MAYNOOTH UNIVERSITY

A thesis submitted to Maynooth University
for the degree of Doctor of Philosophy

Applications of Raman micro-spectroscopy for cancer diagnostics

Laura Therese Kerr

Supervisor: Dr. Bryan Hennelly

Head of Department: Dr. Ronan Farrell

Department of Electronic Engineering

August 2016

Declaration

I hereby certify that the material contained within this thesis is entirely my own work, does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____

ID Number: _____

Date: _____

Abstract

Bladder cancer has the highest recurrence rate of any cancer, and as with most solid organ malignancies, early diagnosis, detection, and treatment are imperative for good clinical outcomes. Cystoscopy is the cornerstone of bladder diagnostics for real-time visualization of the bladder mucosa. However, it is an uncomfortable, invasive procedure, and is not without significant risk and potential complications for the patient. Urine cytology is currently the only non-invasive diagnostic tool available for the diagnosis of bladder cancer; this method is highly sensitive for high grade tumours, but has low sensitivity for low grade tumours, which accounts for the majority of cases. Therefore, there exists a clinical need to develop and integrate a non-invasive, accurate technique to assist in the diagnosis of bladder cancer.

The combination of Raman micro-spectroscopy and voided urine cytology may provide an ideal platform to replace cystoscopy for bladder cancer diagnostics. By recording Raman spectra from cells obtained from urine cytology, it is possible to analyse the spectral differences associated with the biomolecular continuum of disease progression, as well as being able to classify between different pathological subgroups. Previous studies to date have shown promising results in the application of Raman based urine cytology; however, there appears a high degree of variability across experimental protocols, which is believed to have hindered the advancement of this technique into the clinic.

This thesis involves the design and building of a confocal Raman micro-spectrometer to be utilised for the analysis of urine cytology samples, with a key emphasis on the translation of Raman based urine cytology into the clinic. In order to achieve this, a range of traditional protocols and consumables are systematically examined in terms of their compatibility with Raman micro-spectroscopy, as well as comparing the differences between Raman micro-spectroscopy and another form of vibrational spectroscopy for bladder and prostate cancer diagnostics. Although no patient urine cytology samples are used in this thesis, simulated samples are generated using bladder and prostate cell lines along with commercially available synthetic urine. Additional experimentation is provided in order to investigate the impact of hypoxia and exosomal communication on cellular biochemistry.

Acknowledgements

This thesis represents a culmination of work that was achieved over four years, and like all great pieces of work, it would not have been possible without great support from others. It has been a long journey to get to this point, with a few bumps along the way, but as journeys go, it was quite an enjoyable one.

Primarily, I would like to thank my supervisor, Dr. Bryan Hennelly, for his support and guidance throughout these four years, and for providing me with the opportunity to work on this exciting project.

As with all interdisciplinary research projects, this work would not have been possible without the support from my (many) collaborators and their research institutes, particularly Hugh Byrne, who taught me how to record my first Raman spectrum, Alison Malkin for teaching me about clinical cytology, Therese Lynn and Luke Gubbins for growing countless batches of cells, Ivor Cullen for his wonderful insight into the clinical aspects of bladder cancer diagnostics, and Damien King for helping with some last minute experiments. Additionally, thank you to Amanda McCann, Jen McIntyre, Katarina Domijan, Shirley O'Dea, and Marion Butler for your help and expertise.

An additional thank you to all of the students, interns, and staff at Maynooth University, who have helped make every day a brighter experience, and to the basketball gang for helping me to de-stress every week! And a special thank you to Sinead Barton, whose friendship, optimism, and support helped me to make it through to the end.

I will be forever in debt to my family for the kindness, support, and encouragement that they have provided me with over the last few years – to my mother, brothers and their partners, nephews and nieces, thank you all.

And finally, a massive thank you to the most important person in my life – Niall. I never would have finished this PhD without you, *ich bin dir unendlich sehr dankbar für dein Unterstützung, Liebe und Ermutigung.*

Publications

1. **L.T. Kerr**, B.M. Hennelly. *A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides.* Chemometr. Intell. Lab., 158, 61-68, 2016.

2. **L.T. Kerr**, T.M. Lynn, I.M. Cullen, P.J. Daly, N. Shah, S. O'Dea, A. Malkin, B.M. Hennelly. *Methodologies for bladder cancer detection with Raman based urine cytology.* Anal. Methods, 25(8), 4991-5000, 2016.

3. S.J. Barton, **L.T. Kerr**, K. Domijan, B.M. Hennelly. *On the effect of experimental noise on the classification of biological samples using Raman micro-spectroscopy.* SPIE Photonics Europe, Proc. SPIE 9887, 2016.

4. **L.T. Kerr**, H.J. Byrne, B.M. Hennelly. *Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen.* Anal. Methods, 7(12), 5041-5052, 2015.

5. **L.T. Kerr**, A. Adams, S. O'Dea, K. Domijan, I. Cullen, B.M. Hennelly. *Classification of bladder cancer cell lines using Raman spectroscopy; A comparison of excitation wavelength, sample substrate and statistical algorithms.* SPIE Photonics Europe, Proc. SPIE 9129, 2014.

6. **L.T. Kerr**, L. Gubbins, K. Weiner-Gorzel, S. Sharma, M. Kell, A. McCann, B.M. Hennelly. *Raman spectroscopy and SERS analysis of ovarian tumour derived exosomes (TEXs): A preliminary study.* SPIE Photonics Europe, Proc. SPIE 9129, 2014.

7. **L.T. Kerr**, K. Domijan, I. Cullen, B.M. Hennelly. *Applications of Raman spectroscopy to the urinary bladder for cancer diagnostics.* Photon. Lasers Med., 3(3), 193-224, 2014.

Manuscripts – in preparation

1. **L.T. Kerr**, D. King, J. McIntyre, I.M. Cullen, P.J. Daly, M. Bulter, A. Malkin, J. Ducree, H.J. Byrne, B.M. Hennelly. *LiPhos, FTIR, and Raman spectroscopy: A comparison of modalities for the detection of prostate cancer cells.*

2. **L.T. Kerr**, L. Gubbins, A. Malkin, A. McCann, B.M. Hennelly. *Monitoring of changes in cellular biochemistry associated with hypoxia and exosomal communication.*

Contents

List of Figures

XI

List of Tables

List of symbols and abbreviations

The following list gives a short description of the symbols and abbreviations used throughout this thesis.

Abbreviation	Description
α	Polarizability of a molecule
ϵ	Dielectric function
λ	Wavelength (nm)
μ	Dipole moment
μ_t	Attenuation coefficient
ν	Vibrational frequency (Hz)
$\tilde{\nu}$	Vibrational frequency (cm^{-1})
σ	Standard deviation
χ_e	Anharmonicity constant
χ	Aspect ratio of a nanoparticle
ω	Frequency of an electric field
Ω	Angular velocity
a	Radius of a spherical nanoparticle
ALA	Aminolevulinic acid
b	Contamination signal
B	Rotational constant
B_0	Background spectrum
c	Speed of light constant
C	Concentration; Constant scalar value; Condenser
CCD	Charge coupled device
CIS	Carcinoma in situ
D	Diameter of the laser beam
D_{eq}	Depth of the Morse potential
DB	Dichroic beam splitter
DC	Digital camera
DFT	Discrete Fourier transform
DI	Deionized

DNA	Deoxyribonucleic acid
DRE	Digital rectal exam
DS	Dichroic short pass filter
e	Extinction cross section of a single nanoparticle
E	Energy; Electric field
$EMSC$	Extended multiplicative signal correction
ER	Estrogen receptor
f	f number of a spectrograph
F	Focal length of a lens
$F(J)$	Rotational terms of a molecule
$FACS$	Fluorescence-activated cell sorting
FN	False negative
FNA	Fine needle aspiration
FP	False positive
FT	Fourier transform
$FTIR$	Fourier transform infrared
G	Grade
$G(n)$	Vibrational terms of a molecule
H_2	Hydrogen
h	Planck's constant
HAL	Hexaminolevulinic acid
$HER2$	Human EGF receptor 2
HEx	Exosomes derived from hypoxic cells
HL	Halogen lamp
Hz	Hertz
I	Moment of inertia; Intensity
ICA	Independent components analysis
IR	Infrared
J	Rotational quantum number
L	Lens
LDA	Linear discriminant analysis
LRA	Linear regression analysis
LP	Line pass filter; long pass filter
M	Mirror; Width of Gaussian function; Metastases
MM	Multi-mode laser
MO	Microscope objective
MRI	Magnetic resonance imaging
$mRNA$	Messenger RNA
$miRNA$	Micro RNA

MRS	Modulated Raman spectroscopy
MSE	Mean square error
MVE	Multivesicular endosomes
n	Electronic energy level of a molecule
N	Order of a polynomial; Nodes
N_2	Nitrogen
N_A	Area density of a nanoparticle
N_D	Neutral density filter
NA	Numerical aperture
NEx	Exosomes derived from normoxic cells
NHS	National Health Service (UK)
NIR	Near-infrared
$NIST$	National Institute of Standards and Technology
$NMIBC$	Non-muscle-invasive bladder cancer
O_2	Oxygen
OCT	Optical coherence tomography
OI	Optical isolator
OLS	Ordinary least squares
P	Slowly varying baseline signal
PBS	Phosphate buffered saline
PCA	Principal components analysis
$PET - CT$	Positron emission tomography - computed tomography
PR	Progesterone receptor
PSA	Prostate specific antigen
r_s	Fresnal specular reflection for non-polarized light
R	Residual spectrum; Reference spectrum
RNA	Ribonucleic acid
$S(n, J)$	Combined rotational-vibrational terms of a molecule
SEM	Scanning electron microscopy
$SERS$	Surface enhanced Raman spectroscopy
$SLIM$	Spatial light interference microscopy
SM	Single-mode laser
$SNOM$	Scanning near-field optical microscopy
SNR	Signal-to-noise ratio
SRS	Standard Raman spectroscopy
t	Time
T	Time; Tumour; Pathological status
$T2$	ThinPrep 2000
TCC	Transitional cell carcinoma

TEM	Transverse electromagnetic (laser mode); Transmission electron microscope
TN	True negative
$TNBC$	Triple negative breast cancer
TP	True positive
TS	Translation stage
$TURBT$	Transurethral resection of bladder tumour
$TURP$	Transurethral resection of the prostate
UC	Urothelial carcinoma
WHO	World Health Organisation
X_0	Cell spectrum

Chapter 1

Introduction

Raman spectroscopy is an optical technique based on the inelastic scattering of light, whereby the scattered light is a different colour to the incident light. This phenomenon was discovered by Sir C.V. Raman in 1921. (1) Raman spectroscopy has come a long way since the 1920's, achieving a surge of enthusiasm over the last decade, with many research groups worldwide applying this technique for the identification of the minute biomolecular changes associated with cancer progression. This thesis aims to apply Raman micro-spectroscopy and multivariate statistical algorithms in order to identify, and therefore diagnose, cancer cells that are present within a cytology sample.

From a clinical perspective, the primary focus of this thesis is on the diagnosis of bladder cancer. Bladder cancer is the seventh most common cancer in the UK and Ireland, with around 10,000 people diagnosed every year; the majority of which are diagnosed in males. Bladder cancer has the highest recurrence rate of any cancer, and it has been reported as the most expensive malignancy from diagnosis to death for health care systems worldwide, costing the National Health Service (NHS) in the UK approximately £55 million a year. (2) The American Urological Association recommends cystoscopic surveillance every 3–6 months for 3 years, and at least once a year thereafter, particularly for high risk patients; however, cystoscopy is invasive, involving passing an endoscopic probe through the urethra into the urinary bladder, which is not without significant risk to the patient. It is also expensive, time consuming, and can result in patient anxiety and morbidity. Urine cytology is often used as an adjunct to cystoscopy; this is where a urine sample is obtained from the patient, and bladder cells that have naturally shed into the urine are analysed under a microscope by a cytopathologist. However, the process of isolating and identifying bladder cancer cells from a voided urine sample can be challenging, particularly for early stage bladder cancers, with diagnostic accuracies as low as 28% being reported. (3) Furthermore, long waiting lists, which are prominent within public health care systems, often lead to diagnostic delays, and therefore delays in the management of the cancer, with resultant poorer outcomes for the patient.

Within this thesis, Raman micro-spectroscopy is applied to cell lines that simulate urine cytology samples in order to develop an approach for detecting bladder cancer in a completely

non-invasive manner. If successful, this technique has the potential to revolutionize urological diagnostics, particularly for patients who are undergoing lifelong cystoscopic surveillance for bladder cancer. When developing a diagnostic device, it is important to design one that can be integrated into a busy clinic as efficiently as possible. In order to achieve this, the research within this thesis focuses on translating Raman micro-spectroscopy into a urology clinic, and is involved in understanding the needs and demands of such an environment. By systematically examining a range of consumables, protocols, lasers, and other instrumentation, this project aims to design and build a Raman system that can analyse the cells obtained from a urine sample using standard clinical protocols, such as the use of glass slides and common cell preserving agents, as well as being able to remove any blood cells present within urine. The removal of blood cells is particularly important as over 90% of patients presenting with bladder cancer have blood in their urine, and unfortunately blood has a large Raman signal that often hinders the ability to detect the biomolecular changes associated with cancer in bladder cells. By incorporating standard consumables and protocols into a Raman micro-spectroscopy system, the resultant cost to health care systems could be reduced, as well as allowing for further clinical tests to be applied to the same patient sample post-Raman acquisition, resulting in an overall better diagnosis and better patient outcome.

Additional experimentation on the comparison of Raman micro-spectroscopy with Fourier transform infrared (FTIR) spectroscopy for the classification of bladder and prostate cells obtained from urine cytology is also provided. And finally, Raman micro-spectroscopy is applied to investigate the impact of hypoxia on cellular biochemistry, and the role of exosomes in cancer communication.

1.1 Thesis outline

This thesis is intended to be legible to both the biomedical physicist and the clinician, and in order to achieve that, Chapters 2–5 provide a detailed background on the physics, engineering, numerical methods, and biology involved in each of the experiments presented throughout this thesis. A further break-down of the thesis is as follows:

Chapter 2: Raman spectroscopy

Chapter 2 provides information on the physics involved in the Raman scattering process; this includes classical and quantum physics descriptions of rotational, vibrational, and rotational-vibrational Raman spectroscopy. The basic components involved in a Raman spectrometer are discussed, along with an overview of the biomedical applications of Raman spectroscopy. Additionally, a brief introduction to the physics of FTIR, another vibrational spectroscopy technique, is provided.

Chapter 3: Design and building a confocal Raman micro-spectrometer

Chapter 3 presents information on the design considerations required when building a confocal Raman micro-spectrometer; these include the source laser, microscope body, spectrometer, CCD detector, and confocal aperture. A detailed description of our in-house custom Raman micro-spectrometer is provided, which was built as part of this PhD project. Additional information is also provided for three alternative Raman system designs.

Chapter 4: Numerical methods

The application of Raman micro-spectroscopy for cell classification involves the use of numerical processing techniques that are designed to reduce the impact of noise, eliminate unwanted background artifacts, and standardise Raman spectra in terms of calibration and normalisation, as well as the application of multivariate statistical algorithms for classification. Chapter 4 provides information on each of the pre-processing techniques that are applied throughout this thesis, along with overview of the two multivariate statistical algorithms that are applied for classification throughout.

Chapter 5: The biology of cells, cancer, and exosomes

Chapter 5 provides the reader with an overview of cell biology, and the formation of epithelial tissues. This is followed by an introduction to cancer biology, the current modalities applied for the diagnosis of many human cancers, and the impact of hypoxia (low oxygen conditions) within a cancerous tumour. The role of exosomes (microvesicles released by cells) within cell biology is discussed, along with information of how exosomes can be isolated from biological fluids. And finally, cytology sample preparation methods are discussed, with a key emphasis on cell deposition methods and fixation agents.

Chapter 6: Bladder cancer diagnostics using Raman spectroscopy: a literature review

A literature review of the previous applications of Raman spectroscopy for bladder cancer diagnostics is provided in Chapter 6. This chapter begins with an introduction to the current clinical approaches for diagnosing bladder cancer. A review of Raman spectroscopy applications to biopsy and tissue samples is provided, along with applications of Raman cystoscopic procedures. Previous approaches for Raman based urine cytology are reviewed, allowing for the identification of protocols and consumables that require further investigation.

Chapter 7: Sample substrates and source wavelengths for Raman based cytology

Due to the high degree of variability across experimental protocols applied for Raman cytology to date, this chapter compares a range of sample substrates and source wavelengths in order to determine their impact on Raman cytology spectra. A comparison is made between five source

wavelengths ranging from 473 nm to 830 nm, with spectra recorded from cells on ten different sample substrates. Results are presented both graphically and numerically, using three different metrics.

Chapter 8: Algorithm for the removal of the glass signal from Raman cytology spectra

Traditional preparation methods for cytology samples involve depositing cells on glass slides; however, the glass signal can sometimes obscure the Raman cell spectrum. Chapter 8 investigates three background algorithms for the removal of this glass signal along with the slowly varying baseline signal that is inherent across Raman cytology spectra. The resultant spectra are compared based on subsequent multivariate analyses. One algorithm in particular is found to effectively remove the glass signal, and it is hoped that this study will help translate Raman micro-spectroscopy into the clinic, where the use of glass slides is commonplace.

Chapter 9: Protocols for Raman based urine cytology

Chapter 9 attempts to move Raman micro-spectroscopy a step closer to the clinic by systematically examining a range of consumables, protocols, and instrumentation that are used in the standard cytopathology laboratory. The use of glass slides, traditional fixing agents, lengthy exposure to urine, red blood cell lysing agents, as well as common cell deposition methods, are investigated in order to gauge their impact on the diagnostic potential of Raman based urine cytology.

Chapter 10: Detection of prostate cancer cells using FTIR and Raman micro-spectroscopy

Chapter 10 investigates the application of FTIR and Raman micro-spectroscopy in order to identify prostate cancer cells that may be present within a urine cytology sample. Synthetic urine samples are prepared that contain prostate and bladder cell lines, and are deposited onto slides using traditional clinical protocols. Both vibrational spectroscopy methods are compared in terms of classification results, as well as in terms of practical implementation and cost.

Chapter 11: Monitoring of changes in cellular biochemistry associated with hypoxia and exosomal communication

Exosomes are nano-sized microvesicles released by cells, and it is believed that they play a significant role in cell signalling and communication. This chapter investigates the biomolecular differences between normoxic and hypoxic (growing under low oxygen conditions) cancer cells using Raman micro-spectroscopy. Further investigations are provided to monitor the ability of exosomes isolated from normoxic and hypoxic cells to induce changes in the biochemistry of neighbouring cancer cells. This study aims to identify the potential application of Raman micro-spectroscopy as a reliable platform for the diagnosis of patients in need of a more intensive

treatment plan.

Chapter 12: Conclusion

Chapter 12 concludes with a summary of the results and ideas presented throughout this thesis. This is followed by a discussion of the potential for future commercialisation, and the clinical translation, of Raman based urine cytology for bladder cancer surveillance. And finally, an overview is provided of the future work that could be achieved within the areas of Raman based urine cytology, exosome communication, as well as future applications of vibrational spectroscopy technologies for cancer diagnostics.

Chapter 2

Raman spectroscopy

2.1   Introduction

In 1921, upon a voyage from England to India, Sir Chandrasekhara Venkata Raman began to question the origin of the blue radiance of the Mediterranean sea and surrounding glaciers, and believed that it could not be explained by Lord Rayleigh's light scattering explanation. (4) When he returned to India, Raman performed many experiments to observe light scattering from water and transparent blocks of ice, and published his own scientific explanation in the journal Nature. (1) In 1927, following news of Professor Compton receiving the Nobel Prize for his explanation on the scattering of x-rays (5; 6), Raman employed a spectrograph to record monochromatic light penetrating transparent materials, resulting in the detection of new spectral lines, which would later become known as "Raman lines". Raman presented his discovery in 1928, leading to his achievement of the Nobel Prize in Physics in 1930. (7)

Raman scattering has come a long way from the first time Sir C.V. Raman asked "why is sea water blue?", to today, where this technology is being applied for cancer diagnostics. The invention of the laser, the double monochromator, and electronic methods of signal detection in the 1960s led to an expansion of the range of applications for which Raman scattering is useful; such as quality control in the pharmaceutical industry, analysis of sample purity in the semi-conductor industry, monitoring of drug/cell interactions, and disease diagnosis, which is the key focus of this thesis.

This chapter provides an insight into the physics of Raman scattering, covering rotational (Section 2.3), vibrational (Section 2.4), and rotational-vibrational Raman spectroscopy (Section 2.5). Section 2.6 discusses the basic components required for a Raman spectrometer, with a more detailed approach provided in Chapter 3. Section 2.7 provides an overview of the biomedical applications of Raman spectroscopy. A brief discussion of FTIR spectroscopy, another form of vibrational spectroscopy, is provided in Section 2.8, followed by a summary in Section 2.9.

2.2 The physics of Raman spectroscopy

Raman spectroscopy is a powerful tool for analysing the biomolecular composition of biological samples. This technique is based on the inelastic scattering of monochromatic (laser) light; this means that when light interacts with a biological sample, the emitted photons have a different energy to the incident photons. Only about 1 in 10^7 photons will undergo Raman scattering, the rest will undergo Rayleigh scattering as normal, with a small number undergoing Mie, Brillouin, Thompson, or Compton scattering. (8) Figure 2.1 shows a Jablonski energy level diagram, which compares the energy differences between photons emitted during Rayleigh scattering and Raman scattering, and illustrates that there is no change in energy between incident photons and Rayleigh scattered photons; however, Raman scattered photons either gain energy (Anti-Stokes) or lose energy (Stokes) during this scattering process. The magnitude of the change in energy observed in the Raman photons is dependent on the rotational and vibrational energies of the molecules in the sample (9), as will be discussed in more detail in Sections 2.3 and 2.4 respectively, and is described by the following equation:

$$\Delta\tilde{\nu}(\text{cm}^{-1}) = \left[\frac{1}{\lambda_{incident}(\text{nm})} - \frac{1}{\lambda_{scattered}(\text{nm})}\right] \times \frac{10^9(\text{nm})}{10^2(\text{cm})} \tag{2.1}$$

where $\Delta\tilde{\nu}$ represents the change in energy in terms of Raman shift, or wavenumbers (cm^{-1}), and λ is the wavelength (nm) of the incident or scattered photons.
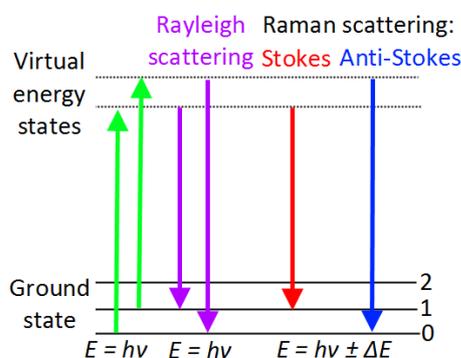
Figure 2.1: *Jablonski energy level diagram for Rayleigh and Raman scattering.*

By analysing the intensity and wavelength of the Raman photons, it is possible to infer information regarding the molecular composition of the sample being investigated. Typically only Stokes photons are measured since these have a higher intensity than Anti-Stokes photons. By recording and measuring the emission of inelastic photons in the form of Raman spectra, it is possible to detect the differences between normal and diseased, or cancerous, cells within a biological sample. Furthermore, Raman spectra can be recorded from cells or tissue samples without photo-damaging the individual cells, making it an ideal technique for diagnostics.

2.3   Rotational Raman spectroscopy

For a molecule to undergo rotational Raman scattering, the molecule must be anisotropically polarizable (i.e. the polarization varies with different crystallographic orientation). (10) Linear molecules that are anisotropically polarizable, but do not possess permanent dipole moments, such as O_2, N_2, and H_2, have a rotational Raman spectrum. In order to determine the polarizability of a molecule, the impact of an applied static electric field on the electron charge distribution of the molecules being analysed must be considered. The induced dipole moment (first order) is proportional to the electric field, and is given by:

$$\mu_{induced} = \alpha E \tag{2.2}$$

where $\mu_{induced}$ is the induced dipole moment, E is the amplitude of the external electric field, and α is the polarizability of the molecule; a property that indicates how readily its electron distribution can be distorted by the external electric field.

If the external electric field is due to an intense beam of light, such as a laser beam, then the induced dipole moment is a second order process. This means that the light initially creates a dipole, and then interacts with that dipole. Transitions created through this process are very weak, but observable. The induced dipole moment (second order) is given by:

$$\mu = \alpha E(t) = \alpha E_0 \cos(\omega_i t) \tag{2.3}$$

where E_0 is the amplitude of the electric field, ω_i is the angular frequency of the electric field, and t is time.

If an anisotropic molecule is rotating at an angular frequency of ω_R, then variations in its polarizability ($\Delta\alpha$) will be time dependent:

$$\alpha(t) = \alpha_0 + \Delta\alpha\cos(2\omega_R t) \tag{2.4}$$

with α ranging from $\alpha_0 - \Delta\alpha$ to $\alpha_0 + \Delta\alpha$ as the molecule rotates through 2π, where α_0 represents the spherical average of the parallel and perpendicular components. By substituting the expression for polarizability (Equation 2.4) into the expression for induced dipole moments (Equation 2.3), then we get:

$$\mu = [\alpha_0 + \Delta\alpha\cos(2\omega_R t)] \times [E_0\cos(\omega_i t)] \tag{2.5}$$

$$\mu = \alpha_0 E_0\cos(\omega_i t) + E_0\Delta\alpha\cos(\omega_R t)\cos(\omega_i t) \tag{2.6}$$

$$\mu = \alpha_0 E_0\cos(\omega_i t) + \frac{E_0\Delta\alpha[\cos(\omega_i t + 2\omega_R t) + \cos(\omega_i t - 2\omega_R t)]}{2} \tag{2.7}$$

where $\alpha_0 E_0\cos(\omega_i t)$ corresponds to Rayleigh scattering, $\frac{1}{2}E_0\Delta\alpha[\cos(\omega_i t + 2\omega_R t)]$ corresponds

to Anti-Stokes Raman scattering, and $\frac{1}{2}E_0\Delta\alpha[\cos(\omega_i t - 2\omega_R t)]$ represents Stokes Raman scattering. If the molecule is not anisotropically polarizable (i.e $\Delta\alpha = 0$), then the Stokes and Anti-Stokes components disappear. The distortion induced in a linear molecule by an external electric field returns to its initial value after a rotation of 180°.

In order to fully understand the physics involved with Raman scattering, it is necessary to consider the quantum effects that are taking place. A diatomic molecule has nuclear, electronic, vibrational, and rotational degrees of freedom. (11) A molecule can simply be approximated as a rigid motor and an anharmonic oscillator. (12) The typical energy level structure of a diatomic molecule is shown in Figure 2.2, represented using Morse curves. Morse curves provide a good approximation for lower vibrational levels, however this is not true for higher vibrational levels. (13) Electronic levels are represented by S, vibrational levels as v, and rotational levels as J. The Maxwell-Boltzmann distribution can be applied to calculate the most populated rotational levels for a given temperature, T; however only the ground electronic level ($n = 0$) is appreciably populated at room temperature. (14; 15; 16)
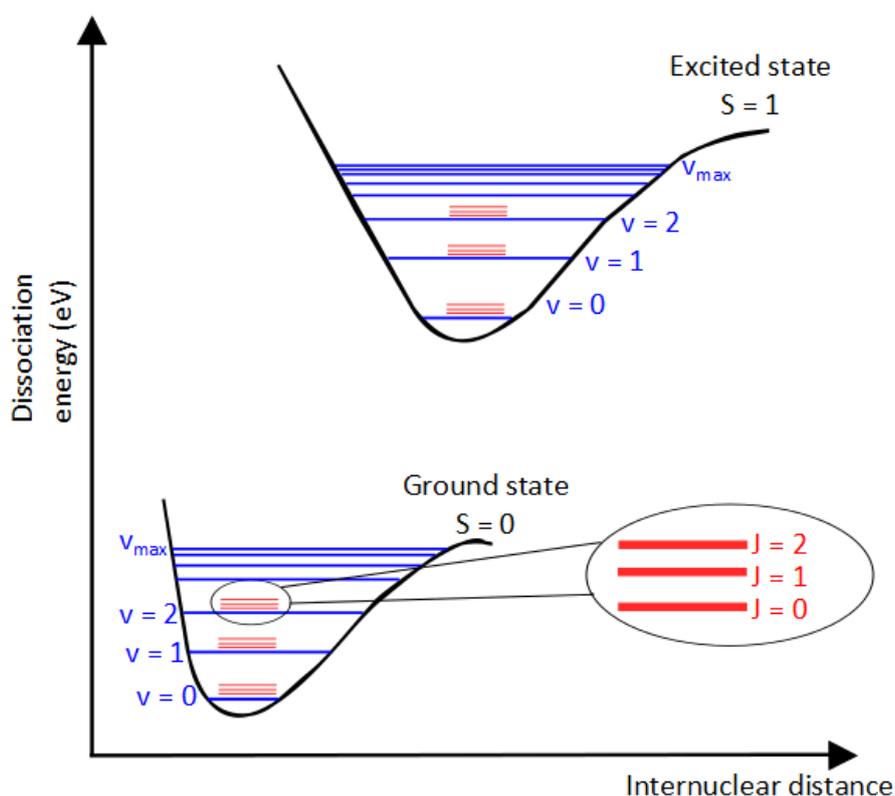
Figure 2.2: *General structure of the energy levels of a diatomic molecule, determined with Morse potential curves, showing the electronic energy levels (S), vibrational levels (v), and rotational levels (J).*

Two photons are involved in the Raman spectroscopy process, and in nuclear physics terms, each photon is a spin-1 particle. (10) This means that a maximum change in angular momentum quantum number of ± 2 is possible. This gives us the selection rule for rotational Raman
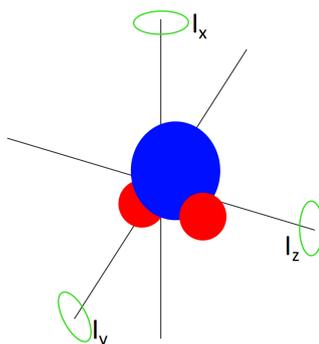
Figure 2.3: *An illustration of the moments of inertia about the three axes for a rotating molecule.*

spectroscopy:

$$\Delta J = \pm 2 \tag{2.8}$$

where J is the rotational angular momentum quantum number of the molecule, and can be described as:

$$J_x = I_x \Omega_x \tag{2.9}$$

where I_x is the moment of inertia and Ω_x is the angular velocity about an axis x. A molecule can have three different moments of inertia about orthogonal axes x, y, and z, as illustrated in Figure 2.3. Therefore, the energy of a rotating molecule, E_J, is given by:

$$E_J = \frac{1}{2}I_x\Omega_x^2 + \frac{1}{2}I_y\Omega_y^2 + \frac{1}{2}I_z\Omega_z^2 \tag{2.10}$$

$$E_J = \frac{J_x^2}{2I_x} + \frac{J_y^2}{2I_y} + \frac{J_z^2}{2I_z} \tag{2.11}$$

Or alternatively,

$$E_J = \frac{h^2 J(J+1)}{8\pi^2 I} \tag{2.12}$$

with h representing Planck's constant.

