Sciences* **1957,** 69, (1), 158-170.
19. Benedict, A. A.; Pollard, M.; Engley, F. B., Infrared Absorption Studies of Virus Preparations. *Tex Rep Biol Med* **1954,** 12, (1), 21-29.

20. Dovbeshko, G.; Repnytska, O.; Pererva, T.; Miruta, A.; Kosenkov, D., Vibrational spectroscopy and principal component analysis for conformational study of virus nucleic acids. *Xvi International Conference on Spectroscopy of Molecules and Crystals* **2004,** 5507, 309-316
420.
21. Erukhimovitch, V.; Mukmanov, I.; Talyshinsky, M.; Souprun, Y.; Huleihel, M., The use of FTIR microscopy for evaluation of herpes viruses infection development kinetics. *Spectrochim Acta A* **2004,** 60, (10), 2355-2361.
22. Aghatabay, N. M.; Bas, A.; Kircali, A.; Sen, G.; Yazicioglu, M. B.; Gucin, F.; Dulger, B., Synthesis, Raman, FT-IR, NMR spectroscopic characterization, antimicrobial activity, cytotoxicity and DNA binding of new mixed aza-oxo-thia macrocyclic compounds. *Eur J Med Chem* **2009,** 44, (11), 4681-4689.
23. Aghatabay, N. M.; Mahmiani, Y.; Cevik, H.; Dulger, B., Synthesis, Raman, FT-IR, NMR spectroscopic data and antimicrobial activity of mixed aza-oxo-thia macrocyclic compounds. *Eur J Med Chem* **2009,** 44, (1), 365-372.
24. Aghatabay, N. M.; Mahmiani, Y.; Cevik, H.; Gucin, F.; Dulger, B., Synthesis, FT-Raman, FT-IR, NMR spectroscopic characterization and antimicrobial activity of new mixed aza-oxo-thia macrocyclic compounds. *Struct Chem* **2008,** 19, (5), 833-842.
25. Aghatabay, N. M.; Somer, M.; Senel, M.; Dulger, B.; Gucin, F., Raman, FT-IR, NMR spectroscopic data and antimicrobial activity of bis[mu(2)-(benzimidazol-2-yl)-2-ethanethiolato-N,S,S-chloro-palladium(II)] dimer, [(mu(2)-CH2CH2NHNCC6H4)PdCl](2)center dot C2H5OH complex. *Eur J Med Chem* **2007,** 42, (8), 1069-1075.
26. Agirre, A.; Flach, C.; Goni, F. M.; Mendelsohn, R.; Valpuesta, J. M.; Wu, F. J.; Nieva, J. L., Interactions of the HIV-1 fusion peptide with large unilamellar vesicles and monolayers. A cryo-TEM and spectroscopic study. *Bba-Biomembranes* **2000,** 1467, (1), 153-164.
27. Ahmed, M. K.; Amiama, F.; Sealy, E. A., Unique spectral features of DNA infrared bands of some microorganisms. *Spectrosc-Int J* **2009,** 23, (5-6), 291-297.
28. Al-Holy, M. A.; Lin, M. S.; Al-Qadiri, H.; Cavinato, A. G.; Rasco, B. A., Classification of foodborne pathogens by Fourier transform infrared spectroscopy and pattern recognition techniques. *J Rapid Meth Aut Mic* **2006,** 14, (2), 189-200.
29. Al-Qadiri, H. M.; Al-Alami, N. I.; Al-Holy, M. A.; Rasco, B. A., Using Fourier transform infrared (FT-IR) absorbance spectroscopy and multivariate analysis to study the effect of chlorine-induced bacterial injury in water. *J Agr Food Chem* **2008,** 56, (19), 8992-8997.
30. Al-Qadiri, H. M.; Al-Holy, M. A.; Lin, M. S.; Alami, N. I.; Cavinato, A. G.; Rasco, B. A., Rapid detection and identification of Pseudomonas aeruginosa and Escherichia coli as pure and mixed cultures in bottled drinking water using Fourier transform infrared spectroscopy and multivariate analysis. *J Agr Food Chem* **2006,** 54, (16), 5749-5754.