The energy of a rotational state is usually given by the Rotational Term $F(J)$:

$$F(J) = \frac{\Delta E_J}{hc} = BJ(J+1) \tag{2.13}$$

with c representing the speed of light constant, and B is the rotational constant (cm^{-1}) of the molecule, given by:

$$B = \frac{h}{8\pi^2 I c} \tag{2.14}$$

Spacing between lines in rotational spectra are often given in terms of B, with Raman lines appearing at displacements of $6B$, $10B$, $14B$, etc. either side of the incident source radiation, as described further in the following paragraphs.

For Stokes scattering, the molecule gains rotational energy of $\Delta J = +2$, and is represented

by the following equations:

$$\tilde{\nu}_{Stokes} = \tilde{\nu}_i - \tilde{\nu}_0 \tag{2.15}$$

where $\tilde{\nu}_{Stokes}$ is the frequency of the Raman shift (wavenumbers) for Stokes scattering, $\tilde{\nu}_i$ corresponds to the incident radiation (e.g. laser source), and $\tilde{\nu}_0$ corresponds to the Raman scattered photons. Therefore,

$$\tilde{\nu}(J+2) - \tilde{\nu}(J) = \tilde{\nu}_i - [F(J+2) - F(J)] \tag{2.16}$$

where $\tilde{\nu}$ is the Raman shift (wavenumbers) for the transition of J to $J+2$. By substituting Equation 2.13 into Equation 2.16, the following equations can be used to represent rotational Stokes scattering:

$$\tilde{\nu}(J+2) - \tilde{\nu}(J) = \tilde{\nu}_i - [B(J+2)(J+3) - BJ(J+1)] \tag{2.17}$$

$$\tilde{\nu}(J+2) - \tilde{\nu}(J) = \tilde{\nu}_i - 2B(2J+3) \tag{2.18}$$

Stokes lines appear on the low frequency side of the incident line source, and at displacements of $6B$, $10B$, $14B$, etc. from $\tilde{\nu}_i$ (Rayleigh position) for $J = 0, 1, 2$, etc., as illustrated in Figure 2.4.
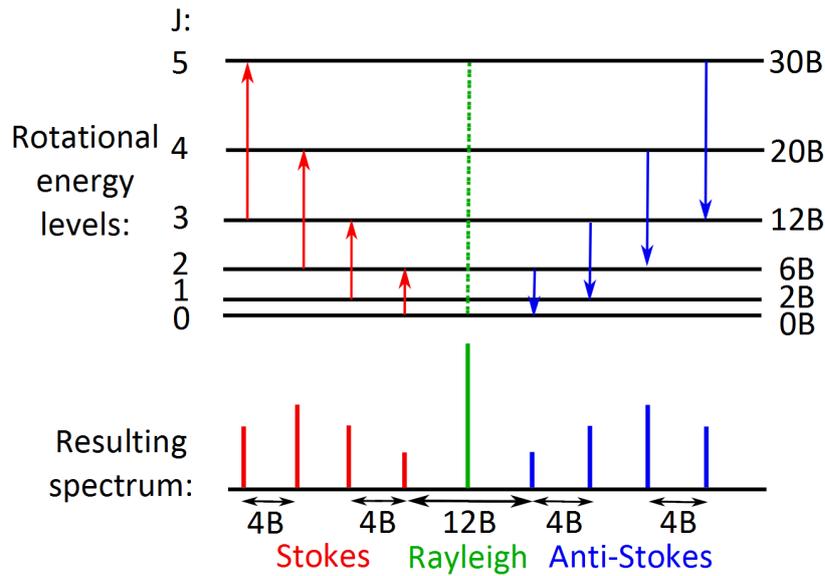
Figure 2.4: *Pure Rotational Raman energy levels and corresponding spectrum, showing Stokes and Anti-Stokes energy transitions and spectral lines.*

When the molecule makes a transition of $\Delta J = -2$, the scattered photon emerges with increased energy. These transitions account for the Anti-Stokes lines of the spectrum, given by the following equations:

$$\tilde{\nu}(J-2) - \tilde{\nu}(J) = \tilde{\nu}_i + [F(J-2) - F(J)] \tag{2.19}$$

$$\tilde{\nu}(J-2) - \tilde{\nu}(J) = \tilde{\nu}_i + [B(J-2)(J-1) - BJ(J+1)] \tag{2.20}$$

$$\tilde{\nu}(J-2) - \tilde{\nu}(J) = \tilde{\nu}_i + 2B(2J-1) \tag{2.21}$$

Anti-Stokes lines occur at displacements of $6B$, $10B$, $14B$, etc. for $J = 2, 3, 4$, etc., to the high frequency side of the incident line source, as shown in Figure 2.4.

2.4 Vibrational Raman spectroscopy

For vibrational Raman spectroscopy to occur, the polarizability of the molecule must change as the molecule vibrates. As diatomic molecules vibrate, the interaction between the electrons and nuclei changes resulting in a change in polarizability. Therefore, all diatomic molecules are vibrationally Raman active. (17) For polyatomic molecules, group theory must be applied to decide if a particular vibrational mode is Raman active. (18) There are two specific selection rules for vibrational Raman spectroscopy:

1. Harmonic approximation: $\Delta n = \pm 1$

2. Anharmonic approximation: $\Delta n = \pm 1, \pm 2, \pm 3...$

The classical description of vibrational Raman spectroscopy is similar to that presented for rotational Raman spectroscopy in Section 2.3. However, for vibrational Raman spectroscopy, the vibration of the molecular bond is accompanied by a periodic stretching and compression of the electron distribution, which gives an oscillation of the polarizability along the direction of the electric field.

Based on Equation 2.4, it can be shown that the induced dipole oscillates in phase with the vibrational motion of the molecule:

$$\alpha(t) = \alpha_0 + \delta\alpha\cos(2\pi\nu_{vib}t) \tag{2.22}$$

where $\delta\alpha$ is the amplitude of the change in the polarizability during one vibrational cycle, α_0 is the average value over one complete cycle, ν_{vib} is the vibrational frequency, and t is time. It should be noted that ν represents the frequency of a wave in hertz (Hz), not wavenumbers, such that:

$$\tilde{\nu} = \frac{\nu}{c} \tag{2.23}$$

Similar to rotational Raman spectroscopy, if the external electric field is due to an intense beam of monochromatic light, the combination of Equations 2.4 and 2.22 gives the following expression for a time dependent induced dipole moment:

$$\mu = \alpha_0 E_0\cos(2\pi\nu_i t) + \frac{E_0\delta\alpha[\cos(2\pi[\nu_0 + \nu_{vib}]t) + \cos(2\pi[\nu_0 - \nu_{vib}]t)]}{2} \tag{2.24}$$

where ν_i is the frequency of the incident beam, ν_o is the oscillating frequency, and ν_{vib} is the vibrating frequency. As for the case of rotational Raman spectroscopy, this equation includes

Rayleigh scattering at the same frequency as the incident light, $\alpha_0 E_0 \cos(2\pi\nu_i t)$, Anti-Stokes scattering at $\nu_0 + \nu_{vib}$, and Stokes scattering at $\nu_0 - \nu_{vib}$. The main difference here is that the frequency shift is $\pm\nu_{vib}$, rather than $\pm 2\nu_{vib}$. As a result, the vibrational Raman selection rule is based on the strong preference for $\Delta n = \pm 1$, which governs normal infrared spectroscopy.

As previously mentioned, almost all molecules are in the ground state ($n = 0$) at room temperature (due to Maxwell-Boltzmann distribution), and the predominant transition is absorption from the $n = 0$ to $n = 1$ state. This explains why the Anti-Stokes lines are very weak compared to the Stokes lines. (19) Interestingly, it is possible to determine the exact temperature of a system based on the ratio of the intensity of the Stokes lines to the intensity of the Anti-Stokes lines. (14; 20)

It is possible to solve the Schrodinger equation, with the Morse potential, to find the allowed energy levels for a vibrational Raman system (21), such that for every vibrational mode of a molecule, the energy of that mode is given by:

$$G(n) = \frac{E_n}{hc} = \left(n + \frac{1}{2}\right)\tilde{\nu}_0 - \left(n + \frac{1}{2}\right)^2 \chi_e \tilde{\nu}_0 \tag{2.25}$$

$$G(n) = \frac{E_n}{hc} = \tilde{\nu}_0 \left[1 - \chi_e \left(\frac{1}{2}\right)\right]\left(n + \frac{1}{2}\right) \tag{2.26}$$

where $G(n)$ represents the vibrational terms of a molecule (cm^{-1}), E_n is the energy level, and χ_e is the anharmonicity constant given by:

$$\chi_e = \frac{\tilde{\nu}_0}{4D_{eq}} \tag{2.27}$$

with D_{eq} representing the depth of the Morse potential. (22)

Diatomic molecules have only one vibrational frequency mode; therefore, the Raman spectrum of a diatomic molecule should consist of only one single Stokes and one Anti-Stokes line (ignoring overtones, hot bands, and rotational fine structure). However, Raman vibrational activity in polyatomics is more complicated, and it is usually necessary to apply group theory to determine if a molecule is Raman active or not. Some general rules can be applied to polyatomics; (i) if the molecule has no symmetry (e.g. HCl), then usually all of its vibrational modes are Raman active; (ii) if the molecule is symmetric (e.g. CO_2, H_2O), then it's symmetric vibrations will produce intense Raman lines, but non-symmetric vibrations are weak and often unobservable.

Fundamental: ($\Delta n = \pm 1$)

For Raman active modes, fundamental transitions can be observed. For Stokes scattering, $\Delta n = +1$, a molecule starts in the $n = 0$ state, is excited by the incident radiation, and de-excites to the $n = 1$ state. Thus the molecule has gained energy, and the scattered photon has lower

energy than the incident photon. $\Delta n = -1$ corresponds to the Anti-Stokes line; this occurs when the molecule starts in the $n = 1$ vibrational state, is excited by the incident radiation, and de-excites to the $n = 0$ vibrational state; the molecule has lost energy, and the scattered photon has a higher energy than the incident photon. A pure vibrational Raman spectrum consists of a series of intense lines to the low wavenumber side of the excitation line (laser), and a weaker mirror image series of lines on the high wavenumber side, as shown in Figure 2.5.
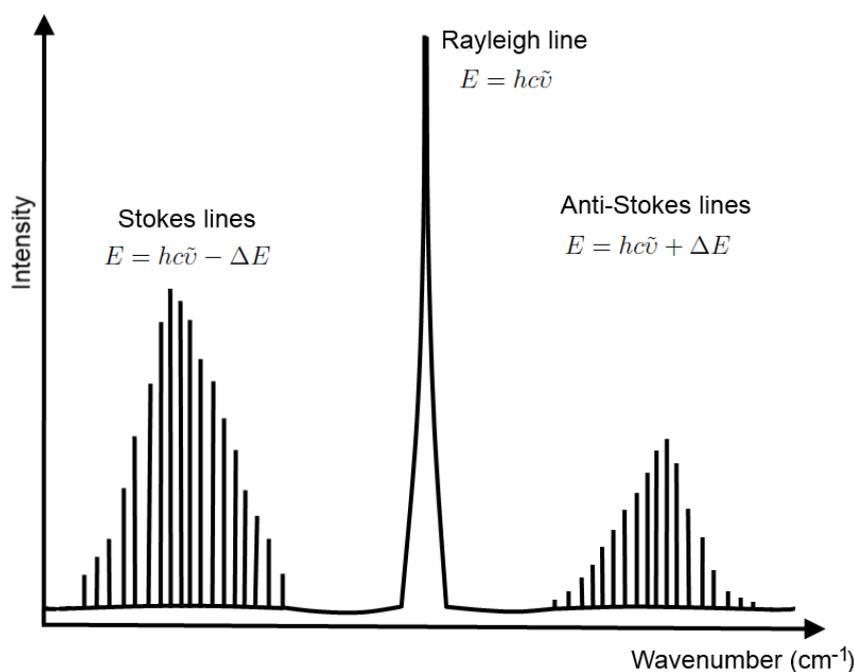
Figure 2.5: *Fundamental vibrational Raman lines, with corresponding energy equations.*

Overtones and hot bands: ($\Delta n = 2, 3, ...$)

It is also possible to observe overtone and hot bands in Raman spectra. Overtones are produced when $\Delta n = 2, 3, ..$, the first overtone is produced as $n = 0 \rightarrow n = 2$, the second for $n = 0 \rightarrow n = 3$, and so on. Hot bands are produced when $n = 1 \rightarrow n = 2, 3, 4, ...$ However, since the Raman scattering process is already a weak technique, we typically can ignore all processes such as overtones and hot bands since these are very weak, even in infrared spectra.

2.5 Rotational-vibrational Raman spectroscopy

Also known as rovibrational, or roto-vibrational, Raman spectroscopy, this topic covers the combination of rotational and vibrational Raman scattering. Since angular momentum must be conserved when a photon is emitted or absorbed, vibrational transitions are often accompanied by a change in the rotational quantum number J, resulting in a rotational-vibrational

transition. This level of detail is rarely resolved in Raman spectra, with the exception of diatomic molecules. Using the Born-Oppenheimer approximation (23), the rotational-vibrational spectrum can be represented in terms of $S(n, J)$ (combined rotational-vibrational term):

$$S(n, J) = F(J) + G(n) \tag{2.28}$$

$$S(n, J) = BJ(J + 1) + \left(n + \frac{1}{2} \right) \tilde{\nu} - \left(n + \frac{1}{2} \right)^2 \chi_e \tilde{\nu} \tag{2.29}$$
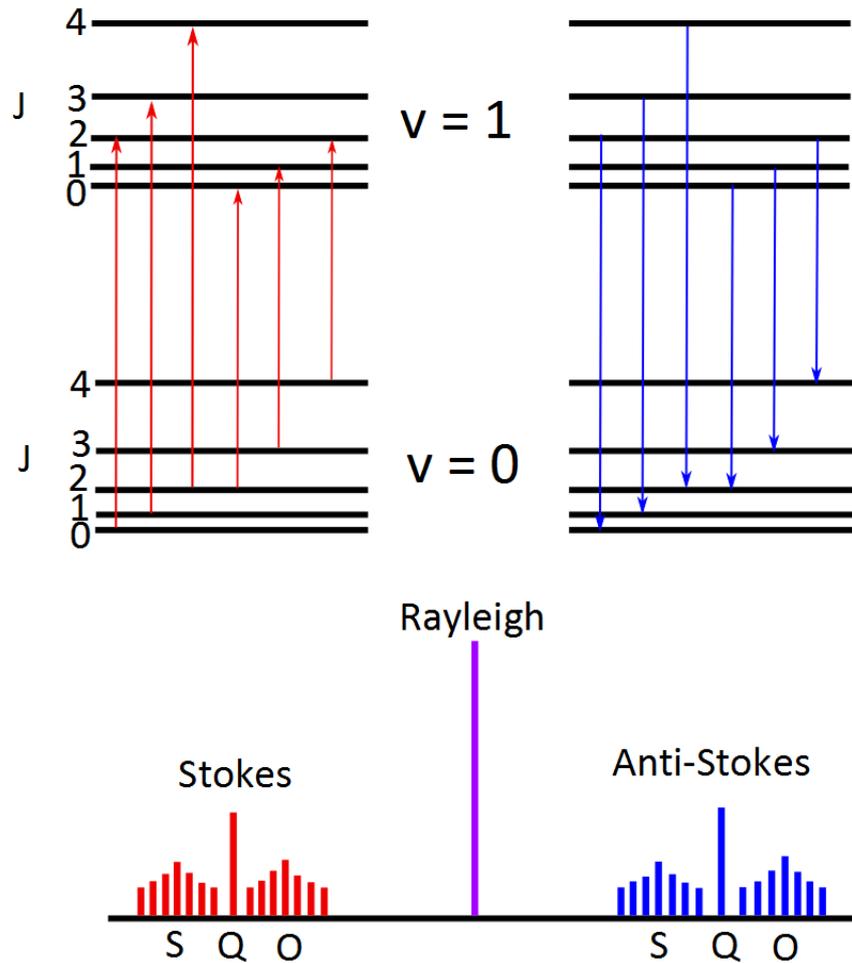
Figure 2.6: *Energy level diagram and corresponding spectrum for rotational-vibrational Raman scattering.*

For diatomic molecules, the following selection rules exist: $\Delta J = 0, \pm 2$, and $n = 0 \rightarrow n = 1$. This gives us three branches of rotational-vibrational Raman: Q-Branch, S-Branch, and O-Branch, (16) as illustrated in Figure 2.6, which are mathematically defined as follows:

- Q-Branch (rotational energy does not change):

$$\Delta J = 0 : \tilde{\nu}_Q = \tilde{\nu}_i - \tilde{\nu}(1 - 2\chi_e) \tag{2.30}$$

- S-Branch (gain rotational energy):

$$\Delta J = +2 : \tilde{\nu_S} = \tilde{\nu}_i - \tilde{\nu}(1 - 2\chi_e) + 2B(2J + 3) \tag{2.31}$$

- O-Branch (lose rotational energy):

$$\Delta J = -2 : \tilde{\nu_O} = \tilde{\nu}_i - \tilde{\nu}(1 - 2\chi_e) - 2B(2J + 3) \tag{2.32}$$

where $\tilde{\nu}_i$ is the frequency of the incident radiation (e.g. laser frequency).

2.6 Basic experimental set-up for Raman spectroscopy

A basic Raman spectroscopic set-up must consist of four main components: (i) an excitation source (i.e. laser), (ii) an optical system to deliver the light to the sample with minimum power loss, (iii) a collection system with highly efficient optics that can collect an optimum amount of Raman scattered light whilst being able to remove all Rayleigh signals, and (iv) a detection system (i.e. spectrograph and detector). A basic design of this optical method is shown in Figure 2.7, where additional optical elements can be inserted to optimise the system.

Since the scattered Raman signal is very weak relative to the laser source beam, it is important to have efficient optical systems for both delivering the laser beam directly to the sample with sufficient power, and for collecting the scattered Raman photons. To achieve this, a series of mirrors and lenses are used to guide the optical path directly to where it is needed, and filters are used to remove unwanted signals. It is also possible to include a confocal aperture to isolate the spectrum from a particular location in the sample. The detection system consists of a spectrograph and a CCD (charge coupled device) detector. Thermal energy can often generate noise and dark current on the CCD device, which can block out the distinct Raman peaks. To avoid this, and to maintain a good signal-to-noise ratio (SNR), the CCD must be cooled. Further information on the design of Raman optical systems is given in Chapter 3.

2.7 Biomedical applications of Raman spectroscopy

The different wavelengths that are emitted from a Raman scattering incident are unique to a particular molecule, which enables fingerprint-like recognition of complex molecular combinations. It has been shown that recording spectra from tissue and cells with a known pathological status allows for the training of statistical classification algorithms that enables the diagnosis of disease from unknown samples. (24; 25; 26) Various supervised and unsupervised statistical techniques have been employed to analyse Raman spectral datasets, and have been so far impressively successful in differentiating normal versus diseased samples; these will be discussed later in Section 4.6.
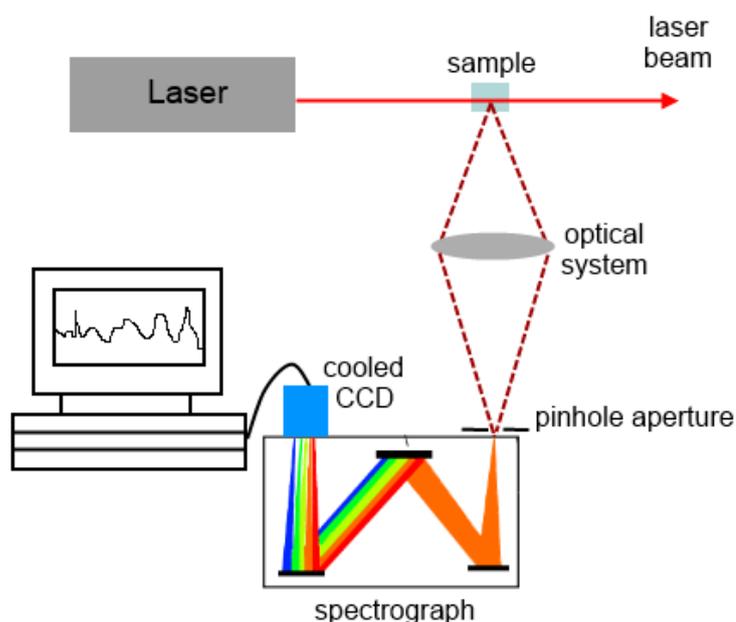
Figure 2.7: *Basic set-up used for Raman spectroscopy.*

Raman spectroscopy has many advantages over other diagnostic techniques; it is fast, giving diagnostic results in real-time, and accurate for diagnosis, but there are also limitations to this technique that need to be considered. Strong background effects are often detected that overlap with the Raman spectral irradiance, making it difficult to distinguish the signals associated with the biological sample. A variety of pre-processing statistical techniques have been developed in order to remove this noise (see Section 4.3), and a range of substrates and laser wavelengths have also been used to minimise background signals. Additionally, background signals vary significantly across source wavelength and sample substrate. (27) Good substrates for Raman spectroscopy should be transparent in the visible/NIR region, produce a low background signal, and should be biocompatible and non-toxic for the cells or tissues placed on them. For example, glass in particular is known to produce a large background signal when used in conjunction with a near-infrared (NIR) laser source, so the use of other optically transparent substrates such as calcium fluoride (CaF_2) or quartz are preferred for Raman spectroscopy in the NIR. A more detailed discussion on pre-processing methods is given in Chapter 4, the choice of lasers in Section 3.2, and sample substrates in Chapter 7.

2.8 Fourier transform infrared (FTIR) spectroscopy

FTIR spectroscopy is another form of vibrational spectroscopy that can be applied to cells and tissues for diagnostics. In Chapter 10, FTIR spectroscopy is applied to bladder and prostate cells for cancer diagnostics, and a comparison is made between the two modalities. In terms of the physics involved, FTIR relies on the absorbance, transmittance, reflection, or transflection of

infrared (IR) light. Vibrating molecules in a sample absorb distinct wavelengths of IR light that correspond directly with the vibrational frequencies of the molecular bonds within the sample. There are fundamental physical differences between FTIR and Raman spectroscopy; Raman is dependent upon a change in polarisation of a molecule, and is based upon the inelastic scattering of radiation, whereas IR spectroscopy is dependent upon a change in the dipole moment of a molecule, and is based upon monitoring of wavelengths of light that are absorbed by the molecule; whereby the absorbance can be modelled using the Beer-Lambert Law. (28)

Similar to Raman spectroscopy, there are specific selection rules that govern the allowed energy transitions that occur within IR spectroscopy. These selection rules are approximated with an anharmonic oscillator:

1. The dipole moment must change, in magnitude or direction, as the molecule oscillates

2. $\Delta n = \pm 1, \pm 2, \pm 3...$ All transitions are allowed, but $\pm n = \pm 1$ transitions are the most common

3. Angular momentum must be conserved when photons are absorbed; therefore, all transitions are rotational-vibrational with $\Delta J = \pm 1$ (although $\Delta J = 0$ is sometimes possible).

The allowed energy levels can be found with these selection rules using Equation 2.29.

An FTIR system measures the intensity of the IR signal as a function of time, and applies a Fourier transform to this signal to produce a spectrum as a function of frequency (or wavelength/wavenumbers). To achieve this, a Michelson interferometer is used (28; 29), as shown in Figure 2.8. The Michelson interferometer introduces a variable path length in the interference measurement; the interference signal is collected at many different path lengths, which is achieved by changing the position of the "movable" mirror. The recorded spectrum is reconstructed using a Fourier transform to produce a frequency domain spectrum.

There are many biomedical applications for FTIR, such as chemical imaging of tissue samples (30; 31) and cells (32; 33), as well as disease diagnostics (34; 35; 36). Applications of FTIR for the diagnosis of prostate and bladder cancer cells is investigated further in Chapter 10.

Raman spectroscopy and FTIR have many differences, these include:

- Ability to analyse aqueous solutions with Raman spectroscopy; water is a strong absorber in the IR, but it is a relatively weak Raman scatterer

- Raman spectroscopy has a higher spatial resolution which allows for subcellular analysis

- FTIR is significantly faster than Raman spectroscopy, allowing for tissue mapping to be achieved with relatively short acquisition times

- Fluorescence, or photoluminescence, can often swamp the weak Raman signal

- Raman lasers can damage tissue and cell samples during acquisition, whereas IR light sources emit within the biological window
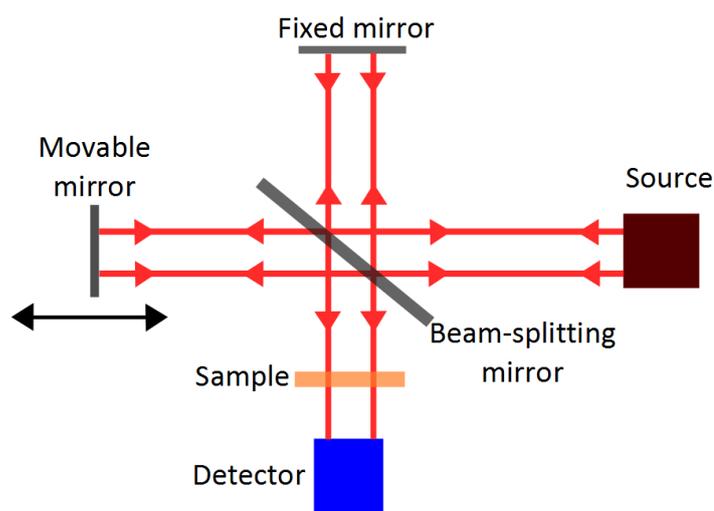
Figure 2.8: *Basic set-up used for an FTIR system, consisting of a Michelson interferometer, a source, and detector.*

- With FTIR, the whole spectrum is obtained across the entire frequency range at once; no need to change gratings to obtain different frequency ranges

- FTIR instrumentation is relatively inexpensive compared to Raman micro-spectroscopy systems

- It is possible to use glass slides with Raman spectroscopy (see Chapter 7), but the fluorescent glass signal is problematic for FTIR analysis, particularly within the fingerprint region (37).

There are advantages and disadvantages associated with each method, but typically Raman spectroscopy is applied for cellular and sub-cellular analysis, and FTIR is applied for the analysis of tissue sections.

2.9 Summary

The physics behind rotational, vibrational, and rotational-vibrational Raman spectroscopy is discussed in this chapter, in terms of both classical and quantum physics. Section 2.8 provides an overview of FTIR, and a short comparison is made between Raman and FTIR spectroscopy. The basic elements of a Raman micro-spectrometer are discussed, with a more detailed description on the design and building of Raman micro-spectrometers given in Chapter 3. The biomedical applications of Raman spectroscopy are introduced here, and are investigated in further detail throughout this thesis.

Chapter 3

Design and building a confocal Raman micro-spectrometer

3.1 Introduction

The basic Raman spectroscopy set-up discussed in Section 2.6 can be integrated into a conventional microscope, such that Raman spectra may be obtained from sub-micron locations within a biological sample, e.g. a cell or tissue sample. The name often used for such an instrument is a Raman micro-spectrometer. This can be further amended with a confocal aperture in the collection path in order to introduce a three dimensional spatial resolution, resulting in a confocal Raman micro-spectrometer; this chapter discusses the important considerations required for the design of one of these systems, particularly for the application of classifying different cell types. The overarching objective of designing this system is for the identification of cancer cells, with an emphasis on the area of urine cytology. As well as discussing the key design questions, and describing in detail the confocal Raman micro-spectrometer that was built during this research project, this chapter also briefly discusses the optical designs of Raman tweezers and Raman fiberoptic systems, which are not directly applied in this thesis, but are referenced throughout.

There are many considerations when designing a Raman micro-spectrometer for applications in cell classification: (i) the first key component is the source laser, for which all optical elements must be optimised, (ii) secondly, the integration of a microscope body and the choice of microscope objective (MO) must be considered, (iii) thirdly, collection of the Raman spectrum is achieved using a spectrograph and CCD, and (iv) fourthly, an appropriate confocal aperture can be included to tailor the spatial resolution of the system. Each of these four considerations are discussed in Section 3.2–3.5, respectively, and in Section 3.6, the design of four different Raman systems is discussed, including the custom built Raman micro-spectrometer used for much of the work presented throughout this thesis.

3.2 Source laser

When choosing a laser for a Raman system, one must consider the Raman scattering efficiency, in terms of both penetration depth and the power of the Raman spectrum, background signals generated from surrounding materials, including fluorescence from the sample itself and the substrate (see Chapter 7 for more information), photodamage, and resonance Raman effects that may be generated. Sections 3.2.1 and 3.2.2 deal with these issues in more detail in terms of the source wavelength.