31. Al-Qadiri, H. M.; Lin, M. S.; Cavinato, A. G.; Rasco, B. A., Fourier transform infrared spectroscopy, detection and identification of Escherichia coli O157 : H7 and Alicyclobacillus strains in apple juice. *Int J Food Microbiol* **2006,** 111, (1), 73-80.
32. Arp, Z.; Autrey, D.; Laane, J.; Overman, S. A.; Thomas, G. J., Structural studies of viruses by Raman spectroscopy part LXXI - Tyrosine Raman signatures of the filamentous virus Ff are diagnostic of non-hydrogen-bonded phenoxyls: Demonstration by Raman and infrared spectroscopy of p-cresol vapor. *Biochemistry-Us* **2001,** 40, (8), 2522-2529.
33. Batard, E.; Jamme, F.; Boutoille, D.; Jacqueline, C.; Caillon, J.; Potel, G.; Dumas, P., Fourier Transform Infrared Microspectroscopy of Endocarditis Vegetation. *Appl Spectrosc* **2010,** 64, (8), 901-906.
34. Becker, K.; Al Laham, N.; Fegeler, W.; Proctor, R. A.; Peters, G.; von Eiff, C., Fourier-Transform infrared spectroscopic analysis is a powerful tool for studying the dynamic

changes in Staphylococcus aureus small-colony variants. *J Clin Microbiol* **2006,** 44, (9), 3274-3278.
35. Beekes, M.; Lasch, P.; Naumann, D., Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. *Vet Microbiol* **2007,** 123, (4), 305-319.
36. Borawska, M. H.; Koczon, P.; Piekut, J.; Swislocka, R.; Lewandowski, W., Vibrational spectra and antimicrobial activity of selected bivalent cation benzoates. *J Mol Struct* **2009,** 919, (1-3), 284-289.
37. Bosch, A.; Minan, A.; Vescina, C.; Degrossi, J.; Gatti, B.; Montanaro, P.; Messina, M.; Franco, M.; Vay, C.; Schmitt, J.; Naumann, D.; Yantorno, O., Fourier transform infrared spectroscopy for rapid identification of nonfermenting gram-negative bacteria isolated from sputum samples from cystic fibrosis patients. *J Clin Microbiol* **2008,** 46, (8), 2535-2546.
38. Erukhimovitch, V.; Talyshinsky, M.; Souprun, Y.; Huleihel, M., Spectroscopic characterization of human and mouse primary cells, cell lines and malignant cells. *Photochem Photobiol* **2002,** 76, (4), 446-451.
39. Forrester, J. B.; Valentine, N. B.; Su, Y. F.; Johnson, T. J., Chemometric analysis of multiple species of Bacillus bacterial endospores using infrared spectroscopy: Discrimination to the strain level. *Anal Chim Acta* **2009,** 651, (1), 24-30.
40. Hassan, M.; Klaunberg, B. A., Biomedical applications of fluorescence imaging in vivo. *Comparative Med* **2004,** 54, (6), 635-644.
41. Kwak, J. T.; Hewitt, S. M.; Sinha, S.; Bhargava, R., Multimodal microscopy for automated histologic analysis of prostate cancer. *BMC Cancer* **2011,** 11, (1), 62.
42. Akbari, H.; Uto, K.; Kosugi, Y.; Kojima, K.; Tanaka, N., Cancer detection using infrared hyperspectral imaging. *Cancer Sci* **2011**.
43. Wang, X.; Qi, Z.; Liu, X.; Wang, S.; Li, C.; Liu, G.; Xiong, Y.; Li, T.; Tao, J.; Tian, Y., The comparison of hair from gastric cancer patients and from healthy persons studied by infrared microspectroscopy and imaging using synchrotron radiation. *Cancer Epidemiol* **2010,** 34, (4), 453-6.
44. Travo, A.; Piot, O.; Wolthuis, R.; Gobinet, C.; Manfait, M.; Bara, J.; Forgue-Lafitte, M. E.; Jeannesson, P., IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas. *Histopathology* **2010,** 56, (7), 921-31.
45. Pezzei, C.; Pallua, J. D.; Schaefer, G.; Seifarth, C.; Huck-Pezzei, V.; Bittner, L. K.; Klocker, H.; Bartsch, G.; Bonn, G. K.; Huck, C. W., Characterization of normal and malignant prostate tissue by Fourier transform infrared microspectroscopy. *Mol Biosyst* **2010,** 6, (11), 2287-95.
46. Beljebbar, A.; Dukic, S.; Amharref, N.; Manfait, M., Screening of biochemical/histological changes associated to C6 glioma tumor development by FTIR/PCA imaging. *Analyst* **2010,** 135, (5), 1090-7.
47. Untereiner, V.; Piot, O.; Diebold, M. D.; Bouche, O.; Scaglia, E.; Manfait, M., Optical diagnosis of peritoneal metastases by infrared microscopic imaging. *Anal Bioanal Chem* **2009,** 393, (6-7), 1619-27.
48. Petter, C. H.; Heigl, N.; Rainer, M.; Bakry, R.; Pallua, J.; Bonn, G. K.; Huck, C. W., Development and application of Fourier-transform infrared chemical imaging of tumour in human tissue. *Curr Med Chem* **2009,** 16, (3), 318-26.
49. Mackanos, M. A.; Contag, C. H., FTIR microspectroscopy for improved prostate cancer diagnosis. *Trends Biotechnol* **2009,** 27, (12), 661-3.
50. Ly, E.; Piot, O.; Durlach, A.; Bernard, P.; Manfait, M., Differential diagnosis of cutaneous carcinomas by infrared spectral micro-imaging combined with pattern recognition. *Analyst* **2009,** 134, (6), 1208-14.

51. Ly, E.; Piot, O.; Wolthuis, R.; Durlach, A.; Bernard, P.; Manfait, M., Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. *Analyst* **2008,** 133, (2), 197-205.
52. Krafft, C.; Kirsch, M.; Beleites, C.; Schackert, G.; Salzer, R., Methodology for fiber-optic Raman mapping and FTIR imaging of metastases in mouse brains. *Anal Bioanal Chem* **2007,** 389, (4), 1133-42.
53. Einenkel, J.; Steller, W.; Braumann, U. D.; Horn, L. C.; Krafft, C., Unrealistic expectations for IR microspectroscopic imaging. *Nat Biotechnol* **2007,** 25, (1), 29-31; author reply 31-3.
54. Chew, S. F.; Wood, B. R.; Kanaan, C.; Browning, J.; MacGregor, D.; Davis, I. D.; Cebon, J.; Tait, B. D.; McNaughton, D., Fourier transform infrared imaging as a method for detection of HLA class I expression in melanoma without the use of antibody. *Tissue Antigens* **2007,** 69 Suppl 1, 252-8.
55. Bhargava, R., Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Anal Bioanal Chem* **2007,** 389, (4), 1155-69.
56. Steller, W.; Einenkel, J.; Horn, L. C.; Braumann, U. D.; Binder, H.; Salzer, R.; Krafft, C., Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. *Anal Bioanal Chem* **2006,** 384, (1), 145-54.
57. Petibois, C.; Deleris, G., Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends Biotechnol* **2006,** 24, (10), 455-62.
58. Krafft, C.; Shapoval, L.; Sobottka, S. B.; Schackert, G.; Salzer, R., Identification of primary tumors of brain metastases by infrared spectroscopic imaging and linear discriminant analysis. *Technol Cancer Res Treat* **2006,** 5, (3), 291-8.
59. Krafft, C.; Shapoval, L.; Sobottka, S. B.; Geiger, K. D.; Schackert, G.; Salzer, R., Identification of primary tumors of brain metastases by SIMCA classification of IR spectroscopic images. *Biochim Biophys Acta* **2006,** 1758, (7), 883-91.
60. Bhargava, R.; Fernandez, D. C.; Hewitt, S. M.; Levin, I. W., High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data. *Biochim Biophys Acta* **2006,** 1758, (7), 830-45.
61. Bambery, K. R.; Schultke, E.; Wood, B. R.; Rigley MacDonald, S. T.; Ataelmannan, K.; Griebel, R. W.; Juurlink, B. H.; McNaughton, D., A Fourier transform infrared microspectroscopic imaging investigation into an animal model exhibiting glioblastoma multiforme. *Biochim Biophys Acta* **2006,** 1758, (7), 900-7.