3.2.1 Source wavelength

There are various advantages and disadvantages associated with a given excitation laser wavelength that must be considered in order to obtain an optimum Raman spectrum. Many materials, including the biological sample, the substrate, and the MO, produce a background signal in particular wavelength bands, which may swamp any weak Raman signals present. Some materials, e.g. glass, emit a large fluorescent background signal as the excitation laser moves from the blue region up to the red or NIR region, making a laser in the lower wavelength region of the visible spectrum preferable. On the other hand, it has been shown that Mie scattering from biological samples can also contribute to an underlying baseline signal. (38) This scattering is minimised at longer wavelengths, as is photodegradation of tissue samples. (39)

The amount of photons scattered is also indirectly related to the excitation wavelength, with the intensity of Raman lines being proportional to the fourth power of the laser frequency: $I \propto \nu^4$. (40) Therefore, when a comparison is made between a 473 nm laser and an 830 nm laser:

$$\left(\frac{\nu_{473}}{\nu_{830}}\right)^4 = \left(\frac{c}{\lambda_{473}} \times \frac{\lambda_{830}}{c}\right)^4 = 9.481 \tag{3.1}$$

it can be seen that the 473 nm laser produces Raman lines that are approximately 9.5 times more intense than those produced by an 830 nm laser for the same laser powers, assuming non-resonant conditions. Therefore, it can be seen that the scattering efficiency is higher at lower wavelengths, resulting in the use of shorter integration times and lower powered lasers in the blue/green regions, although the quantum efficiency of the grating and CCD detector being used must also be taken into consideration. Another important factor is that the optical window for biological tissues exists within the NIR region, where the absorption of light within the tissue sample is minimal. (41) Within this optical window (700–900 nm), biological samples are considered relatively transparent, and measurements in this region result in less tissue damage from the laser. (39) Furthermore, for the same reason, an NIR source can penetrate deeper into tissue and cell samples, resulting in an increase in the number of Raman scattered photons. However, this improvement is not usually significant enough to offset the reduction in scattering efficiency described by Equation 3.1.
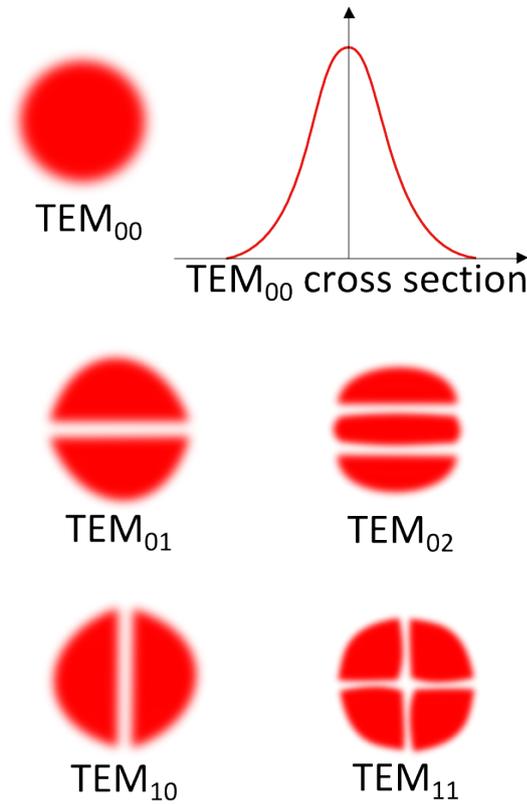
Figure 3.1: *Transverse electromagnetic (TEM) laser mode patterns*

In terms of transverse electromagnetic modes (TEM), a laser with a Gaussian beam profile (TEM$_{00}$) is preferable for a Raman micro-spectrometer, as illustrated in Figure 3.1. A TEM$_{00}$ laser allows for easier alignment, and optimum deposition of laser power into a specific targeted region, such as a cell nucleus; it also allows for a diffraction limited spatial resolution on the order of half a wavelength, provided that the optics of the delivery path are used to expand the laser beam such that $1/e^2$ width of the Gaussian profile matches the diameter of the back of the MO. (42) The effect of the source optical mode for Raman spectroscopic measurement of human tissues has previously been analysed by Li et al. (2014) (43), whereby the influence of single-mode (SM) and multi-mode (MM) excitation lasers at 785 nm and 830 nm were compared in terms of the background signal intensity generated by tissue autofluorescence and the Raman signal intensity measured. The authors reported a reduction in background, an increase in SNR, and a reduction in Mie scattering for the 785 nm SM source when compared to the 785 nm MM source. In terms of source wavelength, 785 nm spectra had a lower background intensity than that at 830 nm, and a SNR between 1.2–1.6 times higher was reported for the 785 nm MM source than that recorded from the 830 nm MM source. (43)

3.2.2 Laser – tissue interactions

The two main reactions of concern when tissues and cells are exposed to laser radiation are either thermal or optical. One form of irreversible thermal damage is coagulation due to the denaturization of cells and proteins, similar to the process in which the albumin of an egg becomes white upon cooking. For all tissue and cell types, coagulation is visibly apparent. (44) When the temperature at the laser-tissue interface exceeds that required for coagulation, vaporization occurs. Typically this happens at temperatures above 100^oC since the major component of cells and tissues is water. (45) While these effects are desirable in laser surgical procedures, such as an angioplasty, they should be avoided in Raman spectral analysis.

To consider the optical response of cells to laser radiation, Beers Law and the laser fluence rate must be investigated. Beers Law states that the fluence rate decreases exponentially as it passes through cells or tissue. (46) This attenuation of the laser beam is due to absorption and scattering, given by the following equation for a continuous laser beam:

$$E_z = (1 - r_s)E_0\exp(-\mu_t z) \tag{3.2}$$

where E_z is the attenuated beam, E_0 is the irradiance, μ_t is the attenuation coefficient, z is the thickness of the sample, and r_s is the Fresnel specular reflection for non-polarized light, which is approximately 2% for light that is incident along the normal to the air-tissue interface. (46) Since it is possible to have a significant amount of scattering and absorption, particularly at the air-cell interface due to the difference in refractive indices, a water or oil-immersion MO is sometimes used to reduce this effect. (38)

The laser power should be selected to optimise the Raman signal whilst minimising the photodamage within the biological sample. A suitable laser power should be applied, and the response of tissues and cells should be monitored upon laser irradiation for any form of photodamage. For valuable samples, the laser power should initially be turned down low, and increased carefully until an acceptable point is achieved, resulting in a maximum Raman signal without any sample damage. (47)

3.3 Microscope body

Raman measurements of biological samples sometimes require long acquisition times, in the order of 10-60 s. The stability of the laser focus position at the sample therefore needs special care. When building a Raman micro-spectrometer, one can make use of a commercially available microscope body, such as the inverted Nikon phase contrast microscope (Nikon, Toyko, Japan; Diaphot 300), which we have incorporated into our system (Section 3.6.1, Figure 3.4). In this case, the laser is directed into the microscope using the fluorescence cube cavity directly below the MO. In order to maintain full imaging functionality of the microscope after the addition of Raman optical filters into the system, it is important to choose a microscope that

employs infinity MOs. The reason for this is that the addition of filters into the imaging path of the microscope changes the optical path length, and if this occurs in any domain other than in an infinity domain, the focal plane of the microscope will be slightly altered, and the image will appear out of focus.

The Nikon Diaphot 300 was one of the first research grade microscopes that adapted the concept of infinity optics, such that fluorescence cubes could be readily inserted into the cavity immediately below the MO turret. However, the microscope did not actually employ infinity MOs that are commonplace today. Instead, it was the last of its kind to employ fixed focal length MOs with a focal length of 160 mm. A convex lens was glued into the MO turret immediately below the MO in order to transform the light into an infinity domain before entering the cavity. In order to employ modern infinity MOs, the entire turret of the microscope is replaced with a turret that does not contain a convex lens.

With a Raman micro-spectrometer, the biological sample should be placed on a stable translation stage; this translation stage should be movable in the horizontal xy plane, as well as in the z direction. To achieve movement in the z plane, it is possible to move either the translation stage or the MO. For our custom-built system, we have incorporated the former approach; more details about the specific translation stage used are provided in Section 3.6.1.

Alternatively, one could build a stable Raman system using cage optics, whereby the optical elements (lenses, mirrors, filters, etc.), cameras, and light sources are the same as those employed in a standard custom built Raman system, with the distinct difference of having them connected with cage optics. (48) There are many advantages of using cage optics such as ease of laser alignment, stability, ease of construction, and convenience. However, the associated disadvantages include lack of freedom and flexibility (cannot send beams at arbitrary angles). A cage optics Raman micro-spectrometer was also built during the course of this PhD, but the details of that system are not provided here.

It should be noted that when using an excitation wavelength in the NIR, it may be preferable to use a fluorite MO in order to reduce the fluorescent background signal emanating from the objective. These MOs, which are produced by the large optics companies, often have various names, e.g. fluor, fluorite, fluortar, neofluar, and are manufactured using a mineral form of CaF_2. These MOs have the desirable property of optimal correction for chromatic aberration (allowing different wavelengths to come into focus at the same plane), which is important in the context of focusing the Raman scattered light onto the spectrograph slit. The background signal from the MO can also be reduced by using a suitable confocal aperture (27), as discussed further in Section 3.4.

3.4 Spectrometer and CCD detector

The performance of a Raman micro-spectrometer depends, among other things, on its sensitivity and spectral resolution. The sensitivity is defined as the probability of detecting a scattered

photon, and determines how long it will take to record a spectrum with a given SNR. The sensitivity is determined by numerous factors, including (i) the grating reflectivity, also known as the diffraction efficiency, (ii) the optical transmission of the various optical elements in the collection system such as lenses, filters, beamsplitters, etc., and (iii) the quantum efficiency of the CCD detector. The spectral resolution determines the accuracy of extracting subtle information from a Raman spectrum, as discussed further in Section 3.3.1. The spectral resolution is determined by the focal length of the spectrometer and the line density of the grating used to disperse the light. (49)
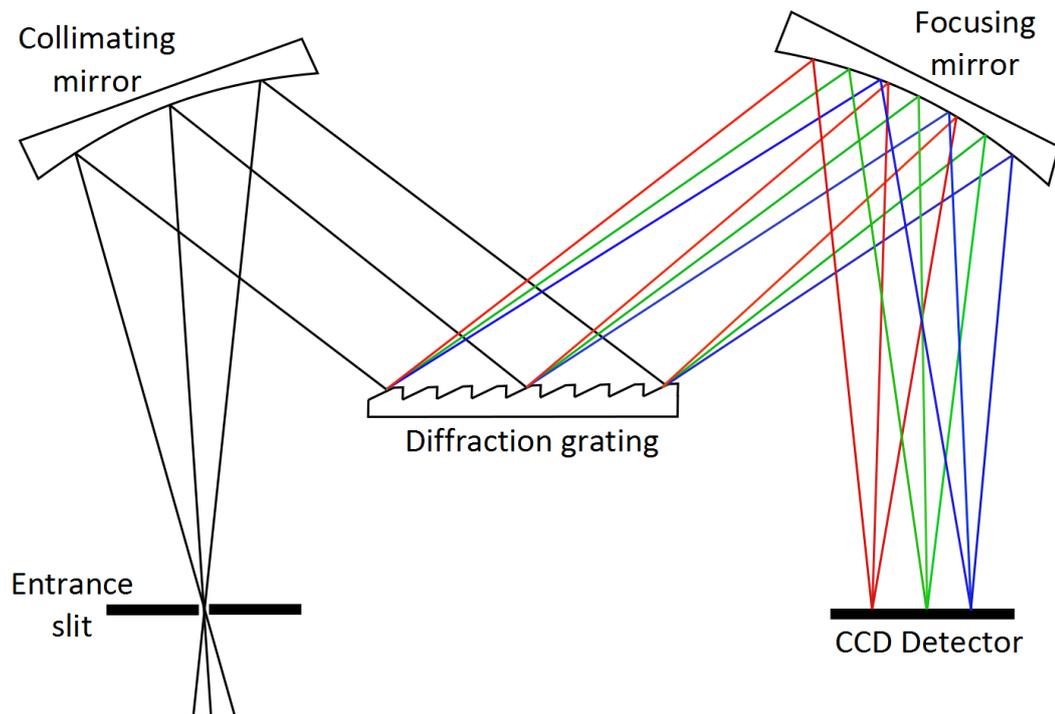
Figure 3.2: *Schematic of a Czerny-Turner spectrograph.*

A spectrograph is designed to accept light with a broad spectrum of wavelengths, to separate the wavelengths spatially, and to detect each wavelength on a multichannel detector, which today is synonymous with a CCD. The spectrograph used for our custom Raman microspectroscopy system is the Andor Shamrock 500 (Andor Technology Ltd., Belfast, UK), in association with an Andor iDus cooled CCD camera (DU420A-BR-DD, Andor Technology Ltd., UK) operating at -80°C. This spectrograph is based on a Czerny-Turner system, as shown in Figure 3.2. Raman signals are focused onto the entrance slit of the spectrograph, where it diverges until it reaches a concave (or collimating) mirror whose focal length corresponds to the distance between the mirror and slit; after being reflected by the mirror, the light is collimated. This collimated beam then hits the diffraction grating, which is an array of finely spaced lines on a reflective surface, where constructive and destructive interference occurs, a phenomenon which is both wavelength and angle dependent. Consequently, each wavelength is reflected at a

slightly different angle onto the focusing mirror. This results in each wavelength reaching a different pixel on the array detector (CCD). The CCD consists of an array of photodiodes (i.e. light sensitive pixels) that can convert photons into electrons, allowing for accurate measurement of the number of photons incident onto the CCD for a particular wavelength.

Additionally, it is important to align, and expand the laser beam, such that it matches the f number of the spectrograph. To achieve this, the f number can be calculated as follows:

$$f = \frac{F}{D} \tag{3.3}$$

where F is the focal length of the lens, and D is the diameter of the laser beam that is illuminated onto the lens, as shown in Figure 3.3.
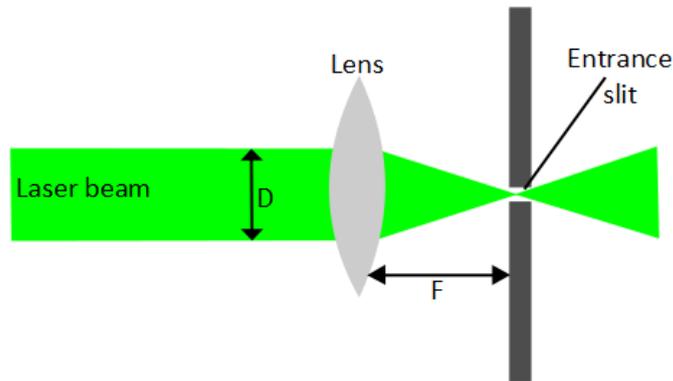
Figure 3.3: *An illustration for the calculation of the f number for a given spectrometer.*

3.4.1 Spectral resolution

As mentioned in the previous section, the spectral resolution of the system is determined by the diffraction grating, the detector, and the focal length of the spectrometer. The wavelength bandwidth that can be observed at the output of a spectrometer is a function of the angular dispersion and the focal length of the spectrometer. Since the angular dispersion is itself a function of wavelength, the observable bandwidth differs depending on the centre wavelength that is chosen. The Andor Shamrock 500 spectrograph enables rotation of the grating, which allows for the selection of different regions of interest. Precise calculation of this bandwidth for a particular centre wavelength can be achieved using the grating equation, and requires knowledge of the various spectrograph parameters such as grating angle, focal length, output aperture size, and grating period. (50) For gratings with relatively low dispersion, e.g. 300 grooves per mm (gr/mm), the observable wavelength bandwidth (and therefore resolution in nm) is approximately constant for different centre wavelengths ranging from 500 nm to 900 nm, though this approximation fails for highly dispersive gratings. However, in Raman spectroscopy, where

units such as Raman shift or wavenumbers are more often used, the spectral resolution (in wavenumbers) of the system increases with the source wavelength. When the wavelength is shifted through Raman spectroscopy from the incident wavelength λ_i to the scattered wavelength λ_s, the shift in wavelength $\Delta\lambda$ is given by:

$$\Delta\lambda = \lambda_i - \lambda_s \tag{3.4}$$

but the corresponding wavenumber shift ($\Delta\tilde{\nu}$) is given by Equation 2.1, therefore,

$$\Delta\tilde{\nu} = \left[\frac{\Delta\lambda}{\lambda_i(\lambda_i - \Delta\lambda)} \right] \tag{3.5}$$

Assuming that the wavelength bandwidth, $\Delta\lambda$, remains constant regardless of the centre wavelength chosen, it follows that the ratio of spectral bandwidths for two different incident wavelengths, λ_{i1} and λ_{i2}, will be given by:

$$\frac{\Delta\tilde{\nu}_1}{\Delta\tilde{\nu}_2} = \frac{\lambda_{i2}(\lambda_{i2} - \Delta\lambda)}{\lambda_{e1}(\lambda_{e1} - \Delta\lambda)} \tag{3.6}$$

Thus, for a sample case of $\Delta\lambda = 150$ nm, it can be concluded that 473 nm excitation wavelength results in approximately 3.7 times more spectral bandwidth than 830 nm. Conversely, the resolution at 830 nm is $\sim$3.7 times smaller than that at 473 nm. The overall diffraction efficiency of a grating also depends on the blaze angle of the grating, and the polarization of the incident light, both of which are not considered here. Typically, gratings are selected in order to provide the best resolution within a specific acquisition time frame over the desired spectral range for the chosen laser wavelength.

3.5 Confocal aperture

Another important consideration is the use of a confocal aperture within a Raman microspectrometer. The confocal aperture controls how much of the Raman spectral irradiance passes through into the spectrograph slit and onto the detector. Its primary function is to achieve spatial resolution for a given spectrum, as well as to remove any background signals emanating from the MO or other optical elements that are common to both the delivery and collection paths. Confocal apertures range in size, typically from small (10–20 μm) up to large (150–300 μm). The larger the aperture, the greater the Raman signal reaching the CCD, however larger apertures also degrade the spatial resolution of the system. In a well designed system, the size of the confocal aperture is matched to the magnification of the MO, but some experimentation can be done to determine the optimum aperture for specific purposes.

The size of the confocal aperture should in turn be approximately matched to the size of one pixel area in the CCD camera. This is achieved with an appropriate design of the imaging

system from the confocal aperture to the spectrograph slit. For example, for a confocal aperture of 100 μm and a pixel size of 25 μm, the magnification factor of the imaging system should be 0.25. The f number of the spectrograph must also be considered when designing this imaging system, such that the angle of light entering the slit of the spectrograph does not exceed the largest acceptance angle for that spectrograph.

3.6 Raman system designs

There are many types of Raman systems, both commercial and custom built, which have many advantages and disadvantages. Careful consideration should be made in order to design a Raman system suitable for specific experimental needs. Here, four different Raman systems are compared, including a custom built system from the Biophotonics Lab in Maynooth University, the commercial Horiba LabRam HR (Horiba Jobin Yvon Inc., US) at the Focas Institute (Dublin Institute of Technology, Ireland) which is used to perform various experiments for this thesis, a standard Raman tweezers system, and a fiberoptic Raman system. The latter two types of Raman system were not employed in any experiments associated with this thesis. However, they are directly relevant to a number of discussions in the thesis, and are often used in the analysis of tissue and cell samples for disease diagnostics. For these reasons, a brief discussion on these system designs is included.

3.6.1 In-house custom Raman micro-spectrometer

A custom Raman micro-spectroscopy system is designed and built as part of this thesis. This system consists of a 532 nm laser (150 mW; TEM$_{00}$; 1 mm beam diameter, 0.4 mrad beam divergence; Torus, Laser Quantum, UK), a spectrograph operating with a 600 lines.mm^{-1} grating, and a cooled CCD camera (outlined in Section 3.4), as illustrated in Figure 3.4. This basic design can be found in many commercial Raman micro-spectrometers from long established companies such as Horiba (Horiba Jobin Yvon Inc., US). A state of the art Raman microscope will cost in the order to \$100k, but an equivalent one can be built from basic elements for less than half this amount.

The design of our system is as follows: the laser, with a narrow bandwidth, is first passed through an optical isolator (OI) which prevents back reflections in the set-up from returning into the laser cavity and damaging the laser head. This is followed by a line pass filter (LP) which removes any spurious lines from the laser that may result from additional modes being present, allowing only the 532 nm line to pass through. A neutral density filter can be inserted after the LP, which can be rotated to vary the power of the laser beam. Two lenses (L1, L2; Thor Labs, US) are used to expand the laser beam such that the diameter of the collimated beam is approximately equal to the back aperture of the MO (Olympus 50x, 0.8 NA, UMPlanFl). The collimated beam is directed onto a dichroic beam splitter (DB; Semrock, US; LPD01–532RS),

which reflects light at 532 nm, but transmits longer wavelengths, i.e. Raman Stokes scattered photons.

The DB reflects the beam towards a dichroic short pass filter (DS; Edmund Optics, US; 69-202) reflecting all wavelengths above 500 nm, and transmitting shorter wavelengths. The DS is located within the fluorescence cube cavity of an inverted Nikon phase contrast microscope (Nikon Diaphot 300), and is carefully aligned such that it reflects the beam directly into the back of the MO. The MO is chosen such that it has a high numerical aperture (NA), which is essential to recover as many Raman back scattered photons as possible. It is also essential that the MO efficiently transmits visible light as this will also impact on the system's ability to recover a strong Raman signal. The sample is placed on an xyz translation stage (TS, ASI, LS–2000 and LS–50) that enables sample positioning relative to the laser spot. In the case of Raman diagnostics, it is desirable to position the cell such that the spot is located within the cell nucleus, or nucleolus, as this results in optimum diagnostic sensitivity. (51)

The back scattered Raman photons (532 nm – 650 nm) propagate back through the MO, are reflected by the DS, and then pass through the DB. Lens L3 focuses the light to a 100 μm confocal aperture (Thor Labs). Lens L4 collimates the beam, with lens L5 focusing the light onto the entrance slit of the spectrograph. A long pass filter (LgP; Semrock, UK; LP03–532RU–25) is placed near the entrance slit in order to further reduce any Rayleigh scattered/laser light that has reached the spectrograph. Via a Czerny-Turner spectrometer, the image of the slit is projected onto the cooled CCD camera, and the Raman spectrum is recorded using the Andor Solis software system. It is essential to use a high quantum efficiency, low noise, cooled CCD in order to recover the weak Raman signal with minimal noise.

The various lenses in this system are chosen such that (i) a laser spot size of ∼3 μm is located at the object plane of the microscope, (ii) the 100 μm confocal aperture is imaged to the spectrograph slit with a demagnification factor of 4, such that the confocal aperture is matched to a single camera pixel (26 μm), and the angle of the light entering the slit falls within that allowed by the f number of the spectrograph (for the Andor Shamrock, $f = 6.5$).

Independently, light from the 100 W halogen lamp (HL; Nikon, Japan) is focused onto the sample and into the MO. Since some of this light has a shorter wavelength than 500 nm, it passes through the DS. This light passes into the body of the microscope where it reflects from a mirror through an internal imaging system, and images onto a digital camera (DC; Basler AG, Germany; acA2000–340km) fixed to the front exit port of the microscope.
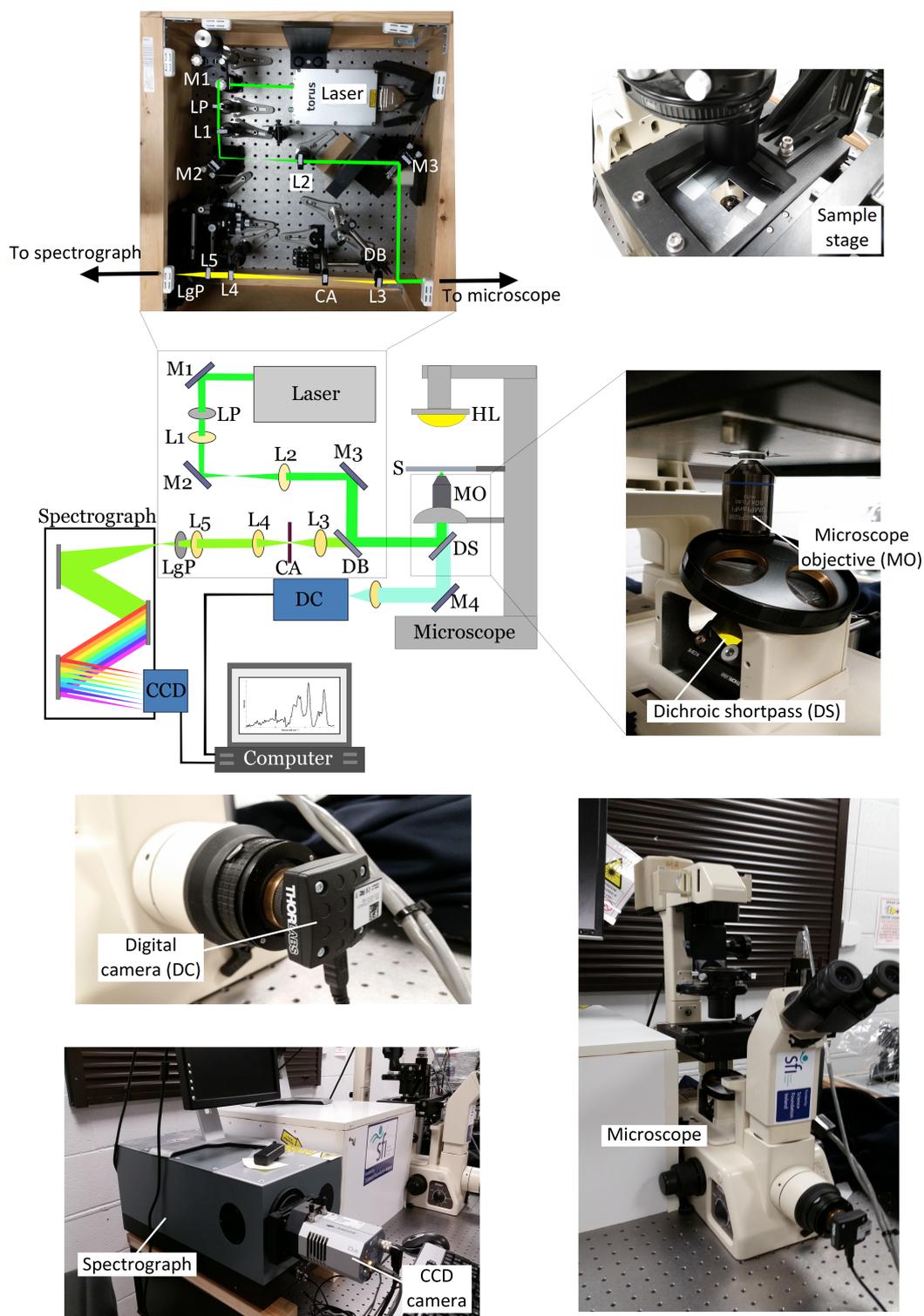
Figure 3.4: *Schematic of the custom built Raman spectroscopy set-up in the Biophotonics Lab in Maynooth University, along with images of individual components. LP, line pass filter; L, lens; M, mirror; DB, dichroic beamsplitter; DS, dichroic short pass filter; MO, microscope objective; S, sample stage; HL, halogen lamp; CA, confocal aperture; LgP, longpass filter; DC, digital camera.*

3.6.2 Commercial Horiba LabRam HR

Commercial Raman micro-spectroscopy systems are widely available from companies such as Renishaw (Renishaw plc, UK), Horiba (Horiba Jobin Yvon Inc., US), and Ondax (Ondax Inc., US). The Horiba LabRam HR 800 system, which is employed across studies within this thesis, is an integrated, robust, and high performance Raman micro-spectroscopy system, as shown in Figure 3.5. There are many advantages to this commercial device, such as high spectral resolution, multiple laser and detector options, along with a flexible choice of gratings and aperture sizes.
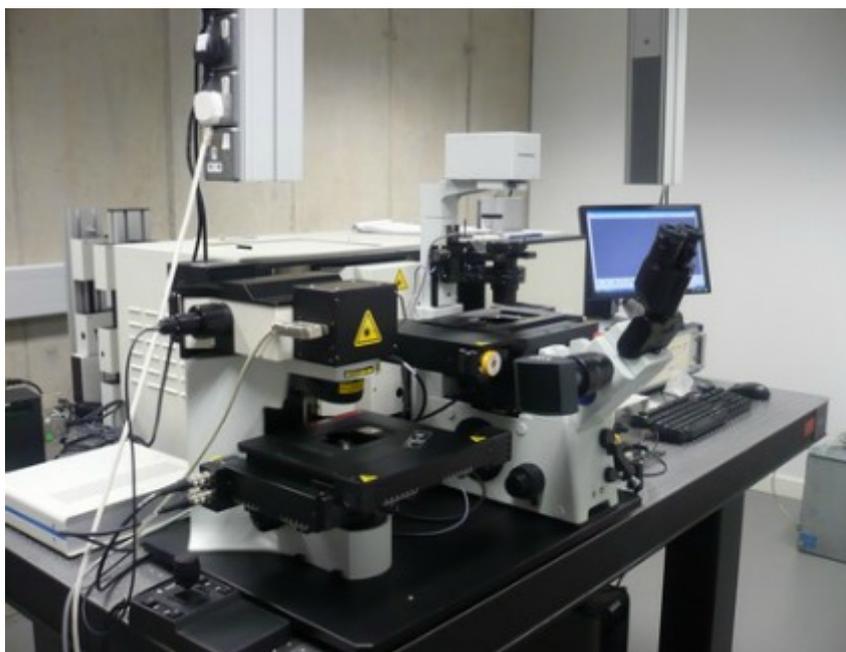
Figure 3.5: *Image of Horiba LabRam HR 800 dual microscope system at the Focas Institute, Dublin Institute of Technology, Ireland.*

The LabRam HR is operable for Raman micro-spectroscopy with five different source wavelengths, as outlined in Table 3.1. These source wavelengths range from 473 nm (blue) through to 830 nm (NIR), allowing for a direct comparison between Raman spectra across different wavelengths, as investigated further in Chapter 7.

Wavelength	Power	Details
473 nm	50 mW	Solid state diode laser
532 nm	50 mW	Solid state diode laser
660 nm	100 mW	Solid state diode laser
785 nm	300 mW	CLDS point mode diode laser
830 nm	200 mW	CLDS point mode diode laser

Table 3.1: *Details of the lasers employed within the Horiba LabRam HR 800 commercial system.*

3.6.3 Raman optical tweezers

Although a Raman tweezers system is not employed in the course of this work, these systems are often used in similar studies, and for this reason, a brief description is included here. Raman tweezers are based on combining the principle of Raman micro-spectroscopy with optical trapping (52; 53), as illustrated in Figure 3.6. Optical trapping, or optical tweezing, is a technique that utilises a high NA MO, usually configured in an inverted microscope, together with a suitable laser, to trap a cell in the waist of focused laser beam. The tightly focused light creates a sharp gradient of intensity which leads to gradient forces trapping the cell. The trapped cell can be suspended in a liquid (or air) environment, and moved in three dimensions by manipulating a component in the optical set-up. By combining Raman micro-spectroscopy with optical trapping, individual cells can be biochemically probed under physiological conditions, or in microfluidic chips.
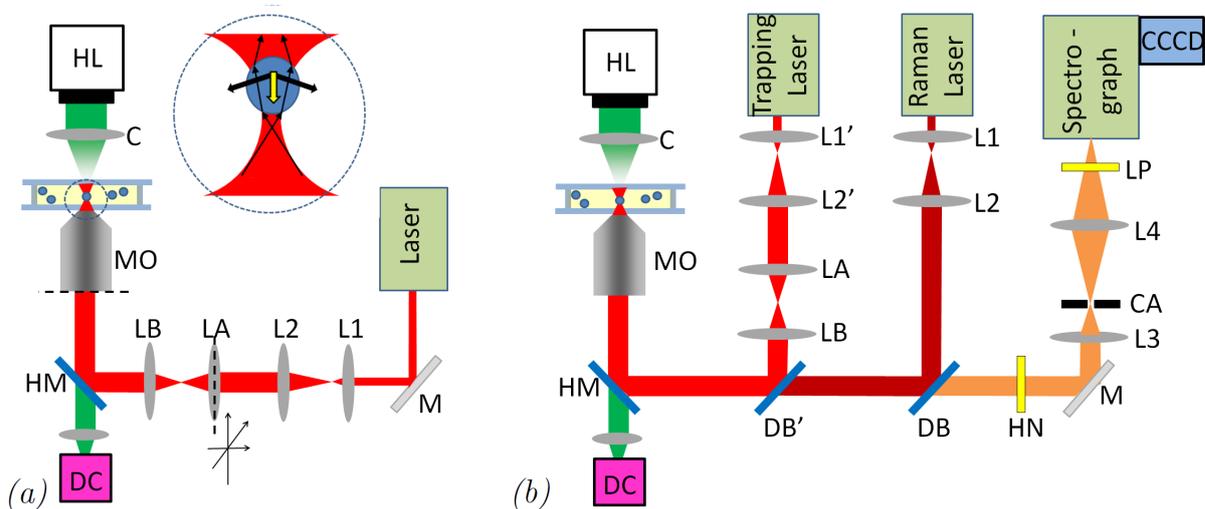
Figure 3.6: *(a) Schematic of a standard Raman optical tweezers set-up, (b) schematic of a dual beam Raman tweezers set-up. HL, halogen lamp; C, condenser; MO, microscope objective; L, lens; M, mirror; DC, digital camera; DB, dichroic beam splitter; CA, confocal aperture, LP, long-pass filter.*

Raman tweezers offer exciting new avenues of research, particularly in the areas of microfluidics and urine cytology. This technique can be used for investigating the biochemistry of living cells in their natural environment with 3D spatial resolution. (54) It can also be used for diagnostic classification of bladder and prostate cancer cells, as shown elsewhere (55; 56; 57; 58).

3.6.4 Raman spectroscopy with a fiberoptic probe

Similar to Raman tweezers, a fiberoptic Raman spectroscopy system is not applied within this thesis, but it is discussed as an alternative diagnostic approach. A fiberoptic Raman set-up consists of a specialised fiber probe, an excitation source (laser), fiber couplers to deliver the light in and out of the fiber, filters to remove unwanted signals, and a detection system (spectrograph

and cooled CCD camera), as can be seen in Figure 3.7. This fiberoptic probe can be inserted during a cystoscopic diagnostic procedure, and be used to give real-time analysis and classifications of the bladder wall. Similar to conventional Raman micro-spectroscopy, the Raman fiberoptic probe can be used to monitor in vitro and ex vivo specimen also. The main advantages of this system include the potential to replace the need for multiple biopsies, reductions in the turn-around time for diagnosis, reductions in histological/pathological costs, and improved surgical procedures by using the probe to identify border regions of diseased or damaged tissue.
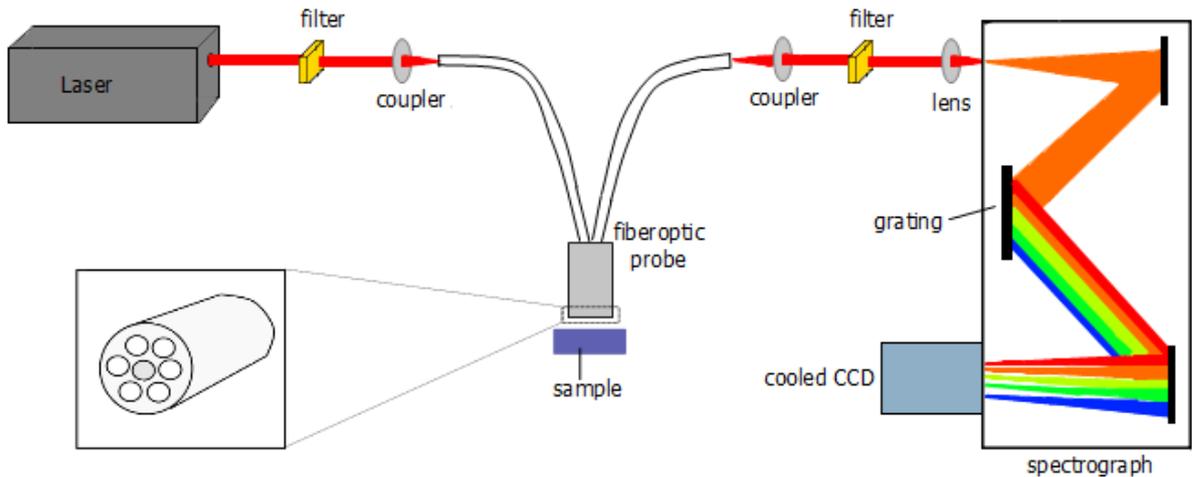
Figure 3.7: *Schematic of a basic Raman fiberoptic probe set-up, which can be applied to biological samples both in vivo and in vitro, with a front view of the design of a fiberoptic probe inset, showing the central delivery fiber which is surrounded by several collection fibers. It is of note that it is often more conventional to integrate the filters into the tip of the fiberoptic probe instead, but for ease of illustration, they have been shown in the positions above.*

A Raman probe for cystoscopic or endoscopic diagnostics must be able to fit into the instrument channel of a standard cystoscope (or endoscope or catheter), be biocompatible, and robust enough to withstand decontamination/disinfection processes. NIR lasers (785–830 nm) are generally used for Raman fiberoptic probes since they minimise thermal damage to the tissue sample being analysed. It is also essential that there are low fluorescence, or photoluminescence, signals generated within the probe. The general design of a Raman fiberoptic probe is shown in Figure 3.7; consisting of a central delivery fiber that is surrounded by several collection fibers, where the diameter of each fiber is between 100 and 400 μm depending on the system design. Typically a bandpass filter is placed at the tip of the excitation fiber to remove unwanted background signals, and a longpass filter is placed at the tip of the collection fibers to prevent Rayleigh scattered light from entering these fibers. Lenses (such as a ball lens, made from sapphire) at the probe tip or tapered fiber tips can also be used to improve the efficiency of delivering both the excitation light to the tissue and Raman scattered light into the collection fibers (further information is available in Ref. (59)).

3.7 Summary

Each of the key components of a Raman micro-spectrometer have been discussed here. Important considerations for the design of a Raman system include the source laser, the microscope body, choice of spectrometer and CCD, and the utilisation of a confocal aperture. As has been highlighted in this chapter, it is particularly important to choose a laser that is suitable for the application involved, such as a laser with a TEM_{00} Gaussian profile. Other considerations involve choosing an NIR laser in order to minimise the possibility of tissue damage, or choosing a laser in the visible wavelength region (e.g. 532 nm) if shorter acquisition times are desired. It is also important to choose a spectrograph and CCD camera that have good quantum efficiencies within the desired acquisition range.

The custom Raman micro-spectrometer that was designed and built in our laboratory has been described in Section 3.6.1, with information given on each of the optical components within the system, such that the design could be replicated in another laboratory. Further details on commercial Raman systems, along with a Raman tweezers set-up and a fiberoptic Raman system have also been provided.

Raman spectroscopy is an opto-electronic technique, resulting in complex spectra that require the development of pre-processing and multivariate statistical algorithms in order to achieve reliable and reproducible spectra, which are necessary for the development of an accurate diagnostic classifier. Such algorithms involve accurate system calibration (post-acquisition), removal of unwanted background signals, and noise reduction, as discussed discussed in further detail in Chapter 4.

Chapter 4

Numerical methods

4.1 Introduction

In the previous two chapters, the physics of Raman spectroscopy was discussed, followed by a detailed discussion on the design and construction of a confocal Raman micro-spectrometer. However, Raman spectroscopy is not a purely optical technique, it also incorporates a suite of numerical processing techniques designed to reduce the impact of noise, eliminate unwanted background artifacts, and standardise the output spectra, in terms of calibration and normalisation, such that the spectra can be reliably compared. As high quality spectroscopic equipment has become available for the analysis of chemical, pharmaceutical, and biological samples, there has been a parallel development of multivariate statistical tools for the analysis of the corresponding spectra. These statistical tools are focused on identifying subtle variations in spectral datasets and exploiting these patterns for the purpose of classification.

Raman spectra consist of three main components: Raman signals, a baseline, and noise. The presence of these unwanted baseline and noise signals is often reduced using pre-processing techniques before the classification process as they can lead to a larger diagnostic error otherwise. Section 4.2 discusses the calibration of Raman datasets, and in Section 4.3 the reduction of noise is discussed. Baseline correction methods are reviewed in Section 4.4, followed by a discussion of normalisation methods in Section 4.5. Finally multivariate classification algorithms are reviewed in Section 4.6, with a summary provided in Section 4.7.

4.2 Calibration

Small changes in the optical set-up (e.g. rotation of a grating), and/or a change in temperature can result in a shift in the Raman spectrum along the wavenumber axis. Furthermore, the optical components, such as the grating and camera, as well as the optical lenses and filters, can have a significant impact on the intensity values of the spectrum recorded by a given Raman micro-spectrometer. Accurate calibration is therefore required to allow for a fair and mean-

ingful comparison of Raman data obtained from different Raman micro-spectroscopic set-ups. This is particularly important when creating standard medical diagnostic databases for disease identification. It is possible to perform calibration on the system itself, or retrospectively, to the spectral datasets after acquisition. To perform calibration on a dataset, the wavelength axis should first be calibrated, followed by intensity calibration, and finally, wavenumber calibration.

Wavelength calibration is performed in order to assign exact wavelength positions to each column of pixels along the CCD camera. There are two causes for nonlinearity in the wavelength dispersion across the CCD caused by the spectrograph grating. The first is inherent in the sine function that describes wavelength dispersion (using the grating equation), while the second is the result of distortions that are present in all spectrograph optical designs. Theoretically, a one- or two-point calibration could be sufficient to compensate for the first cause of nonlinearity, however multi-point calibration is required to correct adequately for distortion and permit spectrograph-to-spectrograph comparisons. (60) To achieve this, the spectrum from a known sample, such as a Neon lamp, is measured. The spectrum of Neon contains many spectral peaks, and the position of each peak is calibrated against a reference spectrum, which contains the precise wavelength position for each spectral line. (61)

Intensity calibration is then performed to correct for the wavelength-dependent transmission efficiency of the optical elements within the system, as well as correcting for the sensitivity of the detector itself. This can be achieved using a National Institute of Standards and Technology (NIST) calibrated white light source (61; 62), or using the fluorescent signal from green glass. (63)

Finally, wavenumber calibration is performed. This step is particularly important if the source wavelength is not accurately known. The purpose of wavenumber calibration is to assign a particular wavenumber shift to each column of pixels in the CCD camera. (64) To perform wavenumber calibration, a known sample such as silicon is recorded, and the position of the peak(s) obtained using Raman spectroscopy are calibrated with the precisely known Raman peak positions for the calibration sample used. Further information on Raman system calibration can be found elsewhere. (61; 65; 66; 67)

It should be noted that whilst such a rigorous approach is necessary for the sharing of data across different systems, the datasets analysed within this thesis have only been wavenumber calibrated based on the position of the silicon peak. In order to minimise intensity artifacts, spectra were recorded within relatively short time periods for each study, i.e. within one or two weeks, without any system adjustments, and with the room temperature maintained as constant insofar as possible.

4.3 Noise

There are four main sources of noise to consider with Raman spectral acquisition when using a CCD detector. These are (i) shot noise, which arises from statistical fluctuations in the

number of photons collected by the CCD pixels; (ii) dark current, caused by thermally generated electrons within the CCD pixels; (iii) readout noise, generated by the electronic circuitry in the CCD; and (iv) cosmic ray artifacts, originating from high energy particles from outer space. (68; 69; 70) In order to achieve a suitable SNR for a given dataset, these sources of noise should be minimised insofar as possible.

Smoothing is often applied to Raman spectra to remove high frequency noise signals. A Savitzky Golay filter is often used, which consists of a moving window based polynomial fitting procedure. (71) A Savitzky Golay filter requires specific parameters such as the size of the moving window and the polynomial order. As the window size increases, some of the Raman bands may disappear, so it is important to choose appropriate parameters for the data being smoothed. Alternatively, a Gaussian or median filter can be used for noise removal instead. (60)

During Raman acquisition, particularly with long duration measurements, it is common to find cosmic ray artifacts in spectra. Due to their nature, cosmic ray artifacts occur randomly in time and space, and can be identified as intense narrow spikes present on Raman spectra. Cosmic rays are often removed from raw spectra by collecting two spectra for each experiment and comparing them on a pixel by pixel basis. If the difference exceeds a predetermined threshold, then the presence of a cosmic ray is assumed, and the lesser value is used for that pixel. (72) Otherwise the average value is taken, thereby taking full advantage of the SNR offered by the full acquisition time of both recordings.

4.4 Baseline correction

The baseline of a spectrum is described as the slowly varying curve seen in the lower range of the spectrum without the distinct Raman peaks. This baseline can be varying or sloped, and is most often ascribed to an auto-fluorescence from the sample itself. (73; 74) However, the source of this signal is debatable, with some authors suggesting that it may originate from sample morphology and Mie scattering of the source laser wavelength. (75) Regardless of its origins, it is possible to characterise the signal as a slowly varying polynomial. It is necessary to remove this baseline signal insofar as possible in order to facilitate an accurate comparison of related spectra, in particular for the application of chemometrics for the classification of a given sample spectrum. The three baseline correction algorithms that have been applied to the data within this thesis are discussed in Sections 4.4.1–4.4.3.

4.4.1 Modified polynomial

The modified polynomial baseline correction method, as proposed by Lieber et al. (73), is based on the subtraction of a polynomial using an ordinary least squares (OLS) fitting approach. (76) In an automated approach, the Raman spectrum is first fitted with an N order polynomial using OLS. The values of the spectrum that lie above the polynomial are set equal to the value of the

polynomial; the resultant signal is again fitted with an N order polynomial, and the process is iteratively repeated, until a point is reached such that the polynomial lies directly underneath the Raman spectral peaks. This method requires two input variables, the order of the polynomial N and the number of iterations for the algorithm to run; both of which are typically dataset dependent.

4.4.2 Modified polynomial plus known background signal

An algorithm is used in Chapter 7 of this thesis which is based on some minor changes to the method proposed by Beier et al. (2009). This algorithm is based on the subtraction of a known background signal and an N order modified polynomial fit (77), which can be used to remove both the baseline and any unwanted background signals present within Raman spectra, such as the background glass signal present in spectra recorded from cells on glass slides. (78; 27)

Firstly, two or three spectra are recorded from the substrate (or another known contaminant), and an average background spectrum $B_0(\lambda)$ calculated. Since this is a discrete numerical algorithm, it is more appropriate to describe the signals in terms of their discrete representations; where $\lambda \rightarrow n\delta\lambda$ and n takes integer values from $0 \rightarrow N - 1$, where $\delta\lambda$ denotes the sampling interval of the recorded spectrum and $N\delta\lambda$ is equal to the bandwidth. A Gaussian smoothing function is applied to the background spectrum to reduce the presence of noise. Equation 4.1 represents the process of discrete convolution with a Gaussian filter that is sampled at the same rate as the spectrum. The width of the Gaussian function is taken to be $M\delta\lambda$, which corresponds to the region where the function has appreciable values, for example if $M\delta\lambda = 6\sigma$, then the region will contain over 99% of the Gaussian signal's energy.

$$B_0(n\delta\lambda) = \sum_{m=0}^{N-1} B_0(m\delta\lambda) \exp\left[\frac{\delta\lambda^2(m-n)^2}{2\sigma^2}\right] \tag{4.1}$$

where σ is the standard deviation, and n takes the same range of values as previously mentioned. This equation can be calculated by applying zero padding followed by multiplication of the Discrete Fourier Transform (DFT) of the signal and that of the Gaussian filter followed by an inverse DFT.

A range of different weights, or concentrations C, of this background signal $B_0(n\delta\lambda)$ are subtracted from the recorded cell spectrum $X_0(n\delta\lambda)$, followed by the subtraction of an N order polynomial $P(n\delta\lambda)$, which is generated using Matlab's $polyfit$ function, resulting in a spectrum, $R(n\delta\lambda)$, consisting of residual values:

$$R(n\delta\lambda) = [X_0(n\delta\lambda) - (C \times B_0(n\delta\lambda))] - P(n\delta\lambda) \tag{4.2}$$

The sum of the square of these residual values was obtained for each C value, until a minimum

$R_1(n\delta\lambda)$ value was obtained, resulting in a new estimate of the background spectrum $B_1(n\delta\lambda)$:

$$B_1(n\delta\lambda) = X_0(n\delta\lambda) - R_1(n\delta\lambda) \tag{4.3}$$

$$B_1(n\delta\lambda) = [C_1 \times B_0(n\delta\lambda)] + P_1(n\delta\lambda) \tag{4.4}$$

A new spectrum $X_1(n\delta\lambda)$ is now defined that is made up of the values of both $X_0(n\delta\lambda)$ and $B_1(n\delta\lambda)$, where for each individual spectral component, the minimum value is taken from the two corresponding components of $X_0(n\delta\lambda)$ and $B_1(n\delta\lambda)$:

$$X_1(n\delta\lambda) = \min[X_0(n\delta\lambda), B_1(n\delta\lambda)] \tag{4.5}$$

This entire process is then repeated by replacing $B_1(n\delta\lambda)$ with $B_i(n\delta\lambda)$ until the optimal fit of $B_i(n\delta\lambda)$ to $X_i(n\delta\lambda)$ is found such that the peaks of the original cell spectrum, $X_0(n\delta\lambda)$, all lie above the modelled background $B_i(n\delta\lambda)$:

$$R_i(n\delta\lambda) = [X_{i-1}(n\delta\lambda) - (B_0(n\delta\lambda) \times C_i)] - P_i(n\delta\lambda) \tag{4.6}$$

$$B_i(n\delta\lambda) = [C_i \times B_0(n\delta\lambda)] - P_i(n\delta\lambda) \tag{4.7}$$

$$X_i(n\delta\lambda) = \min[X_{i-1}(n\delta\lambda), B_i(n\delta\lambda)] \tag{4.8}$$

The Raman peaks can then be isolated from the original cell spectrum by:

$$\text{Raman peaks} = X_0(n\delta\lambda) - X_{final}(n\delta\lambda) \tag{4.9}$$

Figure 4.1 demonstrates the application of this algorithm to one of the spectra recorded in this thesis, whereby different values of C are applied to $B_0(n\delta\lambda)$, resulting in a new $B_i(n\delta\lambda)$ for each proceeding iteration, in order to obtain a residual spectrum containing only Raman peaks without any background contributions.

The background subtraction algorithm described here is based on the method developed by Beier et al., which can be consulted for further information. Some notable changes to this algorithm have been made; firstly, a smoothing filter has been applied to the initially recorded background signal to reduce noise; secondly for the estimate of C, a uniformly sampled range of C is searched in each iteration as opposed to the use of an empirically chosen factor and the $fminsearch$ function in Matlab; and finally, the background is modelled on a combination of C times the recorded background, $B_0(n\delta\lambda)$, instead of C times the recorded cell data, $X_0(n\delta\lambda)$ as suggested by Beier and colleagues. (77) It should be noted that the algorithm applied in Chapter 8 is based on the exact method proposed by Beier et al., not on our modified version.
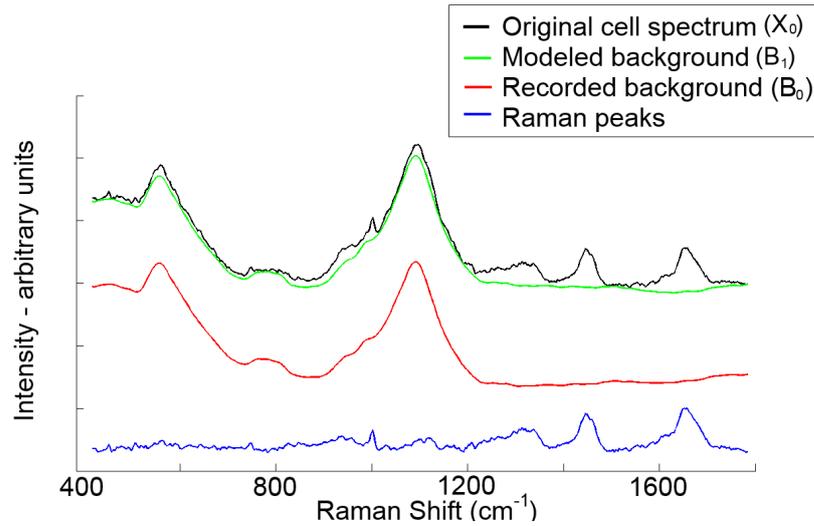
Figure 4.1: *An example of the modified polynomial plus glass signal algorithm being applied for background subtraction - whereby the recorded background signal is combined with a fifth order polynomial until a value of C is found such that the modelled background fits directly under the Raman peaks of the original cell spectrum.*

4.4.3 Extended multiplicative signal correction

An extended multiplicative signal correction (EMSC) algorithm can also be applied to Raman spectra to remove known background contaminants along with the slowly varying baseline signal. (79) This algorithm computes an optimum baseline made up of an N order polynomial and a weighted contaminant signal that is recorded at the beginning of each experiment (e.g. signal from a glass substrate), as described in further detail in Chapter 8. The EMSC algorithm applies a least squares fit to (i) a reference Raman spectrum, (ii) the contaminant signal, and (iii) an N order polynomial. The weight of (i) and (ii), as well as the coefficients of the polynomial are returned by the EMSC algorithm. The reference spectrum provides a basis for all other spectra to be fitted, and for the purposes of this thesis, it is represented by an epithelial cell type recorded on a CaF_2 substrate. Following an OLS determination of the "best fit" for a given Raman spectrum, components (ii) and (iii) are subtracted from the raw spectrum. The value of N is dataset dependent, with higher order polynomials required for accurate modeling of the baseline signal across some datasets. It has been shown elsewhere that the use of high values of N (up to $N = 7$) does not result in over-fitting with EMSC (80), although for many of the cases presented here, only a straight line ($N = 1$) is required (see Chapter 8).

EMSC can also be applied for the removal of Mie scattering artifacts from FTIR datasets, as applied in Chapter 10. This algorithm is based on a reference spectrum either generated from Matrigel or from the mean spectrum for a given dataset. Further information on the application of EMSC for resonant Mie scatter correction is available elsewhere. (81; 82; 83)

Based on the same notation as Section 4.4.2, EMSC can be defined as follows: a raw spectrum, $X_0(n\delta\lambda)$, which can be described as a linear superposition of the Raman spectrum of

interest, $R(n\delta\lambda)$, the baseline signal, $P(n\delta\lambda)$, and the contaminant signal, $B(n\delta\lambda)$:

$$X_0(n\delta\lambda) = R(n\delta\lambda) + B(n\delta\lambda) + P(n\delta\lambda) \tag{4.10}$$

The goal is to estimate the values of $B(n\delta\lambda)$ and $P(n\delta\lambda)$ such that they may be subtracted from the recorded spectrum. Although noise will always be present in the raw spectrum (84), it is assumed that the SNR is sufficiently high such that the noise signal may be ignored.

A reference spectrum, $r(n\delta\lambda)$, is first obtained such that it may be assumed that $R(n\delta\lambda)$ can be approximated by the product of this reference spectrum and a certain weight:

$$R(n\delta\lambda) \approx c_r \times r(n\delta\lambda) \tag{4.11}$$

where c_r is a scalar for a given spectrum.

Similarly, by recording a spectrum directly from a pure contaminant (e.g. a glass slide), $b(n\delta\lambda)$, it is possible to represent the spectral contribution of glass in the recorded cell spectrum, $B(n\delta\lambda)$, as the product of the pure glass spectrum and a certain weight:

$$B(n\delta\lambda) = c_b \times b(n\delta\lambda) \tag{4.12}$$

It should be noted that both c_r and c_b are scalar values that are unique to each cell spectrum, and are dependent on experimental parameters such as the Raman acquisition time.

The slowly varying baseline $P(n\delta\lambda)$ can be represented using an appropriate N order polynomial:

$$P_N(n\delta\lambda) = c_0 + c_1(n\delta\lambda) + c_2(n\delta\lambda)^2 + ... + c_N(n\delta\lambda)^N \tag{4.13}$$

where N is the order of the polynomial, and c_m for $m = 0 \rightarrow N$ represents the various coefficients in the polynomial. (76)

The raw spectrum, $X_0(n\delta\lambda)$, the reference spectrum, $r(n\delta\lambda)$, the contaminant spectrum, $b(n\delta\lambda)$, and the order of the polynomial, N, are all input to the EMSC algorithm, which returns estimates for c_r, c_b, and c_m for $m = 0 \rightarrow N$. These estimates are based on an optimal fit of the various vectors in Equation 4.14 in an OLS sense. (80; 79)

$$X_0(n\delta\lambda) \approx [c_r \times r(n\delta\lambda)] + [c_b \times b(n\delta\lambda)] + \left[\sum_{m=0}^{N} c_m n\delta\lambda^m \right] \tag{4.14}$$

The background corrected cell spectrum, $X_{final}(n\delta\lambda)$, is given by:

$$X_{final} = \frac{X_0(n\delta\lambda) - [c_b \times b(n\delta\lambda)] - \left[\sum_{m=0}^{N} c_m n\delta\lambda^m \right]}{c_r} \tag{4.15}$$

For consistency, the same notation has been used throughout Sections 4.4.2 and 4.4.3, however,

if the reader prefers vector notation to describe this algorithm, this is available in Ref. (85).

4.4.4 Other background correction methods

Various other background algorithms and techniques exist that are not applied within this thesis. The modified polynomial method can sometimes lead to unstable behavior at the end points of the baseline of a spectrum, which led to the development of the "rubberband" method. This method fixes the endpoints of the dataset in order to avoid any such alterations from occurring. (74)

The presence of contaminant signals is highly undesirable in Raman spectra. These signals can arise from unwanted chemical artifacts or from the sample substrate (see Chapter 7). These signals can be removed using the methods highlighted in Sections 4.4.2 and 4.4.3, or using an independent component analysis (ICA) approach. Tfalyi et al. applied ICA along with non-negative least squares to remove residual wax contributions from Raman spectra of biological samples that had previously been stored in embedded paraffin. (86) A similar method was applied for the removal of pharmaceutical drug components (87), and the signal from polystyrene nanoparticles (88), present in Raman cell spectra.

Alternatively, it is possible to record Raman spectra using different Raman system designs in order to reduce the background signals. Such systems include the ultrafast optical Kerr effect (89; 90), or the use of a modulated laser source (9; 91). However, both of these methods are expensive, and experimentally complicated.

4.5 Normalisation

Normalisation in the intensity axis is performed to adjust peak intensity values from each spectrum in order to provide a common scale for comparing Raman peaks across a range of spectra. Normalisation is achieved by dividing each variable in a spectrum by some constant. There are a few normalisation methods that can be utilised for Raman spectra; peak normalisation and vector normalisation are most commonly applied, but area normalisation or min-max normalisation could also be applied. For peak normalisation, the constant is measured as the height difference between the baseline and the maximum point of a chosen peak; vector normalisation obtains the constant value by calculating the sum of the intensity values for each variable in the spectrum, and finding the square root of this value; for area normalisation, the constant corresponds to the sum of the intensity values for each variable in the spectrum; and min-max normalisation involves setting the maximum intensity value to 1, and the lowest intensity value to 0. (92; 93)

4.6 Multivariate statistical algorithms

In the last ten years, there has been growing interest into the diagnostic potential of Raman spectroscopy for classifying healthy and cancerous tissues and cells, as well as identifying the grade (or stage) of cancer present. This approach generally involves the training of a statistical algorithm based on known pathological samples, and as such, some degree of human error is introduced into the method from the beginning. Nonetheless, Raman based diagnostics have shown increasingly impressive results in recent years. This section provides information on the different multivariate techniques that have been used to develop diagnostic algorithms for classification of data obtained from Raman spectroscopy. Techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA) are most commonly employed for classification of biological samples, but other techniques have also been demonstrated.