62. Amharref, N.; Beljebbar, A.; Dukic, S.; Venteo, L.; Schneider, L.; Pluot, M.; Vistelle, R.; Manfait, M., Brain tissue characterisation by infrared imaging in a rat glioma model. *Biochim Biophys Acta* **2006,** 1758, (7), 892-9.
63. Lasch, P.; Haensch, W.; Naumann, D.; Diem, M., Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta* **2004,** 1688, (2), 176-86.
64. Gazi, E.; Dwyer, J.; Lockyer, N.; Gardner, P.; Vickerman, J. C.; Miyan, J.; Hart, C. A.; Brown, M.; Shanks, J. H.; Clarke, N., The combined application of FTIR microspectroscopy and ToF-SIMS imaging in the study of prostate cancer. *Faraday Discuss* **2004,** 126, 41-59; discussion 77-92.
65. Gazi, E.; Dwyer, J.; Gardner, P.; Ghanbari-Siahkali, A.; Wade, A. P.; Miyan, J.; Lockyer, N. P.; Vickerman, J. C.; Clarke, N. W.; Shanks, J. H.; Scott, L. J.; Hart, C. A.; Brown, M., Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study. *J Pathol* **2003,** 201, (1), 99-108.
66. Papamarkakis, K.; Bird, B.; Schubert, J. M.; Miljkovic, M.; Wein, R.; Bedrossian, K.; Laver, N.; Diem, M., Cytopathology by optical methods: spectral cytopathology of the oral mucosa. *Lab Invest* **2010,** 90, (4), 589-598.

67. Diem, T. L. T.; Ngoc, H. L.; Canh, H. N.; Danh, P. T.; Kesara, N. B., Pharmacokinetics of a Five-day Oral Dihydroartemisinin Monotherapy Regimen in Patients with Uncomplicated Falciparum Malaria. *Drug Metab Pharmacok* **2008,** 23, (3), 158-164.
68. Romeo, M.; Mohlenhoff, B.; Diem, M., Infrared micro-spectroscopy of human cells: Causes for the spectral variance of oral mucosa (buccal) cells. *Vib Spectrosc* **2006,** 42, (1), 9-14.
69. Jamin, N.; Dumas, P.; Moncuit, J.; Fridman, W. H.; Teillaud, J. L.; Carr, G. L.; Williams, G. P., Highly resolved chemical imaging of living cells by using synchrotron infrared microspectrometry. *P Natl Acad Sci USA* **1998,** 95, (9), 4837-4840.
70. Gazi, E.; Gardner, P.; Lockyer, N. P.; Hart, C. A.; Brown, M. D.; Clarke, N. W., Direct evidence of lipid translocation between adipocytes and prostate cancer cells with imaging FTIR microspectroscopy. *J Lipid Res* **2007,** 48, (8), 1846-1856.
71. Gazi, E.; Dwyer, J.; Lockyer, N. P.; Miyan, J.; Gardner, P.; Hart, C. A.; Brown, M. D.; Clarke, N. W., A study of cytokinetic and motile prostate cancer cells using synchrotron-based FTIR micro spectroscopic imaging. *Vib Spectrosc* **2005,** 38, (1-2), 193-201.
72. Gazi, E.; Dwyer, J.; Lockyer, N. P.; Miyan, J.; Gardner, P.; Hart, C.; Brown, M.; Clarke, N. W., Fixation protocols for subcellular imaging by synchrotron-based Fourier transform infrared microspectroscopy. *Biopolymers* **2005,** 77, (1), 18-30.
73. Mohlenhoff, B.; Romeo, M.; Diem, M.; Woody, B. R., Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy. *Biophys J* **2005,** 88, (5), 3635-3640.
74. Holman, H. N.; Martin, M. C.; Blakely, E. A.; Bjornstad, K.; McKinney, W. R., IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based fourier transform IR spectromicroscopy. *Biopolymers* **2000,** 57, (6), 329-335.
75. Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Byrne, H. J.; Dumas, P.; Gazi, E.; Brown, M.; Clarke, N.; Gardner, P., Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst* **2010,** 135, (2), 268-277.
76. Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Jackson, E.; Lockyer, N.; Dumas, P.; Brown, M.; Clarke, N.; Gardner, P., RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing. *J Biophotonics* **2010,** 3, (8-9), 609-620.