4.6.1 Principal component analysis

PCA is an unsupervised statistical tool used to reduce the number of variables within a data set. To do this, PCA transforms the spectral data into a set of variables called principal components (PC), whereby all PCs are orthogonal to each other and they are generated in such a way as to represent as much variance within the dataset as possible. This is achieved by finding the direction through the data set that explains the maximum variance, which is known as the first PC coefficient, and is equivalent to finding the line of best fit through plotted data points. (94) Additional PC's must be orthogonal to the previous ones, whilst describing the maximum variance remaining in the data, as shown in Figure 4.2.
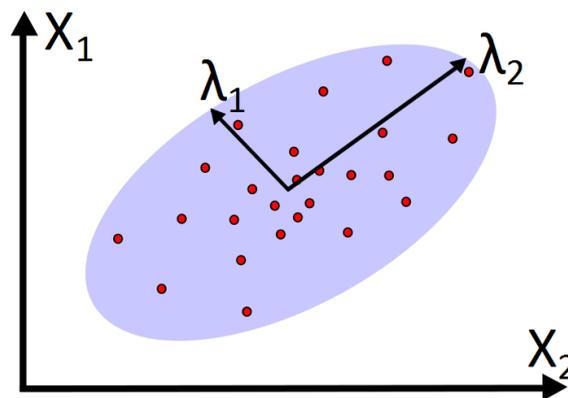
Figure 4.2: *A representation of PCA, whereby component axes are used to maximise the variance within a dataset*

Computationally, PCs are found by determining the covariance matrix of the data set, and calculating the eigenvectors and eigenvalues of this covariance matrix. The eigenvector, or "loading", with the largest eigenvalue represents the first PC, and the eigenvector with the second largest eigenvalue is the second PC, and so on. When the values of each PC are calculated,

it is then possible to use basic geometry to express each sample data set in terms of linear summations multiplied by a scalar value, known as a "score". (95) With this technique it is possible to significantly reduce the number of variables in a dataset whilst still maintaining almost all of the spectral information; for example, in an average Raman spectral measurement, data is collected between 400–1800 cm^{-1}, with measurements taken approximately every 3 cm^{-1}, thus resulting in $\sim$500 variables in the original data set, which can be reduced to $\sim$20 using PCA. Further analysis can then be applied to these PCs to organise them into groups, or clusters, representing different pathologies; to achieve this, techniques such as LDA are often utilised. The main advantage of this tool is that it provides a simpler representation of the data and allows for faster classification algorithms to be designed.

4.6.2 Linear discriminant analysis

LDA, also known as Fisher's discriminant analysis, is a supervised multivariate technique used to optimise class separability by finding the direction that provides the best separation for two or more groups of data. LDA is often applied to PC scores to further reduce the dimensionality of the data set. Similar to PCA, this is achieved by finding a linear combination of vectors that maximise the variance of the dataset, but with the addition of finding the component vectors that maximise the separation between multiple classes, as shown in Figure 4.3. (64) By maximising the variance in this manner, LDA is able to provide the optimum separation for each group, thus improving classification results.
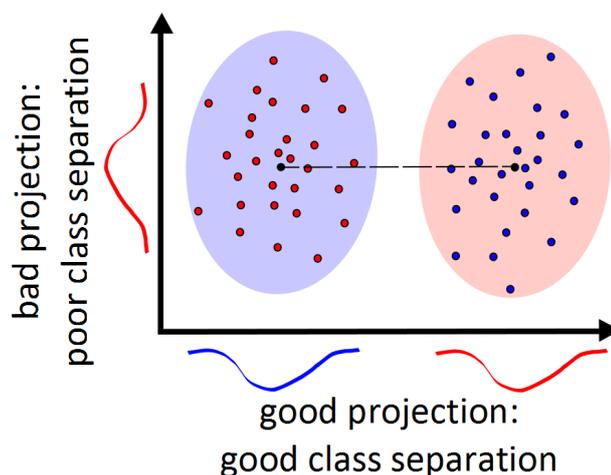
Figure 4.3: *Representation of LDA, which maximises the component axes to provide class-separation*

So, in short, PCA is an "unsupervised" algorithm, which means it "ignores" class labels, and its goal is to find the directions (PCs) that maximise the variance in a dataset. In contrast, LDA is "supervised" and computes the directions (LDs) that represent the vectors that maximise the separation between multiple classes. In general, dimensionality reduction does not only help with reducing computational costs for a given classification task, but it can also be helpful to

avoid overfitting by minimising the error in parameter estimation.

4.6.3   Cross validation

Cross validation is often used to estimate how accurately a diagnostic classification model will perform. This is achieved by assessing the results of the statistical algorithm when applied to a validation set of data. The most common method used for assessing Raman spectra models is leave-one-out cross validation. Leave-one-out is based on using a single spectrum as the validation set, and the remaining spectra are used as the training set for the algorithm. This is repeated to test each spectrum in the dataset iteratively, and can be used to determine how accurate the model is at predicting the pathological status of the sample. Alternative variations include k-fold cross validation, blind testing, or double blind testing. K-fold cross validation involves partitioning the datasets, such that k spectra are used for validation with the remaining spectra used for training, blind and double blind testing are based on concealing pathological information from the data in order to remove possible observer bias.

| | Classified as: | |
	Cancerous	Healthy
Cancerous samples	TP	FN
Healthy samples	FP	TN

Table 4.1: *Basic demonstration how samples may be classified following PC-LDA with a cross validation approach. TP, true positive; FN, false negative; FP, false positive; TN, true negative.*

Diagnostic classification results are often presented in terms of sensitivity and specificity, which provides a good representation of the performance of the algorithm. Based on the values shown in Table 4.1 for a basic classifier, the sensitivity and specificity values can be calculated as follows:

- Sensitivity:
  Refers to the proportion of things that are being looking for that are found, basically how good the model is at getting things right:

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (4.16)$$

  whereby TP is the true positive, and FN is false negative.

- Specificity:
  The proportion of things that are not being looked for that are not found, basically how good the model is at making sure the wrong things aren't found:

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \qquad (4.17)$$

whereby TN is the true negative, and FP is false positive.

4.7 Summary

Raman micro-spectroscopy is a complex modality that can be applied for disease classification. The physics of Raman spectroscopy, and the design of Raman micro-spectrometers, have been discussed in Chapters 2 and 3, respectively. However, due to the presence of noise, unwanted background signals, as well as calibration errors, it is necessary to apply a range of numerical methods to produce Raman spectra that are reliable and reproducible. This chapter presents various pre-processing tools that are available for calibration, noise reduction, and normalisation. Additionally, baseline correction methods and algorithms for the removal of unwanted contamination signals are discussed in Section 4.4. These techniques are essential to produce spectra that can be used for the accurate diagnosis of disease pathologies.

Following pre-processing, Raman data collected from different biological samples can be classified using multivariate statistical algorithms. The most common algorithms applied for classification of Raman spectra involve PCA and LDA, as discussed in Section 4.6, with sensitivity and specificity results often calculated based on a leave-one-out cross validation approach. Furthermore, in order to produce reliable classification algorithms, it is important to record Raman spectra from biological samples that have been prepared using suitable sample preparation methods, as discussed in further detail in Chapter 5.

Chapter 5

The biology of cells, cancer, and exosomes

5.1 Introduction

In order to appreciate the application of Raman micro-spectroscopy to biological specimen for disease diagnostics, it is essential to understand the fundamental biology involved in cells and tissues, the origin and progression of cancer, and the role of exosomes within the body. Section 5.2 provides an overview of cell biology, Section 5.3 discusses how cancer occurs, and Section 5.4 investigates the origin of exosomes and their potential role in disease progression. The methods involved in the preparation of cytology samples during the course of this thesis are discussed in Section 5.5. The main focus of this chapter is on the biology of the urinary bladder, as the development of Raman based urine cytology protocols are the primary focus of this thesis. This chapter is followed by a more detailed discussion of bladder cancer, and recent work in the area of bladder cancer diagnostics using Raman spectroscopy.

5.2 Cell biology

Cells are single units that work together to carry out all of the necessary functions of an organism. (96) Whilst cells all have similar components, individual subgroups exist that can do specialised jobs. There are two main types of cells: prokaryotic (bacteria) and eukaryotic (animals and plants); however, this thesis is only focused on eukaryotic cells. Cells share the following common organelles: nucleus, nucleolus, endoplasmic reticulum (smooth and rough), golgi apparatus, lysosomes, peroxisomes, and mitochondria, which are shown in Figure 5.1.

Cell nuclei vary in size depending on the cell type, but typically have a diameter of $\sim 5 \ \mu$m. The nucleus is the 'control centre' of a cell, where genetic information is stored in the form of chromatin. The nucleus is surrounded by a nuclear membrane consisting of a phospholipid bilayer (similar to that found in the cell's plasma membrane), with the outer membrane attached to the rough endoplasmic reticulum, and the inner membrane carrying unique proteins that are specific to the nucleus. Nuclear pores can be found along the nuclear membrane to allow for
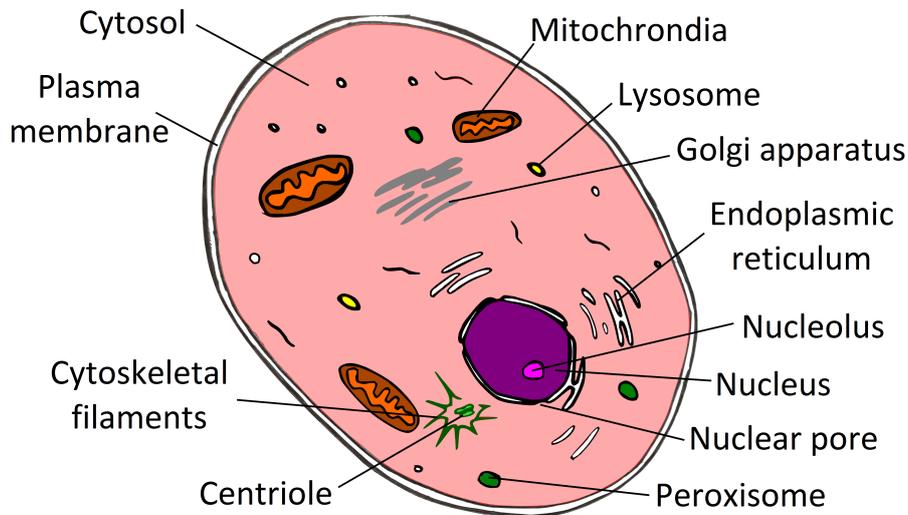
Figure 5.1: *Schematic of a typical eukaryotic cell*

the transport of RNA from the nucleus into the cytoplasm for protein synthesis, and for the transport of specific proteins into the nucleus. The nucleolus is found within the nucleus, and is the location of DNA, and many types of RNA and proteins. Further information on the roles of other cellular organelle can be found elsewhere. (97)

Four levels of organisation can be found within vertebrate organisms: (i) cells, (ii) tissues, (iii) organs, and (iv) organ systems. Groups of cells which have a similar structure and function form tissues (e.g. epithelial tissue); organs are composed of tissues that form a structural and functional unit (e.g. bladder); and an organ system consists of a group of organs that work together to perform the major activities of a body (e.g. urinary system).

5.2.1 Epithelial tissue

Epithelial tissue, otherwise known as an epithelium, is a continuous layer of tightly packed cells that covers the surface of the body and inner cavities. External epithelial surfaces protect the body from invasion, injury, and dehydration. The internal surface mainly offers protection, with some cell types secreting mucus (e.g. digestive tract), removing impurities (e.g. respiratory tract), or absorbing molecules (e.g. kidney tubules). (98)

There are three types of epithelial tissue: (i) squamous: composed of flattened cells, and provides protection, diffusion, and filtration; these cells are mainly found in the lining of the lungs and blood vessels; (ii) cuboidal: these are cube shaped cells which provide protection, secretion, and adsorption, and are found in the lining of kidney tubules; and (iii) columnar: composed of rectangular or column-shaped cells, which provide protection, secretion, and adsorption, and are mainly found along the digestive tract. Additionally, the epithelium can be

simple (i.e. single layer of cells), stratified (i.e. layers of cells on top of each other), or pseudos-tratified (i.e. appears to be layered, but all cells touch the base line). (98) Figure 5.2 illustrates the differences between each type of epithelial tissue.

The urinary bladder is lined with transitional epithelium, otherwise known as the urothe-lium or bladder mucosa, which consists of multiple layers of epithelial cells that can contract and expand to accommodate large volumes of urine. The urothelium provides protection to surrounding tissues from acidic or alkaline urine.

5.3 Cancer biology

Cancer is defined as a disease that involves a group of cells growing uncontrollably by disregard-ing the normal rules of cell division (i.e. mitosis). (99) Cells are constantly subject to signals that dictate whether the cell should divide, differentiate, or die. Cancer cells do not respond cor-rectly to these signals, resulting in uncontrolled growth and proliferation, which can eventually become fatal. Cancer is a disease that involves changes or mutations in the cell genome, and these changes disrupt the delicate cellular division process, resulting in cells that keep dividing and form cancerous tumours. Malignant tumours must not be confused with benign tumours, which are slow growing and remain confined to their original location, and typically are not life-threatening. (100)

Approximately one in three people will suffer from cancer during their lifetime, with inci-dence rates varying dramatically across countries and populations. (100) Cancer is a multi-gene, multi-step disease originating from a single abnormal cell with a DNA mutation. Subsequent rounds of mutations and expansion of these cells results in the formation of a tumour. (101) Death as a result of cancer is due to the invasion, erosion, and spreading of tumours into sur-rounding tissues (i.e. metastasis). Many factors are involved in the initiation and progression of cancer, including the environment (tobacco, chemicals, radiation) and factors within the cell it-self (inherited mutations, hormones, immunity). It is important to note that it takes a significant amount of time for these DNA mutations to accumulate and result in a detectable cancer. (102)

There are four distinct types of cancer: (i) carcinoma: this is a malignancy found within epithelial cells and tissues; (ii) sarcoma: this is a solid tumour found within connective tissues; (iii) leukemia: a cancer that arises from blood-forming cells; and (iv) lymphoma: a cancer formed within immune-system cells. Further subgroup classifications are also made to define the tissue and cell types involved in the cancer. (102)

Although cancerous tumour cells share common characteristics, every type of cancer has its own particular sequence of mutations and uncontrolled growth. These common character-istics include a lack of differentiation, abnormal enlarged nuclei with an abnormal number of chromosomes, and the ability to induce angiogenesis and metastasis. Additionally, cancer cells can form tumours, this is particularly evident when growing cells in a petri dish; normal cells grow to cover the bottom of the dish in a single layer, and require a growth factor to induce
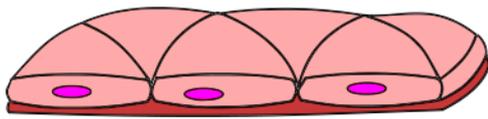
(a) Simple squamous epithelial tissue
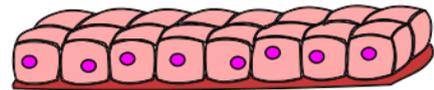
(b) Simple cuboidal epithelial tissue
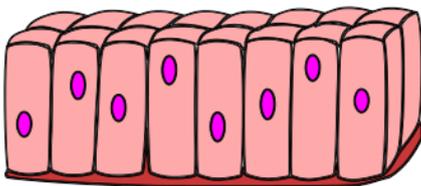
(c) Simple columnar epithelial tissue
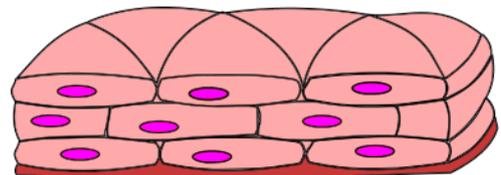
(d) Stratified squamous epithelial tissue
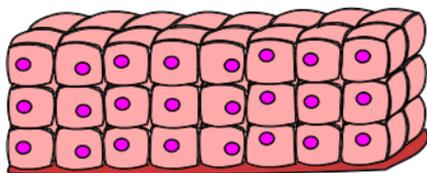
(e) Stratified cuboidal epithelial tissue
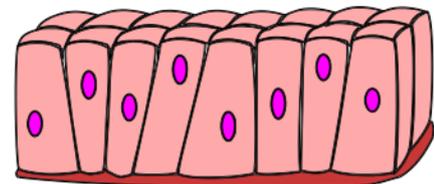
(f) Pseudostratified columnar epithelial tissue
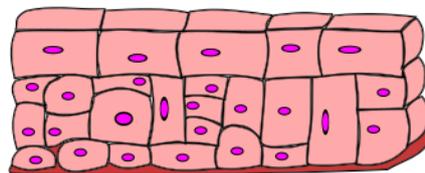
(g) Transitional epithelial tissue

Figure 5.2: *A comparison of each of the different epithelial tissue subgroups.*

Normal cells:
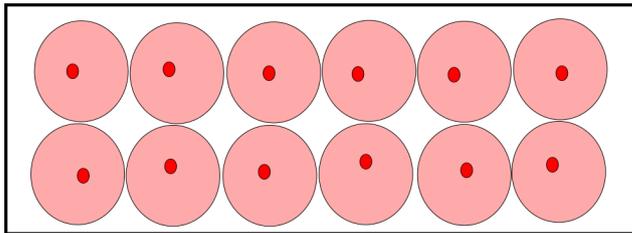
- controlled growth
- contact inhibition
- one organised layer
- differentiated cells

Cancer cells:
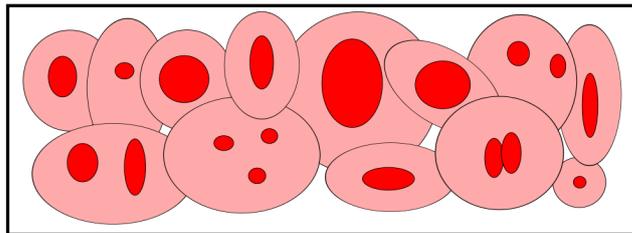
- uncontrolled growth
- no contact inhibition
- disorganised, multilayered
- non-differentiated cells
- abnormal nuclei

Figure 5.3: *Typical differences found between normal and cancerous cells.*

division, whereas cancer cells continue growing on top of each other within the dish. Figure 5.3 illustrates the typical differences seen between normal and cancerous cells, with large, variable nuclei, small cytoplasm to nuclei ratios, variations in cell size and shape, and a disorganised arrangement of cells being the most obvious differences. (100)

5.3.1 Cancer diagnostics

A myriad of diagnostic tests are available for the detection and evaluation of human cancers, ranging from whole body PET-CT (positron emission tomography-computed tomography) scans, invasive endoscopic procedures, cytological smears, to blood tumour marker tests. Many non-invasive and highly accurate diagnostic modalities have been identified, however despite these significant research efforts, many organ-specific cancers remain without an accurate diagnostic tool. One cancer in particular is urothelial carcinoma (i.e. bladder cancer), which does not carry any specific symptoms, and can only be accurately identified with an invasive cystoscopic procedure. (103) New optical techniques are currently being developed for cancer diagnostics, such as optical coherence tomography (OCT) (104), diffuse reflectance spectroscopy (105), and spatial light interference microscopy (SLIM). (106) This thesis is interested in the application of vibrational spectroscopic techniques for highly accurate, non-invasive cancer diagnostics.

5.3.2 Hypoxic cancer tumours

Hypoxia is a key feature of solid tumours and occurs when the tumour has outgrown its vasculature supply. (107) Tumour hypoxia is associated with several aspects of aggressive tumour behaviour including increased invasiveness and proliferation, the formation of metastasis and

poor patient survival rate. Importantly, hypoxic tumours are known to be highly resistant to radiation and chemotherapy treatments. (108; 109) It has been shown by Watson et al. (110) that hypoxia can cause cells to become senescent (i.e. they are still viable, but now cease to divide), which is problematic for treatment since many chemotherapies target actively dividing cells, and therefore senescent cells may remain unaffected when chemotherapeutically challenged. Paradoxically, chemotherapeutics themselves can also cause senescence, termed therapeutic induced senescence (TIS). (111) Tumour hypoxia results in a poor patient survival rate, so knowing the cause and extent of hypoxia is invaluable for treatment planning.

5.4 Exosome biology

Exosomes are microvesicles (40–100 nm in diameter) of endocytic origin released into the extracellular environment by most cell types and play a significant role in cell signalling and communication by virtue of their capacity to traffic miRNAs, mRNAs, and proteins to recipient target cells. (112) Exosomal content originates in the cytoplasm of the parent cell, as illustrated in Figure 5.4, whereby compartments are formed within cells known as multivesicular endosomes (MVE). These compartments contain membrane-bound proteins, cytoplasmic proteins, mRNAs, and miRNAs, and are released when the MVE fuses with the cell's membrane. (113; 114)
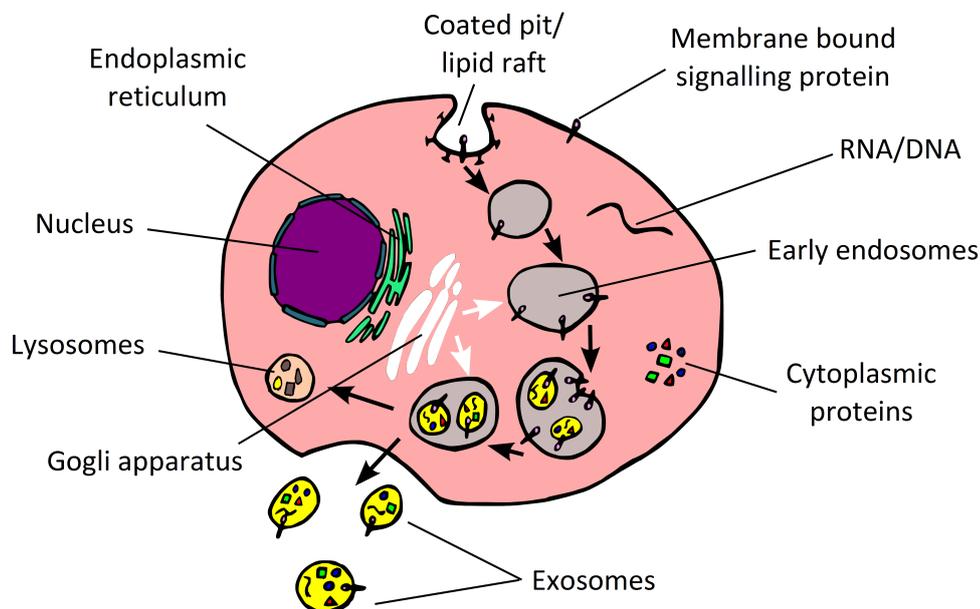
Figure 5.4: *Illustration of the process of exosome formation, whereby compartments are formed within the cell containing proteins and RNA, which are then released from the cell when fused with the cell's membrane.*

Exosomes can be found in biological fluids, and it is believed they can play an important role

in the pathogenesis of disease, which will be discussed further in Chapter 12. Recent evidence has shown the role of tumour derived exosomes in cancer progression, whereby exosomes have the ability to suppress the immune system and thus hinder the development of anti-tumour responses. (114; 115) On the other hand, it has also been shown that exosomes can act as an acellular source of antigenic determinants that could be exploited in the design of cancer vaccines. (116; 117) While all cells have the ability to produce exosomes, either constitutively or upon activation/stress, it has been reported that the levels of exosomes in patients with cancer are significantly higher than those in normal plasma. (118) It has also been shown that cancer cells growing in hypoxia produce more exosomes in comparison to cancer cells growing in normal oxygen conditions. (119) Exosomes have received increasing attention in recent years because of their potential use as non-invasive biomarkers for the detection of diseases such as cancer.

5.4.1 Exosome isolation

Exosomes can be found in biological fluids such as blood, urine, and cell culture medium, however, due to the complex nature of biological fluids, it is extremely difficult to isolate pure samples of exosomes. The most common procedure for the isolation of exosomes from fluid involves a series of centrifugations, which removes cells and large debris, followed by a high speed ultracentrifugation step to pellet the exosomes. However, this does not isolate exosomes from other microvesicles or large proteins, so further testing is often needed to identify the presence of exosomes. Instruments such as the NanoSight (Malvern Instruments, UK) can be used to visualise and measure the size of exosomes, as well as counting the number of exosomes present in the solution. When analysed with a transmission electron microscope (TEM), exosomes have a characteristic 'saucer-like' morphology, as shown in Figure 5.5. (120) The molecular composition of exosomes can be determined using techniques such as western blotting, fluorescence-activated cell sorting (FACS), or Raman spectroscopy. (120; 121; 122)

5.5 Cytology sample preparation methods

There are three main considerations involved with the preparation of cytology samples for Raman micro-spectroscopy, (i) the method of depositing cells onto the microscope slide, (ii) cell fixation, and (iii) the potential re-use of expensive spectroscopic substrates, which are discussed in Sections 5.5.1 – 5.5.3, respectively.

5.5.1 Cell deposition

Clinical cytology samples are often prepared within the cytopathology laboratory using liquid based processing techniques, such as SurePath (Becton Dickson and Company, USA) or Thin-
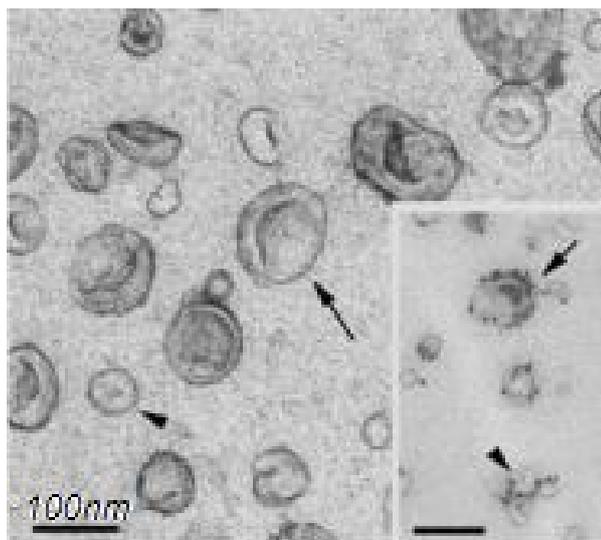
Figure 5.5: *TEM image of exosomes, showing their characteristic 'saucer-like' morphology. Image obtained by our collaborator Luke Gubbins at the Conway Institute, University College Dublin, Ireland.*

Prep (Hologic, USA). These systems are designed to prepare uniform monolayer cells onto glass slides with minimal cell debris or blood residue in the background. SurePath is a density gradient based cell enrichment process that fixes cells with an ethanol based solution, whereas ThinPrep is a filter based cell concentration technique that uses the methanol based solutions PreservCyt and CytoLyt. (123; 124; 125) The ThinPrep method is applied in the preparation of many of the cell samples throughout this thesis.

Figure 5.6 shows the ThinPrep 2000 (T2), which can be used to prepare both gynaecology and non-gynaecology cytology samples. As this thesis is focused only on non-gynaecology samples, the following description applies to the preparation of these samples only. For urine cytology, a vial of urine is collected from the patient, and mixed in a 2:1 ratio of urine to PreservCyt, this prevents the cells from degrading within the urine solution. The cell pellet can be obtained at a later time following centrifugation. The cell pellet should be exposed to CytoLyt, a red blood cell lysing agent, to remove any residual blood present within the sample. A few drops of the cell pellet is then suspended in a ThinPrep vial of PreservCyt, such that a slightly cloudy solution is formed. The cells remain in the PreservCyt solution for at least 15 min to allow for adequate fixation. The vial is placed into the T2, as shown in Figure 5.6, along with a ThinPrep glass slide, and a filter tube.

In order to transfer the cells from the PreservCyt vial onto the glass slide, the filter tube is inserted into the vial, where it rotates, creating currents within the fluid, allowing for the separation of debris and mucus, without damaging cell integrity, as illustrated in Figure 5.7(a). A gentle vacuum is then applied within the filter tube, allowing for the collection of cells onto the exterior surface of the filter membrane, as shown in Figure 5.7(b), whilst allowing residual PreservCyt solution to flow up through the filter tube, where it is passed along to an exterior
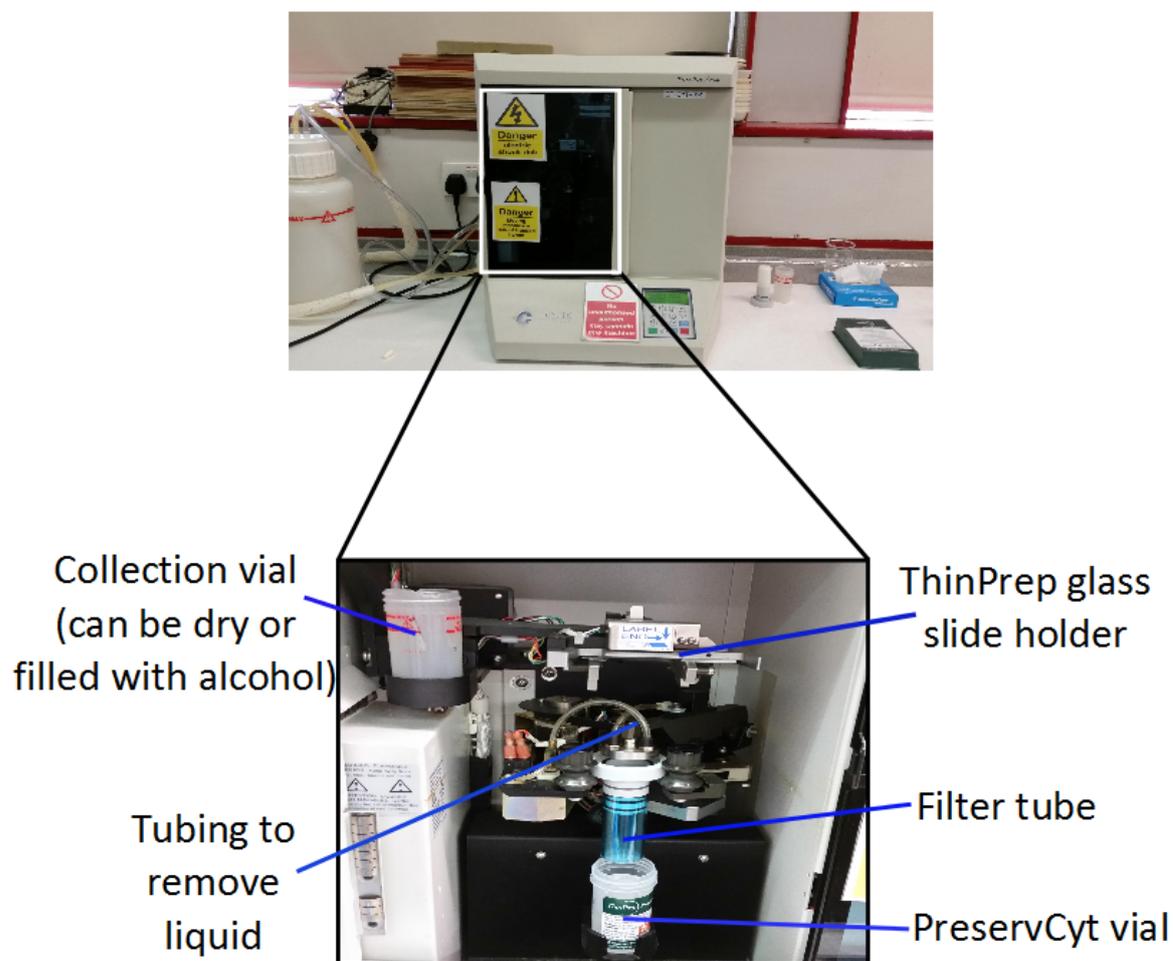
Figure 5.6: *Image of the ThinPrep 2000 (T2) system, with the key internal components identified, which is located in the School of Biological Sciences, Dublin Institute of Technology, Ireland.*

waste collection vial. The rate of flow through the filter tube is controlled by the T2 software, and varies across cytology samples. When an adequate number of cells have collected onto the filter membrane, the filter tube is removed from the PreservCyt vial, and rotated 180°, such that it can come into contact with the microscope slide, as shown in Figure 5.7(c). Natural attractive forces, and slight positive air pressure, causes the cells to detach from the filter membrane and adhere to the glass microscope slide. This method of depositing cells onto a slide allows for an even distribution of cells within a specific region of the slide.