77. Martens, H.; Stark, E., Extended Multiplicative Signal Correction and Spectral Interference Subtraction - New Preprocessing Methods for near-Infrared Spectroscopy. *J Pharmaceut Biomed* **1991,** 9, (8), 625-635.
78. Hulst, H. C. v. d., *Light scattering by small particles*. Wiley: New York, 1957.
79. Mie, G., Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen der Physik* **1908,** 330, (3), 377-445.
80. Davis, B. J.; Carney, P. S.; Bhargava, R., Theory of Mid-infrared Absorption Microspectroscopy: II. Heterogeneous Samples. *Anal Chem* **2010,** 82, (9), 3487-3499.
81. Davis, B. J.; Carney, P. S.; Bhargava, R., Theory of Midinfrared Absorption Microspectroscopy: I. Homogeneous Samples. *Anal Chem* **2010,** 82, (9), 3474-3486.
82. Lee, J.; Gazi, E.; Dwyer, J.; Brown, M. D.; Clarke, N. W.; Nicholson, J. M.; Gardner, P., Optical artefacts in transflection mode FTIR microspectroscopic images of single cells on a biological support: the effect of back-scattering into collection optics. *Analyst* **2007,** 132, (8), 750-755.
83. Mader, K. T.; Tetteh, J.; McAuley, W. J.; Lane, M. E.; Hadgraft, J.; Andanson, J. M.; Kazarian, S. G., Investigation of skin permeation using ATR-FTIR spectroscopic imaging and multivariate target factor analysis. *J Pharm Pharmacol* **2010,** 62, (10), 1279-1280.
84. Kazarian, S. G.; Chan, K. L. A., Micro- and Macro-Attenuated Total Reflection Fourier Transform Infrared Spectroscopic Imaging. *Appl Spectrosc* **2010,** 64, (5), 135a-152a.

85. Palombo, F.; Cremers, S. G.; Weinberg, P. D.; Kazarian, S. G., Application of Fourier transform infrared spectroscopic imaging to the study of effects of age and dietary L-arginine on aortic lesion composition in cholesterol-fed rabbits. *J R Soc Interface* **2009,** 6, (37), 669-80.
86. Andanson, J. M.; Chan, K. L.; Kazarian, S. G., High-throughput spectroscopic imaging applied to permeation through the skin. *Appl Spectrosc* **2009,** 63, (5), 512-7.
87. Miller, L. M.; Dumas, P., Chemical imaging of biological tissue with synchrotron infrared light. *Bba-Biomembranes* **2006,** 1758, (7), 846-857.
88. Guo, S. L.; Li, P. L.; Fang, F.; Huang, H.; Cheng, C. G., [FTIR spectra-principal component analysis of phenetic relationships of Huperzia serrata and its closely related species]. *Guang Pu Xue Yu Guang Pu Fen Xi* **2005,** 25, (5), 693-7.
89. Kohler, A.; Bertrand, D.; Martens, H.; Hannesson, K.; Kirschner, C.; Ofstad, R., Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Anal Bioanal Chem* **2007,** 389, (4), 1143-53.
90. Kohler, A.; Host, V.; Ofstad, R., Image analysis of particle dispersions in microscopy images of cryo-sectioned sausages. *Scanning* **2001,** 23, (3), 165-174.
91. Boydston-White, S.; Romeo, M.; Chernenko, T.; Regina, A.; Miljkovic, M.; Diem, M., Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: principal component and artificial neural network analysis. *Biochim Biophys Acta* **2006,** 1758, (7), 908-14.
92. Griffiths, P. R.; Yang, H. S.; Li, Q. B.; Ling, X. F.; Wang, J. S.; Yang, L. M.; Xu, Y. Z.; Weng, S. F.; Wu, J. G., Discrimination of normal and malignant gastric tissues with FTIR spectroscopy and principal component analysis. *Guang Pu Xue Yu Guang Pu Fen Xi* **2004,** 24, (9), 1025-7.
93. Kohler, A.; Host, V.; Enersen, G.; Ofstad, R., Identification of fat, protein matrix, and water/starch on microscopy images of sausages by a principal component analysis-based segmentation scheme. *Scanning* **2003,** 25, (3), 109-115.
94. Navas, N.; Romero-Pastor, J.; Manzano, E.; Cardell, C., Benefits of applying combined diffuse reflectance FTIR spectroscopy and principal component analysis for the study of blue tempera historical painting. *Anal Chim Acta* **2008,** 630, (2), 141-9.
95. Wang, J. J.; Qiu, Q. Y.; Liu, W., [Quality assessment of tobacco flavor by classification of principal component analysis-mahalanobis distance combined with FTIR-ATR fingerprint]. *Guang Pu Xue Yu Guang Pu Fen Xi* **2007,** 27, (5), 895-8.
96. 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. *J Agric Food Chem* **2005,** 53, (18), 7115-27.
97. Zhang, Z. F.; Liu, Y.; Zhang, H., [FTIR spectra-principal component analysis of Erigeron breviscapus and Erigeron multiradiatus from different areas]. *Guang Pu Xue Yu Guang Pu Fen Xi* **2009,** 29, (12), 3263-6.
98. Zhang, Z. X.; Liu, P.; Kang, H. J.; Liao, C. C.; Chen, Z. L.; Xu, G. D., [A study of the diversity of different geographical populations of Emmenopterys henryi using FTIR based on principal component analysis and cluster analysis]. *Guang Pu Xue Yu Guang Pu Fen Xi* **2008,** 28, (9), 2081-6.
99. Zuo, K.; Li, D. T.; Guo, S. L.; Chen, J. H., [FTIR spectra-principal component analysis of roots of Polygonum cuspidatum from different areas]. *Guang Pu Xue Yu Guang Pu Fen Xi* **2007,** 27, (10), 1989-92.
100. Lasch, P.; Beyer, W.; Nattermann, H.; Stammler, M.; Siegbrecht, E.; Grunow, R.; Naumann, D., Identification of Bacillus anthracis by Using Matrix-Assisted Laser Desorption

Ionization-Time of Flight Mass Spectrometry and Artificial Neural Networks. *Appl Environ Microb* **2009,** 75, (22), 7229-7242.
101. Khanmohammadi, M.; Garmarudi, A. B.; Ghasemi, K., Back-propagation artificial neural network and attenuated total reflectance-Fourier transform infrared spectroscopy for diagnosis of basal cell carcinoma by blood sample analysis. *J Chemometr* **2009,** 23, (9-10), 538-544.
102. Lasch, P.; Diem, M.; Hansch, W.; Naumann, D., Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. *J Chemometr* **2006,** 20, (5), 209-220.
103. Schmitt, J.; Beekes, M.; Brauer, A.; Udelhoven, T.; Lasch, P.; Naumann, D., Identification of scrapie infection from blood serum by Fourier transform infrared spectroscopy. *Anal Chem* **2002,** 74, (15), 3865-3868.
104. Hecht, E., *Optics*. Pearson Education; 4 edition: Addison-Wesley series in physics, 1974.
105. Bassan, P.; Byrne, H. J.; Lee, J.; Bonnier, F.; Clarke, C.; Dumas, P.; Gazi, E.; Brown, M. D.; Clarke, N. W.; Gardner, P., Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells (vol 134, pg 1171, 2009). *Analyst* **2009,** 134, (12), 2484-2484.
106. Bassan, P.; Byrne, H. J.; Bonnier, F.; Lee, J.; Dumas, P.; Gardner, P., Resonant Mie scattering in infrared spectroscopy of biological materials - understanding the 'dispersion artefact'. *Analyst* **2009,** 134, (8), 1586-1593.
107. Dumas, P.; Polack, F.; Lagarde, B.; Chubar, O.; Giorgetta, J. L.; Lefrancois, S., Synchrotron infrared microscopy at the French Synchrotron Facility SOLEIL. *Infrared Phys Techn* **2006,** 49, (1-2), 152-160.
108. Bird, B.; Miljkovic, M.; Diem, M., Two step resonant Mie scattering correction of infrared micro-spectral data: human lymph node tissue. *J Biophotonics* **2010,** 3, (8-9), 597-608.
109. Fernandez, D. C.; Bhargava, R.; Hewitt, S. M.; Levin, I. W., Infrared spectroscopic imaging for histopathologic recognition. *Nat Biotechnol* **2005,** 23, (4), 469-474.