Alternative methods for cell deposition include (i) a basic drop-dry method, where the cell pellet is simply dropped onto a microscope slide and left to dry at room temperature; (ii) cytospin, whereby the cell suspensions are spun onto a microscope slide via centrifugation forces (126); or (iii) growing of cells directly onto microscope slides, however this is only possible for sterile, biocompatible substrates. (127)
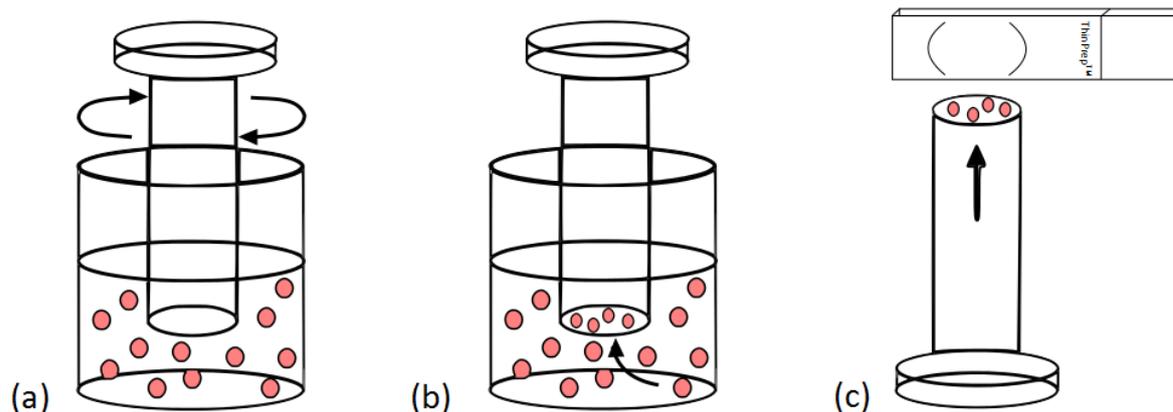
Figure 5.7: *ThinPrep sample preparation process; (a) rotation of the ThinPrep filter to separate debris and mucus; (b) collection of cells on the filter membrane; and (c) transfer of cells from the filter membrane onto the ThinPrep microscope slide.*

5.5.2 Fixation methods

It is important to fix, or preserve, cytology samples in order to prevent morphological or spectroscopic changes occurring during Raman spectral acquisition. Additionally, it is essential to fix samples that have been obtained from the clinic, where the presence of infectious pathogens is unknown. With the ThinPrep method, as discussed in Section 5.5.1, the cells are fixed with a methanol solution prior to deposition onto the slide. Alternative fixation methods include (i) dry fixation, where the cells are allowed to dry onto the slide; or (ii) formalin fixation, whereby the cells are placed into a solution of 10% buffered formaldehyde for approximately 10 min to allow for cross-linking of the collagen amine groups. (128)

After fixation, it is often necessary to wash the slide to remove any residual fixation solution, although it should be noted that washing is not necessary following ThinPrep sample preparation due to its filter-based properties. Washing of cells should be performed with sterile, filtered PBS (phosphate buffered saline) or deionized (DI) water. PBS is the preferred option since DI water can cause cell lysis, however, the salt crystals within PBS can result in additional contamination peaks within the cell spectra, therefore it is recommended that the PBS is filtered prior to any washing steps.

5.5.3 Cleaning of sample substrates

Spectroscopic substrates such as CaF_2, quartz, or MgF_2 are often very expensive, and therefore, it is necessary to have a protocol in place to allow for adequate cleaning and re-use of such slides. Within this thesis, this is achieved by (i) autoclaving, whereby pressurised steam is applied to the substrates to sterilise them, (ii) gently cleaning with 2-Propanol (isopropanol) to remove any residual dirt, (iii) dipping the substrates into acetone, followed by DI water, and (iv) placing the substrates into a bath of ethanol for 30 min. Step (iv) is repeated three times,

using fresh ethanol each time. The substrates should be quickly dried, and stored in a sterile, sealed container until needed.

5.6 Summary

The cellular and subcellular components involved in cancer progression are discussed within this chapter. Information on cell biology is given in Section 5.2, along with further information on how cells form tissues and organs, as well as identifying the different subtypes of epithelial tissues that exist within the human body.

Section 5.3 defines cancer as an uncontrolled growth process that results in cells forming malignant tumours. It has been noted that although cancer cells share many common characteristics, every cancer subtype is the result of a unique sequence of mutations. It is important to be able to identify if a cancerous tumour has outgrown its vasculature supply, resulting in a hypoxic tumour. Hypoxic tumours have poor patient survival rates and require specific treatment plans, which will be discussed further in Chapter 12.

The role of exosomes in cancerous tumours is briefly discussed in Section 5.4. These microvesicles are the topic of many recent research investigations, with increasing evidence showing that exosomes have the ability to communicate with neighbouring cells. Chapter 12 applies Raman micro-spectroscopy to monitor changes in cells that have been exposed to exosomes.

An important consideration when preparing cytology samples for Raman micro-spectroscopy is the choice of cell deposition method. Section 5.5.1 focuses on the liquid based processing technique known as ThinPrep, which is applied for the preparation of many cytology samples throughout this thesis. The ThinPrep method provides monolayer cell distributions onto glass slides, and is routinely applied in the cytopathology laboratory. Fixation agents and the potential re-use of expensive spectroscopic substrates are other important considerations that need to be factored into the design of Raman micro-spectroscopic experiments involving cytology samples.

Chapter 6

Bladder cancer diagnostics using Raman spectroscopy: a literature review

This chapter has been published as part of the following paper:

- L.T. Kerr, K. Domijan, I. Cullen, B.M. Hennelly. *Applications of Raman spectroscopy to the urinary bladder for cancer diagnostics.* Photon. Lasers Med., 3(3), 193-224, 2014.

6.1   Introduction

The main focus of this thesis is on the diagnosis of bladder cancer cells that have been shed into urine using Raman micro-spectroscopy; therefore, this chapter provides a literature review of this topic, and provides an analysis of previous studies that have been carried out in this area. Additional investigations into the presence of prostate cells in urine, and the ability to classify between bladder and prostate cells, are discussed in Chapters 10 and 11. Information on the role of exosomes in cancerous tumours can be found in Chapter 12, and although this chapter focuses on exosomes found in breast cancer tumours, similar exosomes could be isolated and analysed from urine samples instead.

Bladder cancer is the fourth most common cancer in men, with approximately 200 new cases diagnosed in men and women every day in the USA. (129) As with most solid organ malignancies, early diagnosis, detection, and treatment are imperative for good clinical outcomes. Cystoscopy is the cornerstone of bladder diagnostics for real-time visualisation of the bladder mucosa. However, by its nature, it is an uncomfortable invasive procedure, and is not without significant risk or potential complications. Urine cytology often serves as a useful non-invasive diagnostic adjunct in the diagnosis of bladder cancer. Examination of voided urine or bladder washing specimens for exfoliated cancer cells has high sensitivity in high grade tumours, but low sensitivity in low grade tumours. (130) It has been shown that delays in the process of diagnosis leads to unsatisfactory outcomes and adverse effects for the patient involved. (131; 132) Therefore, there is a real clinical need to develop and integrate a non-invasive, sensitive, and

specific technique to assist in bladder cancer diagnosis.

Optical diagnostic tools in medicine are becoming more popular in recent times because they can provide fast, highly accurate results, without damaging the biological samples with toxic chemical markers. Raman spectroscopy, which was introduced in Chapters 1 and 2, is one of the most promising optical diagnostic tools that can be applied to cells from the urinary bladder as it is able to detect changes in the molecular composition of tissue samples. This is not only useful for discriminating between different grades and stages of disease, but may also provide epigenetic clues as to the etiology of the malignancy. This chapter reviews the contributions over the past decade on the application of Raman spectroscopic techniques for improving the accuracy of detecting bladder cancer.

The breakdown of this chapter is as follows; firstly, we build upon the basic discussion of biology that was introduced in the preceding chapter to provide an introduction to bladder pathology and current diagnostic approaches, which are given in Section 6.2. Section 6.3 summarises all previous applications of Raman spectroscopy to diagnose bladder biopsy and tissue samples in vitro and ex vivo. Section 6.4 focuses on the use of Raman fiberoptic probes integrated into cystoscopic procedures for real-time cancer diagnostics in vivo. And Section 6.5 describes the completely non-invasive technique of combining Raman spectroscopy with urine cytology.

6.2 Bladder cancer

6.2.1 Diagnosis of bladder cancer

Bladder cancer is generally diagnosed using cystoscopy, aided by urine cytology. Cystoscopy is an invasive endoscopic procedure of the urinary bladder carried out via the urethra, as shown in Figure 6.1; however it is sometimes indeterminate, and there are variables that can result in false negative results. Bladder mucosa inflammation (cystitis) is commonly encountered in individuals with a urinary tract infection and individuals with indwelling catheters, and carcinoma in situ (CIS), a particularly aggressive form of non-invasive bladder cancer with a high propensity to progress to muscle invasive disease, also appears as as inflamed bladder lining.

Cytology is useful when a high grade cancer or CIS is present, but its use in the diagnosis of low grade cancers is very limited owing to its low level of sensitivity. (133) As a standard procedure, cystoscopy is performed using white light. However, the use of white light can lead to missing lesions that are present but not visible, which is why new technologies are being developed to aid the clinician. Photodynamic diagnosis (fluorescence cystoscopy) is performed using violet light after intravesical (intra-bladder) instillation of 5-aminolevulinic acid (5-ALA) or hexaminolevulinic acid (HAL). It has been confirmed that fluorescence guided biopsy and resection are more sensitive than conventional procedures for the detection of malignant tumours, particularly for CIS of the bladder. (134)
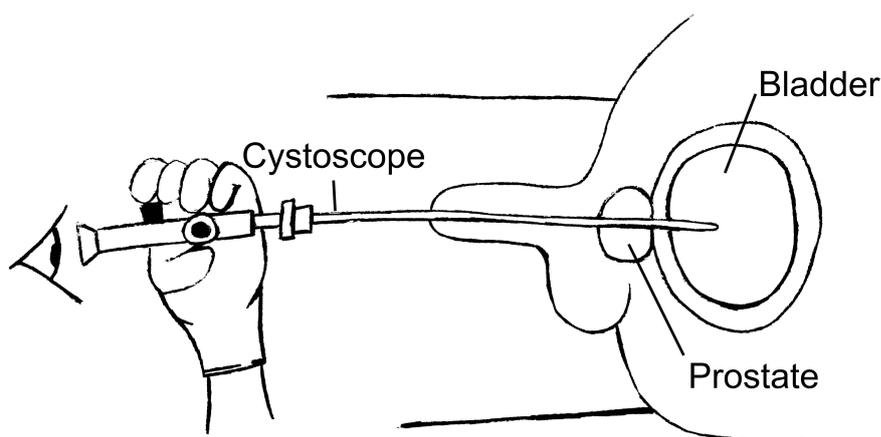
Figure 6.1: *Illustration of a typical cystoscopy procedure, whereby a cystoscope is passed through the urethra into the bladder for the visualisation of bladder disease.*

6.2.2 Classification of bladder carcinoma

Approximately 90% of all bladder cancers are diagnosed as urothelial carcinoma (UC), otherwise known as transitional cell carcinoma (TCC). (135) UC develops in the urothelium of the bladder wall, and is the second most common malignancy of the genitourinary tract. (136) The remaining ∼10% of bladder cancers consist of squamous cell carcinoma, adenocarcinoma, lymphoma, and metastatic tumour cells from different neoplastic primaries. Almost 80% of patients with UC present with non-muscle-invasive, or superficial UC, the remainder being infiltrative, invading the detrousor muscle and potentially spreading to loco-regional lymph nodes and viscera. (137) There are two main diagnostic classification systems for UC based on histological grading and pathological staging. Histological grading represents the degree of differentiation present between the carcinoma cells and normal urothelial cells, which can be used to predict the rate at which the carcinoma cells are multiplying. Pathological staging is used to assess the extent of cancer invasion into the bladder wall in order to determine if the disease has spread.

Histological grading

The histological grade of a UC tumour is determined under a microscope using a specific set of criteria to establish how much the tumour resembles normal (healthy) tissue. Within histological grading, there are two systems in use; the WHO 1973 grading system and the WHO 2004 grading system. Most histopathologists still use the WHO 1973 system, which is based on a three-tiered grading system as summarised in Table 6.1. Grade 1 represents well differentiated cells that are slow growing and look similar to healthy bladder cells, Grade 2 cells are moderately differentiated, and Grade 3 are poorly differentiated, fast growing cells that look very different to healthy cells (138), as illustrated in Figure 6.2. Histological grading is often

criticised given the subjective nature of this test.
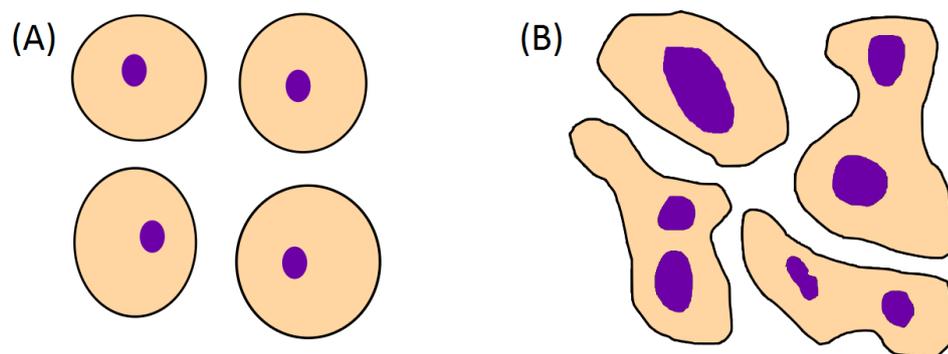
Figure 6.2: *Basic representation of (a) normal (healthy) cells, and (b) poorly differentiated cancer cells.*

Histological grading is very subjective since it is based on the pathologist's perception of cell differentiation. Generally, diagnosis is made by two or more pathologists per sample in order to avoid misclassification; however there is still often inter-observer variation (between pathologists) and intra-observer variation (from one pathologist at different times). (139) Thus, this method of classification is time-consuming with a high degree of variability. The future trend is leaning towards the use of the newer system (WHO 2004), which is based on simply grading cancers as low or high grade, thus reducing inter-observer variation.

Pathological staging

Pathological staging is based on the TNM (tumour, node, metastases) system, whereby the pathologist assesses the location of the tumour (T), if the cancer has spread to nearby lymph nodes (N), and the extent of metastases (M). A number is assigned to each value ranging from 1 to 4, where higher numbers indicate a more aggressive, invasive carcinoma. (103) The tumour (T) value, which is the most important value here, represents the location of the tumour with respect to the inner lining of the bladder wall. Stages Ta, T1, and CIS are non-muscle-invasive UC, with Ta and CIS both confined to the urothelium. Most Ta and T1 tumours are low grade and are less likely to become muscle-invasive. CIS is a flat, high grade, non-invasive UC. Macroscopically, CIS can be missed at cystoscopy or be considered as an inflammatory lesion if not biopsied. A summary of each pathological stage can be seen in Table 6.1, and Figure 6.3 illustrates the extent of the growth of the tumour with respect to the inner bladder wall lining.

6.3 Biopsy and tissue sample diagnostics

In this section, the specific studies relating to the application of Raman spectroscopy to biopsy and tissue samples for the purpose of bladder disease diagnosis are reviewed. This section begins by discussing the sources of tissue and the manner in which they can be stored and
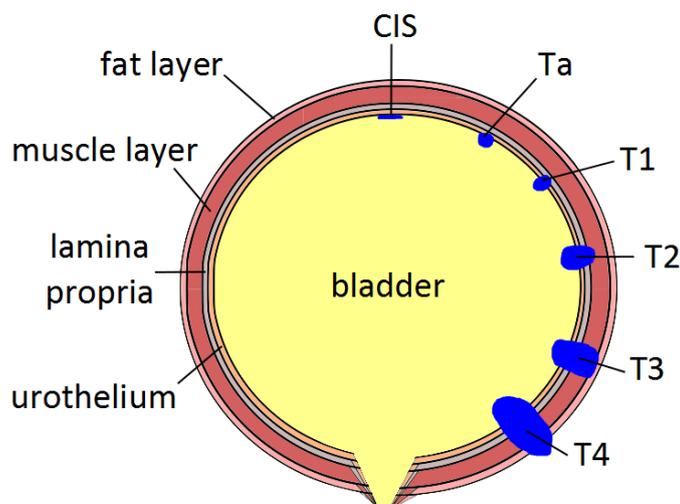
Figure 6.3: *Staging of bladder tumours representing the location of each tumour with respect to the urothelium.*

prepared for Raman spectroscopic analysis. This is followed by a review of each of the major studies in the area, the techniques used, and the resulting sensitivities and specificities that were achieved.

To diagnose bladder tissue using conventional Raman spectroscopy, tissue samples can be removed from tumourous regions during transurethral resection surgical procedures, such TURBT (transurethral resection of bladder tumour) or TURP (transurethral resection of the prostate). Both of these are minimally invasive surgical procedures, carried out via the urethra by passing surgical tools through a cystoscope into the bladder. Large tumours or CIS are more difficult to remove or analyse during TURBT or TURP, so biopsy samples can be taken during cystoscopy procedures instead. A more invasive surgical procedure for the removal of all or part of the bladder is known as a cystectomy, and these tissue samples can also be diagnosed using Raman spectroscopy.

Raman spectroscopy can be effectively applied to both fresh and preserved bladder tissue samples. There are numerous methods of preserving bladder tissues so that they can be stored for long periods of time without largely affecting their natural biochemistry. Tissue samples can be snap-frozen with upon removal from the patient, and then stored at -80°C. The main advantage with freezing is that the biochemistry of the samples are not altered, and they can be analysed simply by returning the samples to room temperature. Bladder tissue could also be stored in paraffin wax blocks; however it is more difficult to analyse tissue samples that have been stored in paraffin with Raman spectroscopy since paraffin has large distinct peaks that appear in the Raman spectra. Additionally, the dewaxing agents applied for the removal of paraffin wax are known to alter the biochemical integrity of the biological sample, and therefore, alter the corresponding Raman spectra. (140) Despite this, with the use of particular deparaffinization

Histological grade	Description of cancer cells
GX	Unable to determine the extent of differentiation
G1 / Low grade	Well differentiated
G2	Moderately differentiated
G3 / High grade	Poorly differentiated
Pathological stage	**Description of tumour**
T0	No primary tumour
Ta	Non-invasive papillary tumour, confined to urothelium
CIS	High-grade, flat tumour, confined to urothelium
T1	Invasive tumour; invades lamina propria (connective tissue region)
T2	Muscle invasive tumour
T3	Tumour invades perivesical fat layer beyond the muscle region
T4	Tumour invades nearby tissues such as the prostate, vagina, uterus, pelvic wall, or abdominal wall.

Table 6.1: *Description of bladder cancer corresponding to histological grade and pathological staging. (103)*

methods or complex data processing techniques, it is sometimes possible to obtain diagnostic results from these samples. (141)

In 2002, in what is thought to be the first Raman based bladder cancer diagnostic study, Stone et al. investigated the application of conventional Raman spectroscopy to classify, or distinguish between, different pathological grades of bladder tissue. (142) Bladder tissue samples were collected during cystoscopy procedures from 12 patients, and classified as being normal, CIS, low grade, medium grade, and high grade carcinoma by histopathologists. Following PC-LDA classification, the spectra separated into five distinct groups with sensitivities between 78% and 98%, and specificities between 96% and 99%. These results show excellent group separation; however, some overlap was found between CIS and low grade cancer, and between medium grade and high grade cancer, which the authors note is either due to the biochemical continuum of disease, or due to difficulty in distinguishing between pathological groups during histopathological classification.

Members of the previous study, further developed this diagnostic model using Raman spectroscopy to distinguish between a larger range of bladder tissue pathologies in vitro. (143) Conventional Raman spectroscopy was applied to a range of bladder tissue pathologies – normal, cystitis, CIS, G1, G2, G3, G3 squamous dysplasia, and adenocarcinoma. These tissues were obtained from 72 patients during TURBT, TURP, and cystectomy procedures, and samples were classified by a consultant uropathologist and a consultant pathologist. All tissues were snap-frozen and stored at -80 ºC, and then brought back to room temperature for measurement. This study resulted in sensitivities and specificities greater than 82% and 95% respectively.

De Jong and colleagues (2006) used Raman spectroscopic mapping (mapping refers to biochemical images obtained using Raman micro-spectroscopy to create a spectral model that

could differentiate between tumour and non-tumour bladder tissue in vitro. (144) Fifteen snap-frozen bladder tissue samples were obtained consisting of normal, cystitis, T2-T3, and CIS, as confirmed by a pathologist. Spectra were obtained from each tissue sample in order to create a 2-D Raman map of the area. PCA was applied to cluster the spectra into 90 separate clusters, whereby the spectra within each cluster showed a similar biochemical composition. Each cluster was assigned a colour code, and in this way, images were created highlighting areas of similar biochemical composition. Following LDA based classification, 100% sensitivity and 95% specificity results were achieved.

And finally, Grimbergen et al. (2009) investigated the potential of a combined diagnostic modality using fluorescence with conventional Raman spectroscopy. (145) This was achieved by applying Raman spectroscopy to bladder biopsy samples, with and without the presence of 5-ALA, in vitro. 5-ALA is a photosensitive dye that leads to an accumulation of photoporphyrin IX (PpIX) in malignant cells, and is often used in fluorescence cystoscopy to improve the detection of bladder cancers. 92 biopsies were obtained from 73 patients without 5-ALA during TURBT, and 38 biopsies with 5-ALA were obtained from 19 patients that underwent fluorescence-guided endoscopy, which were evaluated as normal, cystitis, CIS, G1, G2, and G3 by a histopathologist. Standard Raman spectra were measured from all biopsy samples, and classification of non-5-ALA samples, based on an algorithm trained from non-5-ALA samples, resulted in 88% sensitivity and 80% specificity. However, when applied to the 5-ALA samples, a sensivitiy of 42% and a specificity of 71.1% were found, indicating a change in the biochemistry of bladder tissue in the presence of 5-ALA. Further algorithms were developed based on a combination of Raman spectroscopy and fluorescence prediction measurements, resulting in a sensitivity of 100% and specificity of 80.8% for distinguishing between cancerous and non-cancerous biopsy samples containing 5-ALA. This combined modality of Raman spectroscopy with fluorescence would allow for in vivo analysis of bladder tissue.

6.4 Raman cystoscopy

Raman fiberoptic probes can be easily inserted during a cystoscopy procedure to give real-time analysis and classifications of the urothelium. Similar to conventional Raman microspectroscopy, fiberoptic Raman probes can also be used to measure samples in vitro and ex vivo. This section provides information on the progression of fiberoptic Raman probes applied to the urinary bladder for diagnostics.

In 2005, Crow et al. were the first group to integrate a fiberoptic probe into Raman spectroscopy to differentiate between benign and malignant bladder tissue in vitro. (146) Bladder tissue samples were obtained from 24 patients during TURBT and TURP, and classified as normal, cystitis, or UC by a uropathologist. Raman spectra were obtained by holding the probe in contact with these tissue samples. Based on PC-LDA classification, a sensitivity of 89% and specificity of 79% was achieved for separating benign and malignant tissues. The probe used

in this study was designed to be compatible with the working channel of a standard rigid or flexible cystoscope so that it could easily be integrated into in vivo studies. While it is proposed that the probe could be used in vivo, the reported experimental work focused on bladder tissue in vitro as a proof of concept.

Grimbergen and colleagues (2009) also investigated the potential of using Raman spectroscopy for bladder tissue diagnosis during cystoscopy by examining bladder tissue biopsies ex vivo using an endoscopic probe. (147) 107 bladder tissue samples were obtained from 54 patients during TURBT, where 5-ALA was used to enhance tissue contrast. Samples were graded as normal, cystitis, CIS, G1, G2, G3, and atypia by a pathologist. Raman spectral measurements were obtained from fresh tissue samples, resulting in a sensitivity and specificity of 78.5% and 78.9%, respectively. Whilst this study was based on ex vivo experimentation, these results show the possibility of using a Raman spectroscopic probe for discerning normal from malignant bladder tissue in vivo in the presence of 5-ALA.

Draga et al. (2010) were the first group to investigate the use of a Raman based fiberoptic probe for the diagnosis of bladder cancer in vivo. (148) Raman spectra were obtained during TURBT procedures on 17 patients without any photosensitive dyes, 12 patients with 5-ALA, and 9 patients with HAL, representing a mixture of normal, Ta, T1, and T2 bladder tissues. Spectra were measured with a high-volume Raman probe, and diagnostic algorithms were designed using PC-LDA, resulting in a sensitivity of 85% and a specificity of 79%. Draga et al. successfully demonstrated the use of Raman spectroscopy for the diagnosis of bladder cancer in vivo, both with and without the presence of photosensitive dyes, showing the potential of integrating a Raman probe into fluorescence cystoscopy procedures to improve diagnosis.

In 2012, Barman et al. proposed the use of a confocal fiberoptic Raman probe to increase the specificity (in terms of tissue depth discrimination) for bladder cancer diagnosis in this proof-of-concept study performed ex vivo. (149) Bladder tissue samples were excised from 14 patients during TURBT, and confirmed as being normal or cancerous by a uropathologist. Raman spectra were obtained from each sample (ex vivo) using both a high-volume probe and a confocal probe to compare the sensitivities and specificities achieved for each. The confocal probe was designed by placing a pinhole aperture into the high-volume probe to decrease the depth of field from 520 μm to 280 μm, thus suppressing the spectral information from surrounding regions and from deeper tissue layers beyond the region of interest. Diagnostic algorithms were developed using PCA and logistic regression analysis (LRA), with the high-volume probe producing a sensitivity of 85.7% and specificity of 85.7%, whereas the confocal probe had a sensitivity of 85.7% and specificity of 100%. The significant increase in specificity values of the confocal probe in comparison to the high-volume probe are associated with the smaller depth of field values, giving this particular device an advantage in the application of Raman probes for real-time in vivo diagnosis of bladder disease.

6.5 Raman based urine cytology

Within the last decade, studies have begun investigating the potential use of Raman micro-spectroscopy for analysing and classifying disease from bladder cells rather than tissue, thus moving to the completely non-invasive procedure of detecting bladder cancer from voided urine. This technique has the potential of replacing standard biopsy or cytology procedures. Within this section, we attempt to extrapolate the best practice (in advance of our own contributions as part of this thesis, which are detailed in Chapters 7-10) in terms of the methods and tools used.

The first study to analyse bladder cells using Raman spectroscopy was investigated by Harvey et al. (2008), where Raman optical tweezers were used to trap and analyse both live and chemically fixed bladder cells, and to differentiate between normal and cancerous cells. (57) In the first part of this study, the Raman tweezers technique was applied to discriminate between a malignant bladder cell line (MGH-U1, otherwise known as T24 (150)) and a malignant prostate cell line (PC-3). All of the cells were fresh and stored in PBS before and during measurement. Raman spectra were obtained using a 514.5 nm laser, and all spectra were pre-processed by removing the baseline contribution, followed by EMSC. Spectra were then subjected to PCA analysis to discriminate between both cell types, and the PCA plot of this shows that they can be identified as separate cell groups. The second part of the study investigated the use of Raman tweezers to discriminate between the following alcohol fixed cell lines: malignant bladder cells (MGH-U1), malignant prostate cells (LNCaP, PC-3), and benign prostate cells (BPH). All cells were fixed using SurePath, and stored at $<4^oC$. Raman spectra were recorded and pre-processed the same way as for live cells. Spectra were then subjected to PC-LDA to differentiate between each cell group, resulting in sensitivities of 73–94% and specificities of 95–98%. This initial study based on the application of Raman tweezers to differentiate between different urological cell lines has shown the diagnostic potential of this technique for detecting both bladder and prostate cancer.

Harvey et al. (2009) continued on from this initial study to further investigate the application of Raman tweezers to classify cultured cell lines that were chemically fixed using either SurePath or formalin, as well as investigating the effect of exposure to urine on these cells. (55) Raman tweezers were first used to distinguish between the following cell lines fixed with SurePath: malignant bladder cells (MGH-U1), malignant prostate cells (LNCaP, PC-3), normal prostate cells (PNT2-C2), and benign prostate cells (BPH). Raman spectra were recorded from 100 MGH-U1, 104 BPH, 104 LNCaP, 115 PC-3, and 110 PNT2-C2 cells. A diagnostic model was designed by applying PC-LDA to the spectra, resulting in sensitivities of 73–100% and specificities of 93–98%. The authors note that the MGH-U1 and PC-3 cells were found to be larger in SurePath than when unfixed. The second section of this study was based on the use of Raman tweezers to distinguish between the following formalin fixed cells: malignant bladder cells (MGH-U1), urethral cells (from patients), benign prostate cells (BPH), normal prostate cells (PNT2-C2), and malignant prostate cells (LNCaP, PC-3). Raman spectra were

recorded from 135 cells in each group, and following PC-LDA classification, results showed 88-100% sensitivity and 98-100% specificity. A comparison between both chemical fixing methods showed that formalin resulted in better diagnostic performance and a greater SNR, although this is due to the higher laser power used for these measurements. There was also no contamination in spectra from the formalin solution, whereas a significant contribution was found in the SurePath spectra. And finally, the last part of this study was to determine the effects of exposure to urine on the ability to classify cells. To investigate this, all cells (MGH-U1, BPH, LNCaP, and PC-3) were unfixed and stored at room temperature before being exposed to urine for the following time durations: 15 min, 30 min, 45 min, 1 h, 2 h, 4 h, 6 h, 8 h, 10 h, and 12 h. PC-LDA was applied to the spectra, and it was shown that in general the prediction values do not deteriorate over 12 h. Overall, this study has shown the potential of using Raman tweezers to distinguish between bladder cells and prostate cells, and also between cancerous and non-cancerous cells.

In 2011, Shapiro et al. investigated the possibility of applying Raman spectroscopy to bladder cells from voided urine in order to determine whether or not UC was present. (151) For this study, fresh urine samples were obtained from 340 patients (116 without UC, 92 with low grade UC, and 132 with high grade UC). To prepare the urine samples for Raman spectroscopic measurements, the cell pellet was obtained from the urine sample following centrifugation and filtration. The cells were then cytospin deposited onto aluminium slides. Raman spectra were obtained from an average of five cells per slide using a 532 nm laser. The resulting spectra showed a positive 1584 cm^{-1} Raman peak for cancerous cells that was not present for nonmalignant cells. Shapiro and colleagues developed a model based only on this peak (no PCA or LDA required) that could classify spectra as normal, low grade, or high grade UC by evaluating the height of the 1584 cm^{-1} peak above the baseline at 1500 cm^{-1}, resulting in an overall sensitivity of 92% and a specificity of 91%. A second part of this study involved comparing the Raman spectra from these bladder cells with spectra obtained from bladder tissue samples. These bladder tissue samples were obtained from surgical touch preps and were frozen prior to Raman measurements. This study showed that the spectra from bladder cells and tissue were virtually identical, and that it is possible to determine the histological grade of both cells and tissue based on the height of the distinct Raman peak at 1584 cm^{-1}.

Canetta et al. (2011) developed the first application of modulated Raman spectroscopy (MRS) for the identification of human bladder cell lines in urine, whilst monitoring cell viability after exposure to urine, and the differences between standard Raman spectroscopy (SRS) and MRS for cell classification were also examined. (91) Healthy human urothelial cells (SV-HUC-1) and malignant bladder cells (MGH-U1) were cultured, and then exposed to a solution containing 2 parts fresh urine and 1 part PreservCyt for the following time durations: 0 h (control), 30 min, 3 h, and 6 h. This solution was then centrifuged, and the cells were resuspended in PreservCyt for storage. Prior to Raman spectral measurements, the cells were washed with PBS before being resuspended in PBS, and placed in a sample chamber between a quartz slide and a quartz

coverslip. Cells were allowed ∼30 min to sediment onto the quartz before Raman measurements were obtained. Spectra were recorded from 40 urothelial cells and 40 bladder cancer cells for each of the urine exposure times using a 785 nm modulated laser (modulated with a ramp wave of 50% symmetry, 40 mHz modulation frequency, 60 GHz modulation amplitude - corresponding to 0.2 nm). For each cell, 40 stacked spectra were recorded with a total integration time of 200 s. PCA was then applied to classify these spectra, resulting in sensitivities of 80–95% and specificities of 87–95%. These results show that it is possible to distinguish between both cell groups using MRS, even after exposure to urine for 6 h. The second part of this study involved monitoring cell viability in urine using 1% Fast Green dye (Sigma Chemical Co. Ltd., USA), and it was found that after 1 h, the viability of the cancer cells decreased to ∼60% and ∼64% for healthy cells, and the reproductive capacity of both cells decreased rapidly after exposure to urine. This mirrors the Raman data obtained from different exposure times where changes in the biochemistry can be seen with progressing exposure times. The robustness of MRS in comparison to SRS was then investigated by comparing the Raman spectra obtained from each method for fixed urothelial cells and fixed bladder cancer cells that were not exposed to urine (control group). The SRS spectra showed a strong autofluorescence signal from the quartz substrate, and Raman peaks from the cells are not well defined. On the other hand, the MRS spectra allow for Raman peaks to be clearly observed. PCA was applied to both the SRS and MRS spectra to discriminate between both cell groups, and a Fourier transform was also applied to the SRS spectra to estimate between the slow varying background and the Raman peaks, resulting in a sensitivity of 97% and specificity of 72%. In Fourier space, the background signals were present at low frequencies, whereas the Raman signals were present at high frequencies, therefore the Fourier transform was able to separate the most important Raman signals from the background. Higher sensitivities and specificities were achieved with MRS, which may be due to the better SNR obtained by filtering out the Raman peaks associated with the biological cells from the fluorescence background signals. It has also been noted that the healthy urothelial cells show relatively strong protein and carbohydrate Raman peaks (1100–1300 cm^{-1} region) in comparison to the cancerous cells, and there was an increase in ring breathing modes for the cancerous cell spectra (669, 727, 785, 828, 1095, 1578 cm^{-1}) indicating a higher concentration of DNA in malignant cells.

6.6   Summary

It is clear from the results reviewed in this chapter that there is growing potential for the application of Raman spectroscopy to the study of urinary bladder disease in order to improve patient treatment and patient outcome. This technique can be applied in a number of different forms, including conventional Raman micro-spectroscopy, as well as using Raman tweezers or fiberoptic probes. Raman micro-spectroscopy has been shown to be a strong candidate for improving the diagnostic sensitivity and specificity of urine cytology, with particular interest

for the case of low grade UC cells, for which a recent (and the first) patient study showed a sensitivity and specificity of 92% and 91% respectively. However, this study applied Raman micro-spectroscopy to unfixed cells which is not an ideal method to introduce into a clinic due to the possible presence of highly infectious disease. Furthermore, spectra were classified based on the 1584 cm^{-1} peak (151), an assignment that has not been acknowledged for the identification of bladder cancer elsewhere in the literature.

Raman fibers have been shown to return similar results for both in vitro/ex vivo and in vivo samples, allowing for their potential integration into cystoscopic procedures. The improvement in sensitivity obtained for Raman cystoscopy over existing diagnostic procedures is not only due to the additional biochemical information provided by the Raman spectrum, and the ability to classify this using multivariate diagnostic algorithms, but it is also partially due to the reduction in inter- and intra-observer variation, which can vary by up to 45–50% using current clinical methods. (139; 152)

This chapter highlights the potential for Raman micro-spectroscopy in the diagnosis of bladder cancer, both in vivo via a cystoscopic procedure, and in vitro to urine cytology samples. However, this chapter also highlights the shortcomings that have been observed to date, in terms of the variability of experimental parameters and consumables that have been utilised across different research groups. The lack of standardisation, and in particular the lack of advancement towards the translation of Raman technology towards the clinic, is something that needs to be addressed. In order to identify the optimum methodologies for Raman based urine cytology, Chapter 7 of this thesis investigates a range of source wavelengths and sample substrates, and Chapter 9 systematically compares different cell deposition techniques, fixation agents, and the impact of applying red blood cell lysing agents to urine samples prior to Raman micro-spectroscopic analysis, with the aim of integrating Raman micro-spectroscopy into current clinical protocols insofar as possible.

Chapter 7

Sample substrates and source wavelengths for Raman based cytology

This chapter has been published as part of the following paper:

- L.T. Kerr, H.J. Byrne, B.M. Hennelly. *Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen.* Anal. Methods, 7(12), 5041-5052, 2016.

7.1   Introduction

In the preceding chapters, Raman micro-spectroscopy has been introduced as an experimental technique, and it's application to understanding the etiology and diagnosis of bladder cancer was reviewed. In particular, improving the diagnostic potential of urine cytology was identified as a key area of research. However, our review of progress in this area, and indeed in other related applications of Raman micro-spectroscopy, has highlighted a high degree of variability across experimental protocols, some of which result in large background signals that can often overpower the weak Raman signals being emitted. These protocols need to be standardised before the technique can provide reliable and reproducible experimental results in an everyday clinical environment. The objective of this chapter is to investigate the impact of different experimental parameters involved in the analysis of biological specimen. This study investigates the Raman signals generated from healthy human cheek cells using different source laser wavelengths; 473 nm, 532 nm, 660 nm, 785 nm, and 830 nm, and different sample substrates; Raman-grade calcium fluoride, IR polished calcium fluoride, magnesium fluoride, aluminium (100 nm and 1500 nm thin films on glass), glass, fused silica, potassium bromide, sodium chloride, and zinc selenide, whilst maintaining all other experimental parameters constant throughout the study insofar as possible. The results obtained throughout this study are presented both graphically and numerically. Numerical results are based on the analysis of the spectral data using three different statistical metrics: discrete correlation, normalised covariance, and mean square er-

ror (MSE).

The breakdown of this chapter is as follows; firstly the importance of standardising the experimental parameters associated with Raman micro-spectroscopy is discussed. Section 7.3 provides information about the experimental materials and methods used in our study. Results from all experiments are shown in Section 7.4, followed by a discussion of these results in Section 7.5, and a summary in Section 7.6.

7.1.1   Standardisation of Raman micro-spectroscopy

While Raman micro-spectroscopy has been demonstrated to produce accurate diagnostic results, further development is necessary to ensure that Raman spectroscopic systems are sufficiently robust for everyday clinical usage. (153) The lack of standardisation, in terms of equipment, consumables, and measurement protocols has resulted in the recording of significantly differing spectra across studies to date, and it is believed that this has hindered further advancement of this technique. Even within the application of Raman micro-spectroscopy to the diagnosis of one particular pathological disease (e.g. bladder cancer), a wide range of sample substrates, source laser wavelengths, and integration times have been applied to date, as discussed in Chapter 6; all resulting in moderately varying spectra for the same disease. (154) If Raman micro-spectroscopy is ever to become a commonly used clinical tool, then it is important for standardised procedures to be established in order to overcome these inconsistencies. (155) This chapter aims to identify the optimum source wavelength and sample substrate for Raman micro-spectroscopic measurements of biological samples, and also to identify the similarities, or lack thereof, between spectra recorded using these different experimental parameters.

7.2   Raman sample substrates

Sample substrates for Raman micro-spectroscopy should produce low background signals, be biocompatible, and non-toxic for the cells and tissues placed on them, and be as cost effective as possible. Previous studies have reported the use of a wide range of substrates including calcium fluoride (145; 156; 144; 141), aluminium (151; 147; 157), quartz (91; 89; 158), glass (159; 78), and 3D collagen gels (160). However, substrates that produce low background signals for NIR sources are often expensive, and are available in different levels of purity/film thicknesses, which can produce variable results with Raman micro-spectroscopy. Therefore, this study analyses a wide range of sample substrates and source wavelengths in order to identify which substrates produce the lowest background signals at each wavelength, while monitoring the cost effectiveness associated with each substrate.

An important consideration that is not included in this study is the biocompatibility associated with each of the various substrates. This is particularly important for experimentations involving living cells, which often involve the incubation or growth of cells directly onto the

substrate. However, this has been previously explored by Meade et al. (2006), whereby substrates were coated with fibronectin, laminin, and gelatin, resulting in improved cell proliferation and similar Raman spectra to those achieved without the use of any coatings. (161)

7.3 Materials & methods

7.3.1 Sample preparation

Fresh cheek cells are swabbed from a single healthy human volunteer, placed directly onto each of the substrates before each set of experiments, and are allowed to air dry for 5 minutes before measurement. No additional preserving/fixing agents, or washing steps, are applied to the samples, and therefore some spectral signals relating to dried saliva, or other oral contaminants, may be present across all spectra. It is noted that no debris or contaminants are visible in the images of the cells that are used in this experiment, and there is no obvious contamination observed in any of the spectra recorded. While the presence of saliva and other possible contaminates is acknowledged, the associated Raman signals are not expected to significantly affect the results presented in Section 7.4. It is noted that all experiments were performed in compliance with our institute's Ethics Committee (BSREC-2015-008), and informed written consent was obtained from the donor prior to experimentation.

7.3.2 Raman spectral acquisition

Raman micro-spectroscopy measurements are performed across two similar commercial Raman systems (both Horiba Jobin Yvon LabRam HR 800) with Synapse cooled CCD detectors, as discussed in Section 3.6.2. These systems operate with a 50x MO (Olympus MPlanN, 50x/0.75, ∞/0/FN22), 50 μm confocal aperture, and a 300 gr/mm grating. Spectra are recorded with an acquisition time of 30 s averaged over 2 iterations, allowing for the subsequent removal of cosmic rays. (162) It should be noted that the confocal aperture is opened to 100 μm for measurements with the 830 nm source due to the low photon count and lower quantum efficiency of the CCD in this region. Both systems are calibrated using a silicon wafer prior to measurements.

Background spectra are recorded from each substrate at every wavelength, and are used to isolate the Raman signals from the recorded biological spectra using the background subtraction method that was previously explained in Section 4.4.2.

The following substrates are used for this study:

- calcium fluoride – Raman-grade (Crystran Ltd., UK),
- calcium fluoride – IR grade polished (Crystran Ltd., UK),
- magnesium fluoride (Crystran Ltd., UK),
- 100 nm aluminium thin film on glass (Deposition Research Laboratory Inc., USA),

- 1500 nm aluminium thin film on glass (Deposition Research Laboratory Inc., USA),

- glass,

- fused silica – IR grade polished (Crystran Ltd., UK),

- potassium bromide (Edmund Optics, UK),

- sodium chloride (Edmund Optics, UK),

- zinc selenide (Crystran Ltd., UK).

Measurements made using the 473 nm, 660 nm, and 830 nm lasers are performed on one Raman system, and measurements at 532 nm and 785 nm on the other system, with all other parameters maintained as constant insofar as possible; variances however are present in the output power of each source laser (see Section 3.6.2). The lasers, and their corresponding powers, used for this study are similar to those regularly employed in commercial Raman systems for biological measurements.

7.4   Results

Background signals present in Raman spectra are the result of intrinsic Raman signals and/or fluorescence from the sample substrate and other optical elements within the Raman system, as previously discussed in Chapters 3 and 4. The MO, in particular, is known to produce a not insignificant background signal. In order to gauge the contribution of the optical system alone, this section begins by presenting the spectra obtained from the system without the presence of any samples or substrates. The results are shown in Figure 7.1 for each source wavelength. The majority of this signal contribution is associated with the MO used, with the signal being most significant for the 785 nm source.
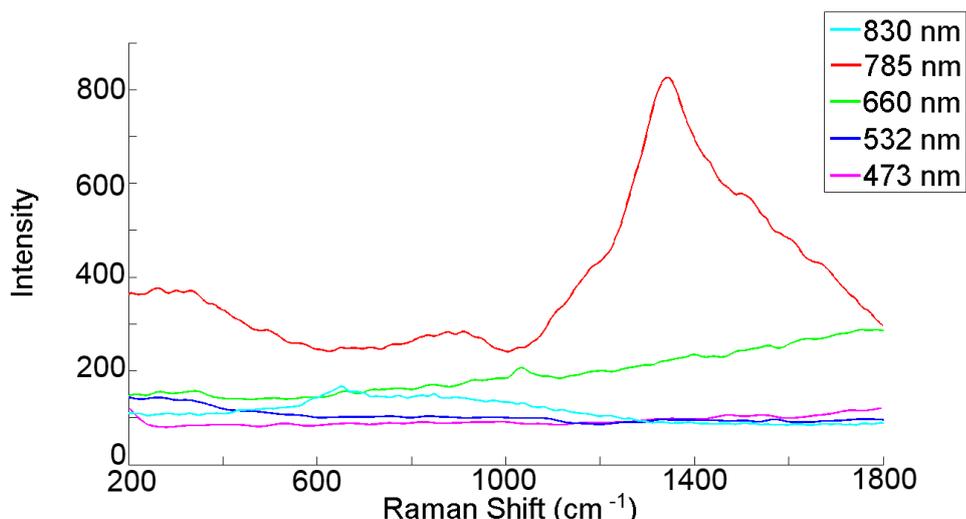
Figure 7.1: *Background signals generated by the optical elements within the Raman system for each of the above wavelengths - a Gaussian smoothing function was applied to all spectra.*

1 - proteins; 2 - nucleic acids; 3 - proteins;
4 - phenylalanine; 5 - amide III; 6 - proteins and nucleic acids;
7 - CH$_2$ mode of proteins and lipids; 8 - amide I
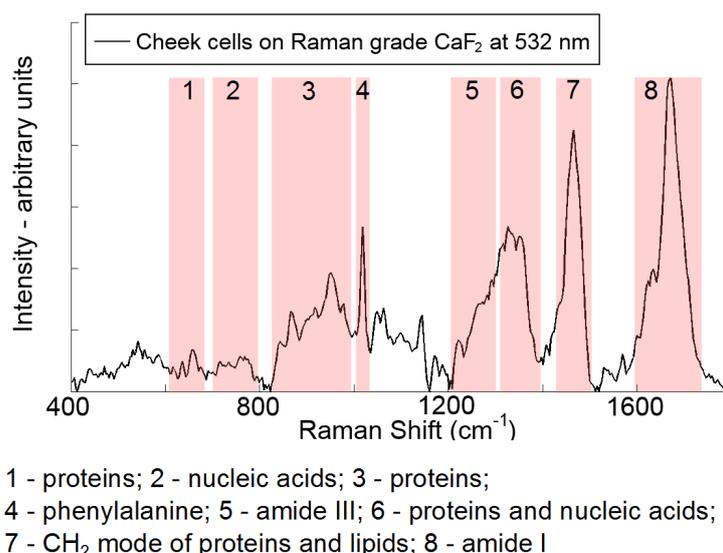
Figure 7.2: *A single Raman spectrum obtained from fresh human cheek cells on Raman-grade calcium fluoride with a 532 nm source laser, with key biomolecular regions highlighted.*

Figure 7.2 represents a single baseline-corrected spectrum recorded from fresh human cheek cells on Raman-grade calcium fluoride using a 532 nm source laser, recorded using the parameters outlined in Section 7.3; this spectrum proved to be the optimum spectrum recorded. Figure 7.2 highlights the key Raman peaks associated with cheek cells, and the important biomolecular regions that are often used in the analysis of cells in Raman based cytology studies have been identified. A more detailed analysis of cellular Raman spectra can be found elsewhere. (163; 164)

The spectra shown in Figure 7.3 to Figure 7.12 represent the background signals and Raman cell peaks associated with each substrate for every wavelength, using the parameters outlined in Section 7.3 (i.e. all spectra are recorded with an acquisition time of 30 s, averaged over 2 iterations). The (a) component of these figures represents the background signals obtained from each substrate at each wavelength, and are shown at their recorded intensity values, after convolution with a Gaussian smoothing function is applied to remove any additional noise contributions. This noise reduction is necessitated by the low photon count, and therefore high shot noise, brought about by the presence of a 50 μm confocal aperture. The corresponding spectra in (b) are recorded from fresh cheek cells on each substrate for each wavelength; these spectra have been baseline corrected using the background subtraction algorithm discussed in Section 4.4.2. Following this, the spectra were area normalised by dividing each spectrum by the sum of its intensity.

Background spectra are recorded between 200–1800 cm^{-1} for all substrates. Cell spectra are generally shown between 400–1800 cm^{-1}. The reason for this is two fold; firstly this is the fingerprint region for biological specimen, and is the most commonly observed spectral region for Raman related biochemical investigations, and secondly, this removes any issues with baseline correction of large background contributions in the lower wavenumber region;

problems can occur when applying the background subtraction algorithm in regions where the background has relatively strong intensities, e.g. the CaF_2 peak at approximately 321 cm^{-1}, which results in a breakdown of the algorithm. This is easily avoided by discarding the lower part of the spectrum for both the cell signal and the background. Additionally, some substrates produce large background signals between 400–600 cm^{-1}, which are not entirely removed by the background subtraction algorithm, such as zinc selenide (for all source wavelengths) and fused silica at 532 nm. For this reason, these spectra (and their corresponding metric values in Tables 7.1–7.3) relate only to the 600–1800 cm^{-1} region.
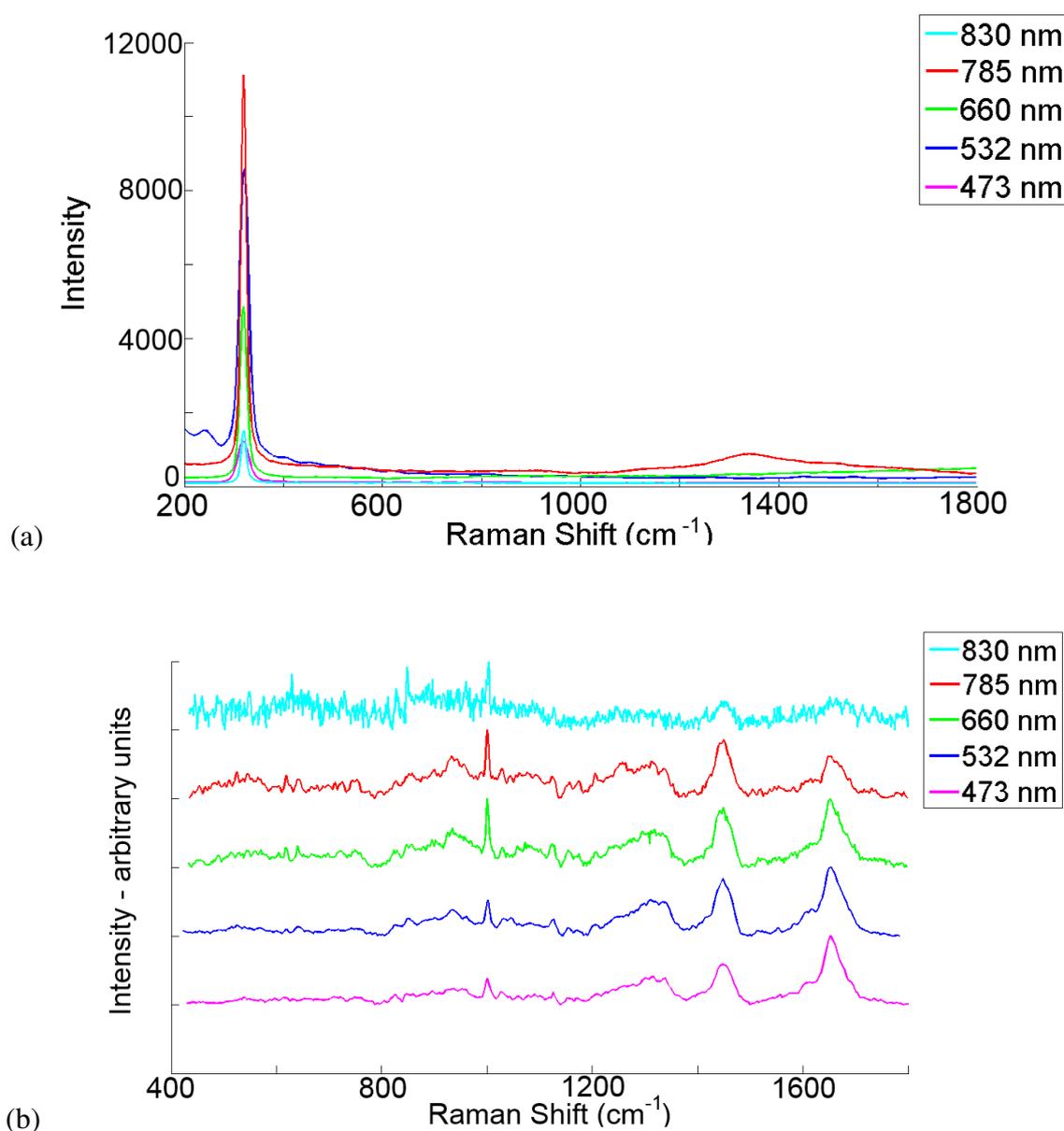
Figure 7.3: *(a) Background signal of **Raman-grade calcium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on Raman-grade calcium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2).*
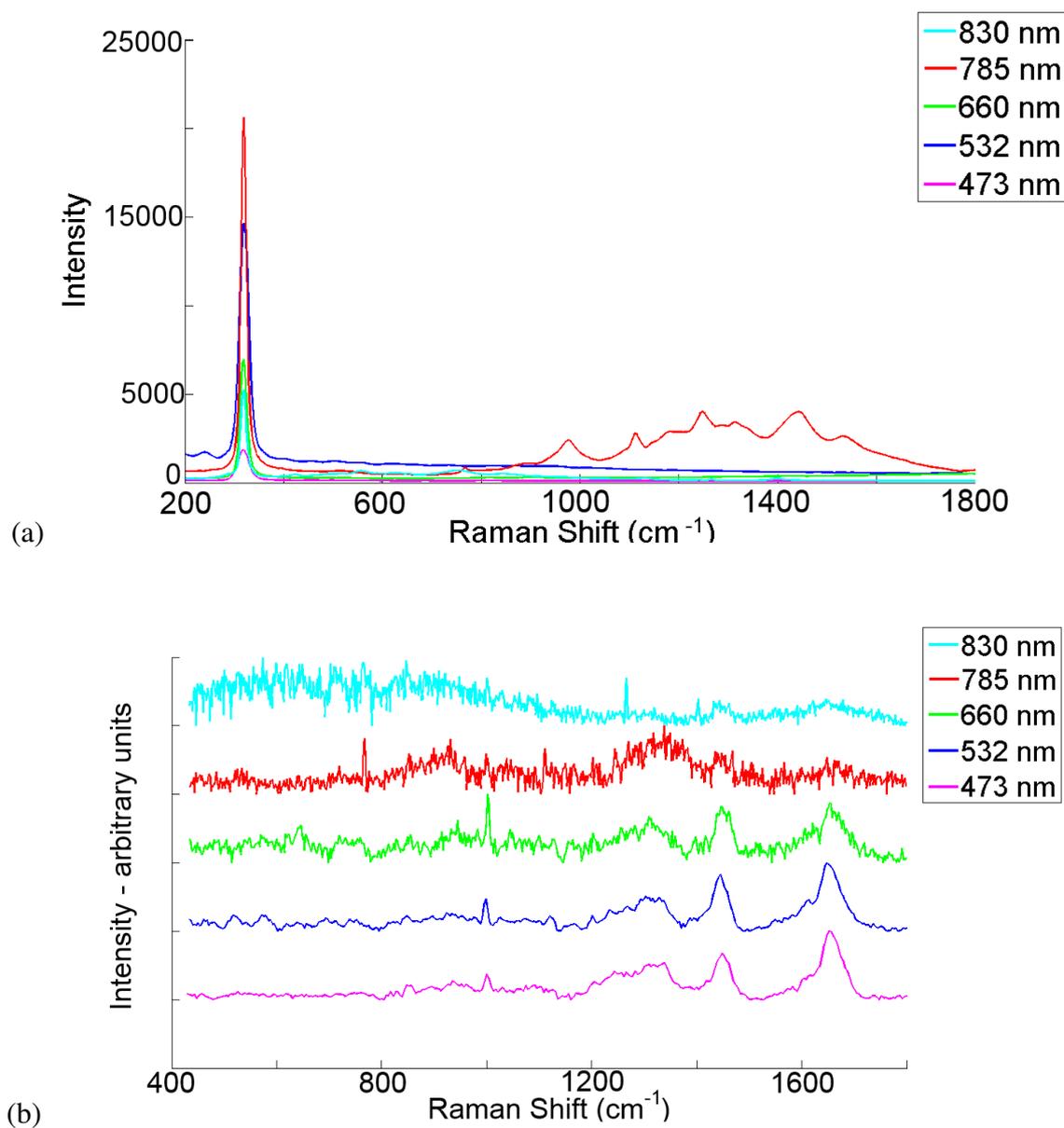
Figure 7.4: *(a) Background signal of **IR polished calcium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on IR polished calcium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2).*
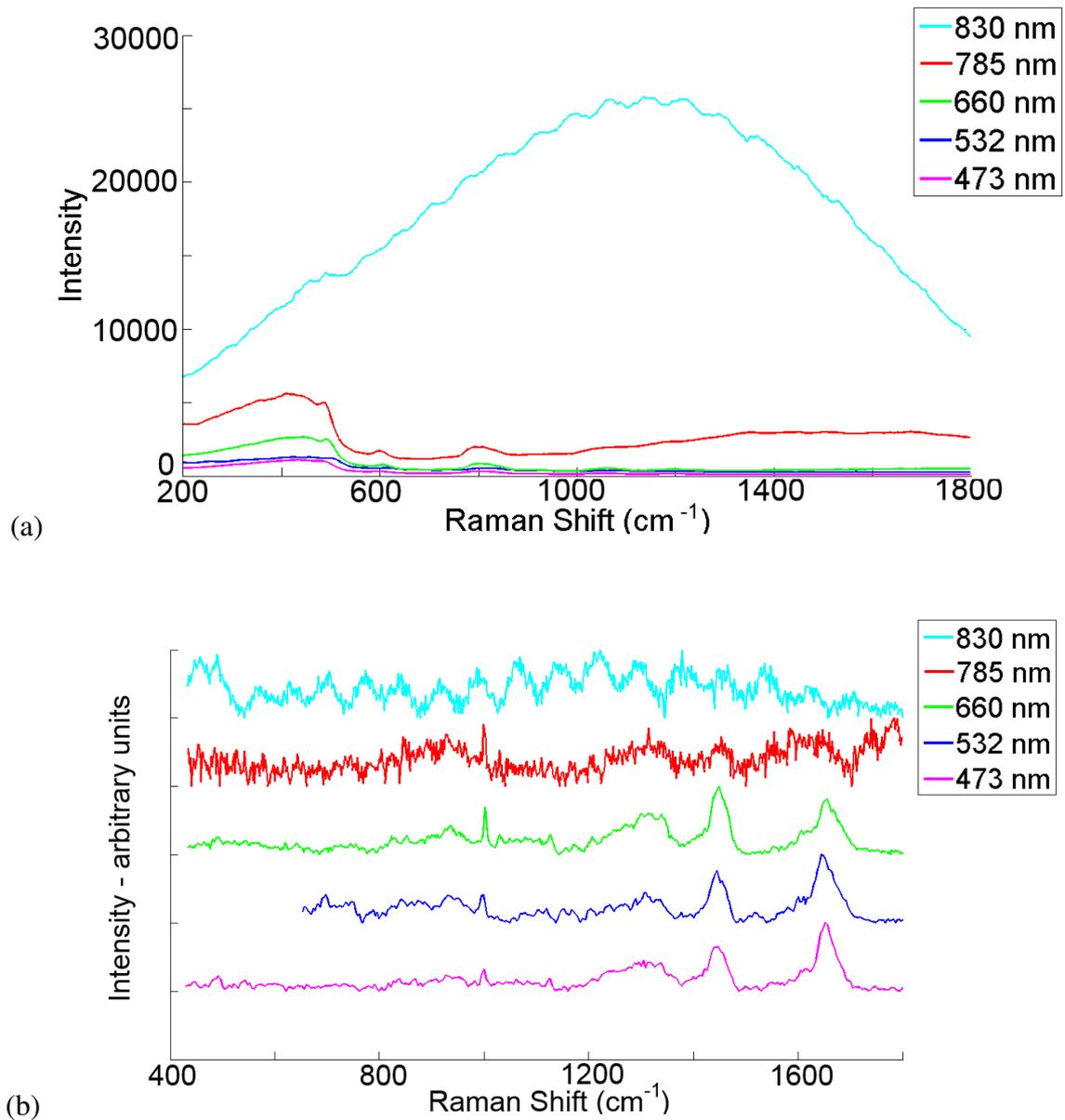
Figure 7.5: *(a) Background signal of **fused silica** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on fused silica for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the cell spectrum at 532 nm is only shown between 600-1800 cm^{-1} due to the large background contribution in the lower wavenumber region.*
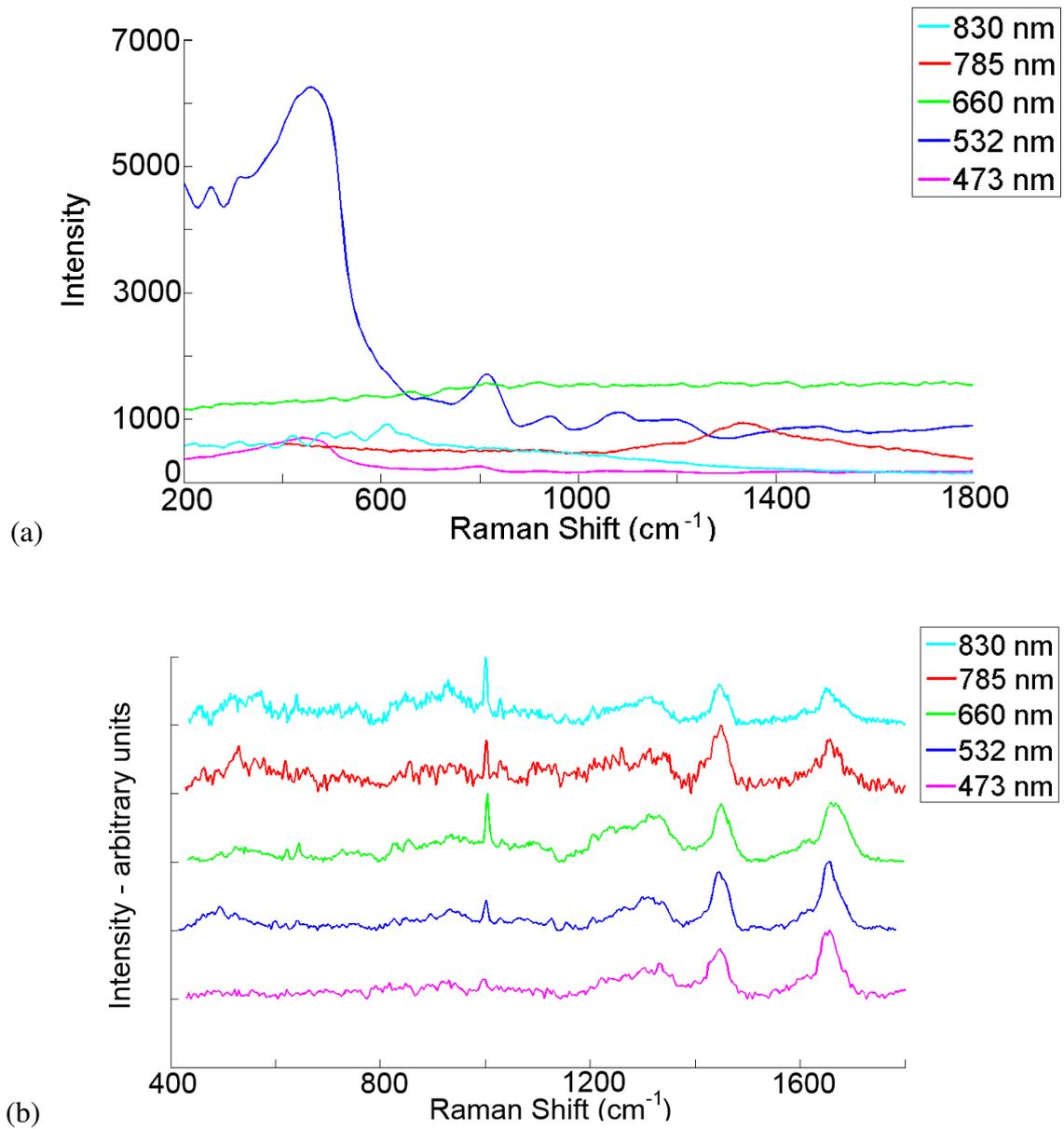
Figure 7.6: *(a) Background signal of **100 nm aluminium thin film on glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on 100 nm aluminium thin film for each of the above wavelengths, with an acquisition time of 30 s (x2).*
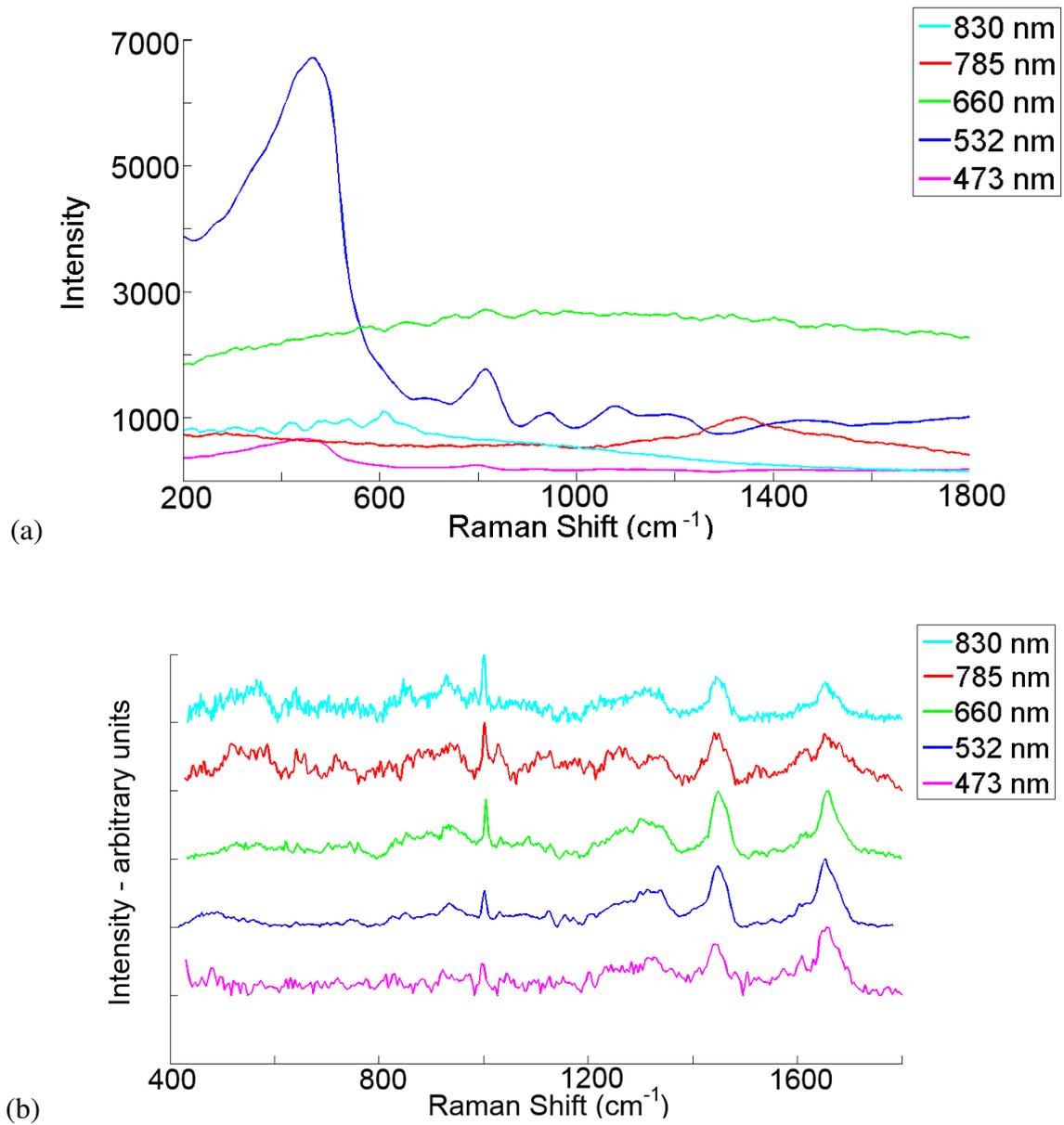
Figure 7.7: *(a) Background signal of **1500 nm aluminium thin film on glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on 1500 nm aluminium thin film for each of the above wavelengths, with an acquisition time of 30 s (x2).*
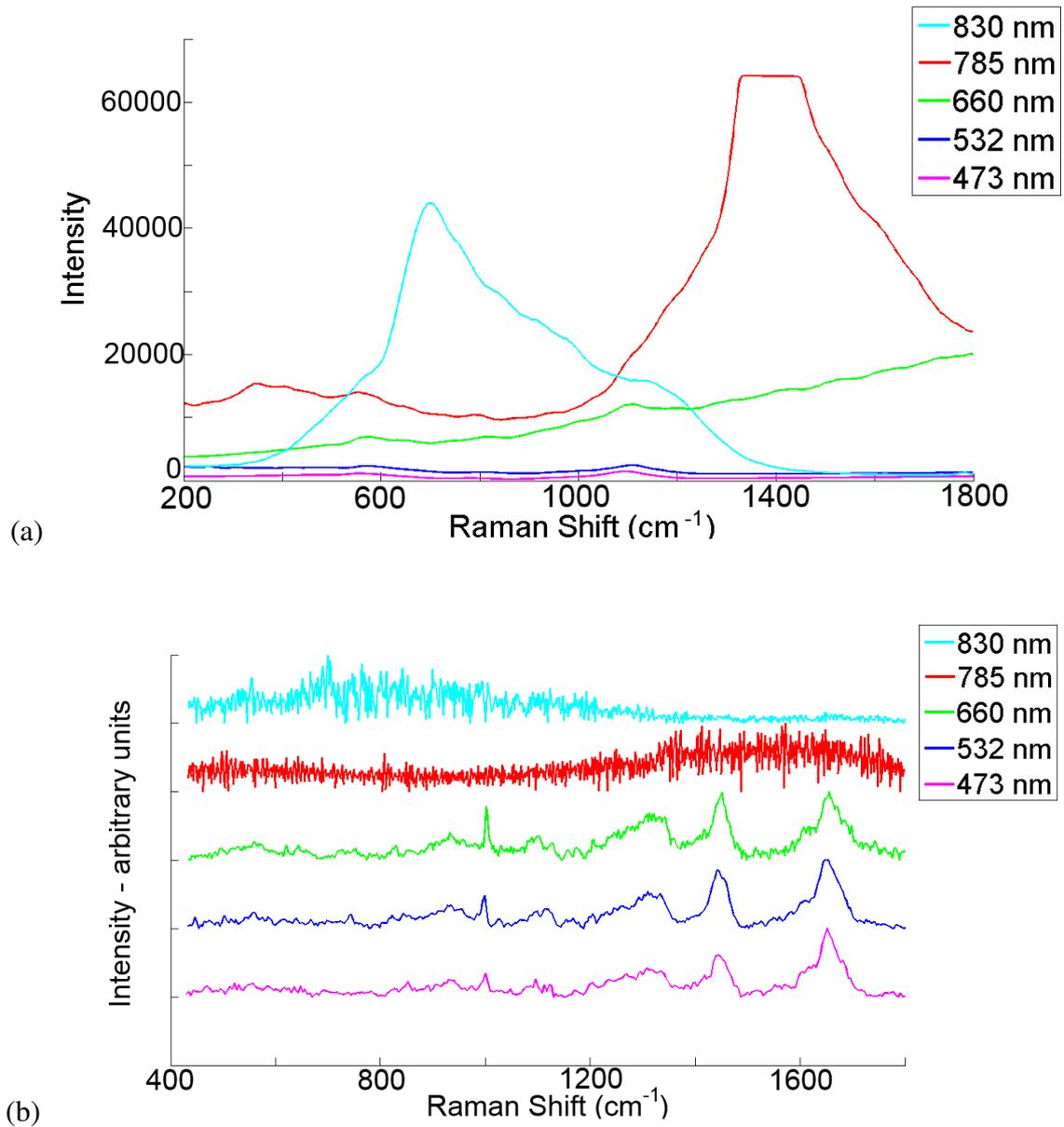
Figure 7.8: *(a) Background signal of **glass** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on glass for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the recorded background signal at 785 nm saturated in the 1400 cm^{-1} region.*
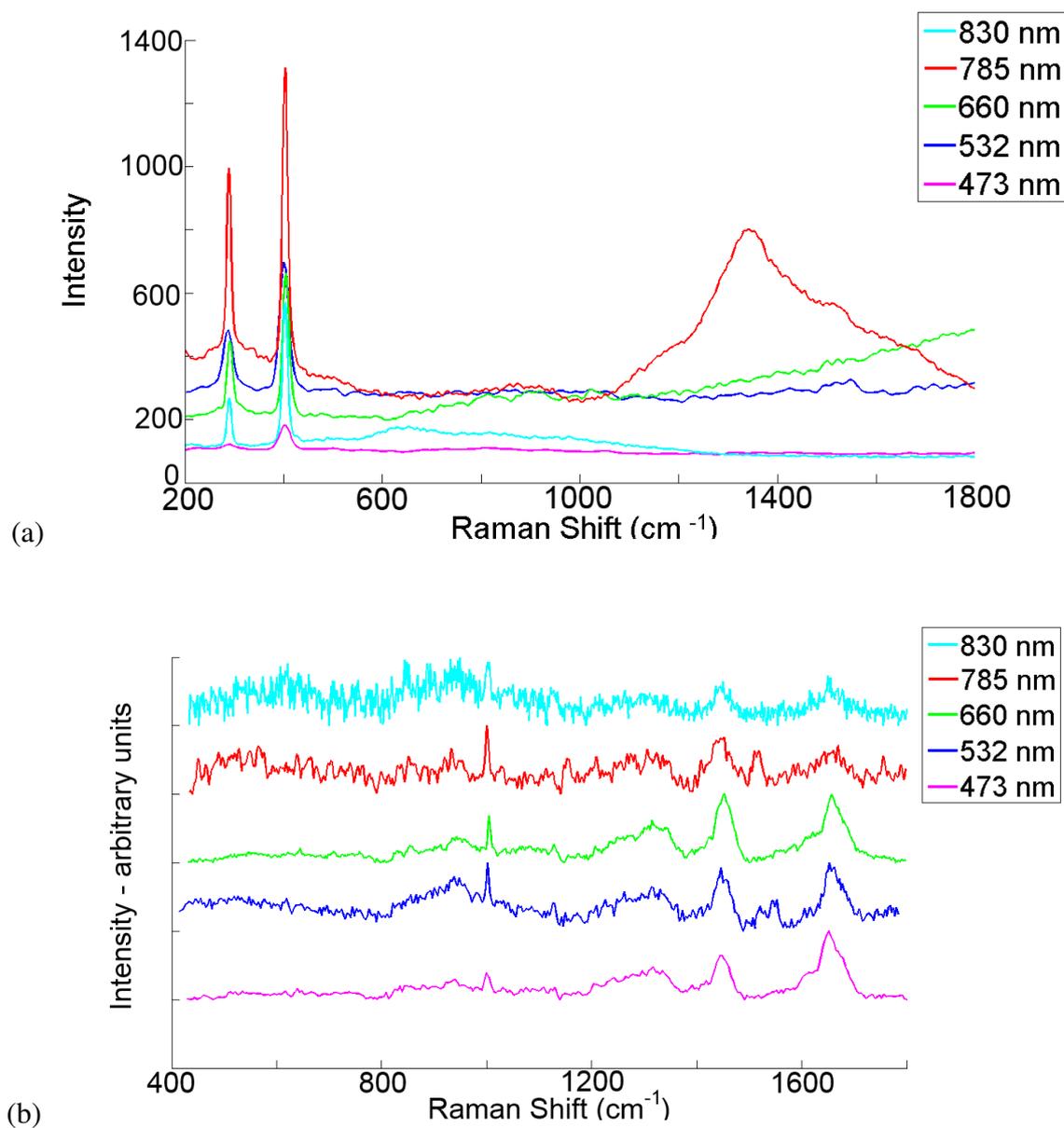
Figure 7.9: *(a) Background signal of **magnesium fluoride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on magnesium fluoride for each of the above wavelengths, with an acquisition time of 30 s (x2).*
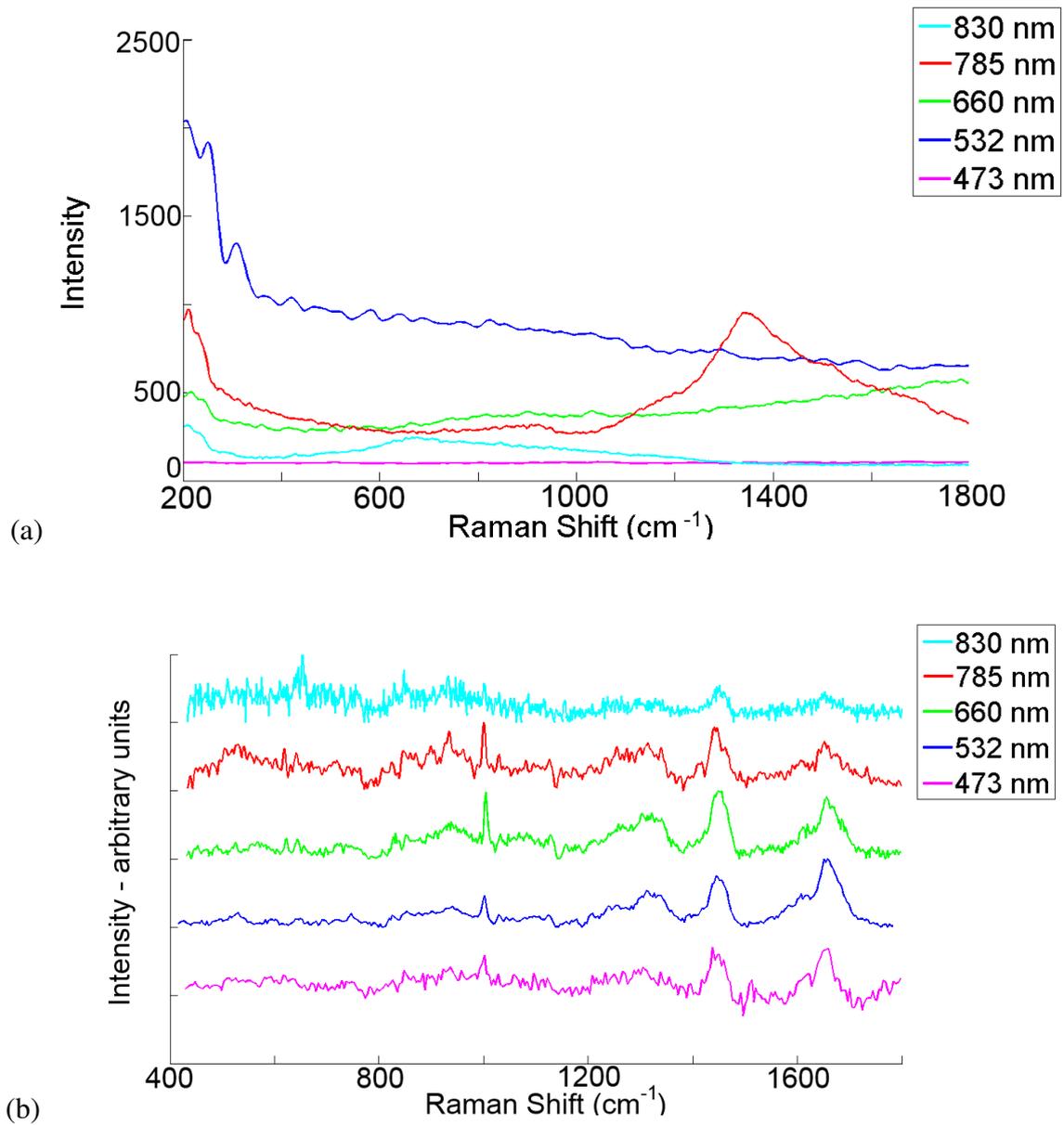
Figure 7.10: *(a) Background signal of* **potassium bromide** *recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on potassium bromide for each of the above wavelengths, with an acquisition time of 30 s (x2).*
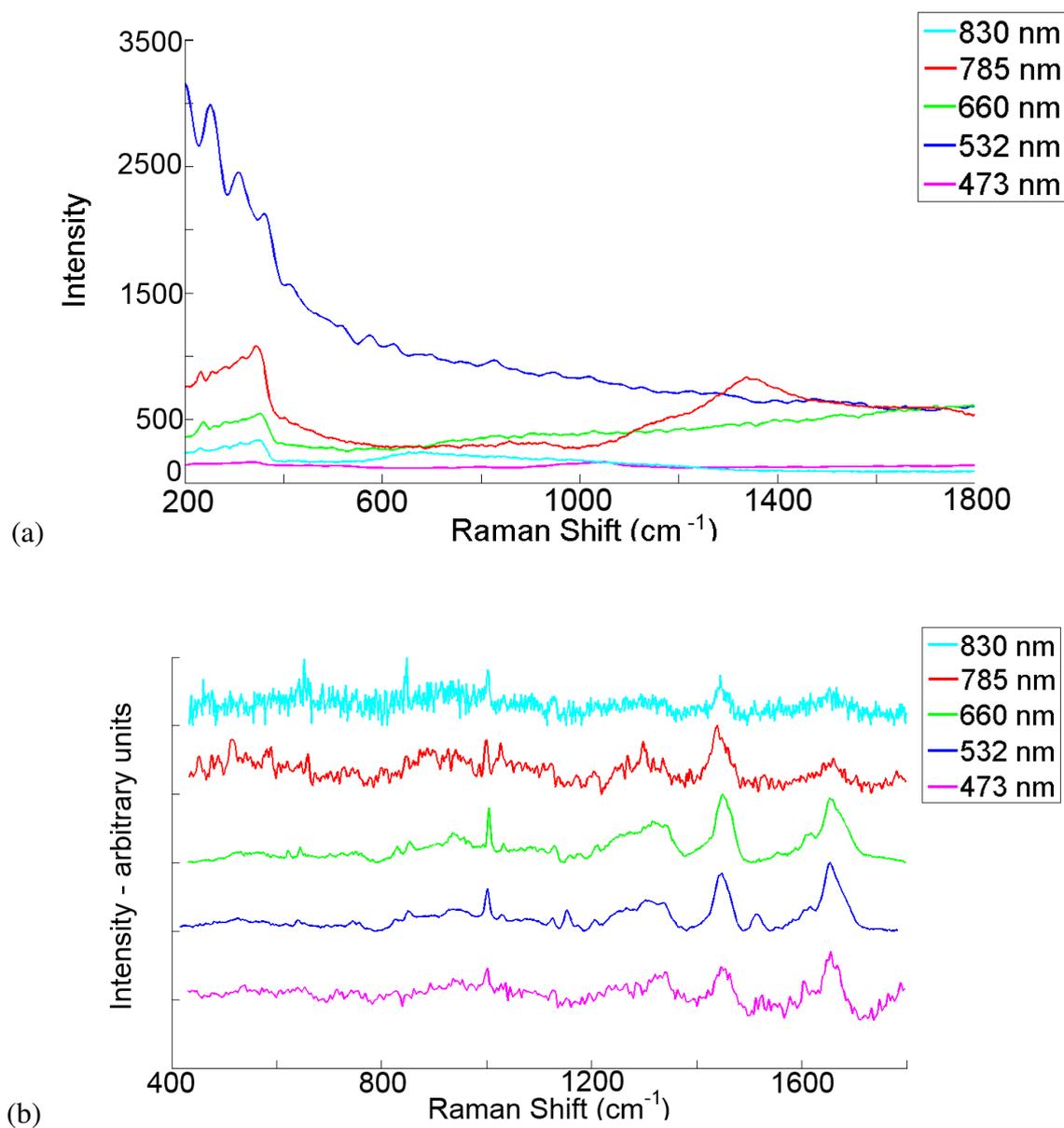
Figure 7.11: *(a) Background signal of **sodium chloride** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on sodium chloride for each of the above wavelengths, with an acquisition time of 30 s (x2).*
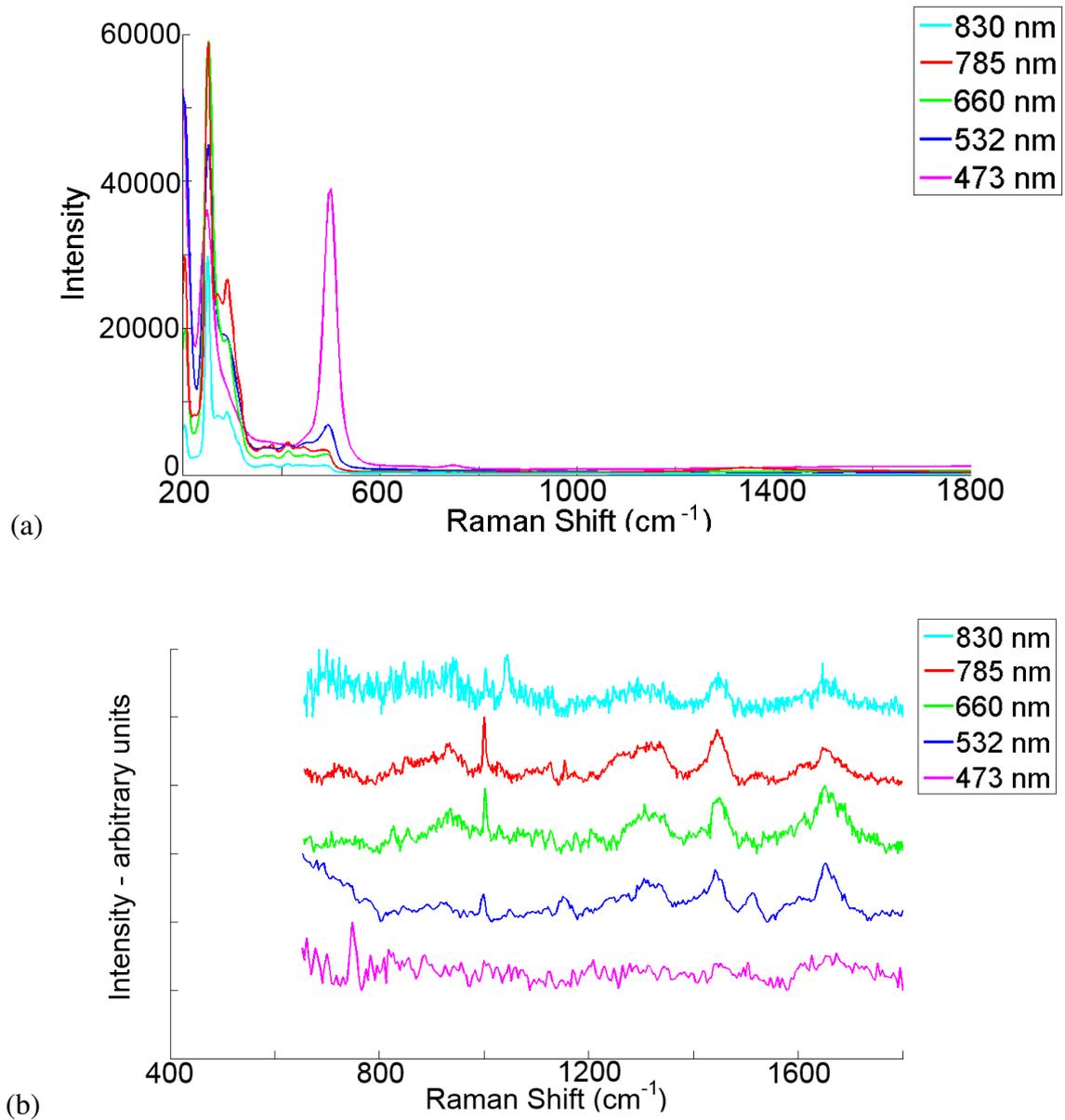
Figure 7.12: *(a) Background signal of **zinc selenide** recorded for each of the above wavelengths, (b) normalised Raman signals from cells recorded on zinc selenide for each of the above wavelengths, with an acquisition time of 30 s (x2). Note, the cell spectra are only shown between 600–1800 cm^{-1} due to the large background contribution in the lower wavenumber region.*

In order to perform a quantitative evaluation of the quality of each of the recorded spectra, the background subtracted spectra are all directly compared to that recorded from cells on Raman-grade calcium fluoride using a 532 nm laser, as seen in Figure 7.2. For the sake of completeness, three separate metrics are employed in this comparison; (i) discrete correlation, (ii) normalised covariance, and (iii) MSE. These three metrics compare overlapping regions of spectra, and thus, produce a value representing how similar the spectra are to that recorded from Raman-grade calcium fluoride at 532 nm. In this way, it is possible to get a good estimate of how similar the Raman peaks, and the associated SNR, are when recorded across a range of substrates and source wavelengths.

Discrete correlation is defined by the following equation:

$$(X \star Y)(n\delta\lambda) = \sum_{m=0}^{2N-1} X^*[n\delta\lambda]Y[(n+m)\delta\lambda] \tag{7.1}$$

where $\delta\lambda$ is the sampling interval of the two signals X and Y, which both have length N, $\star$ represents correlation, and X^* is the complex conjugate of X. Previously, in Section 3.4.1, the relationship between spectral resolution and laser wavelength for an identical spectrograph grating was discussed. In order to employ the discrete correlation of spectra recorded using different source wavelengths, it is necessary to ensure that they have identical sampling intervals. This is achieved by applying an interpolation function in Matlab. Correlation was then implemented using Matlab's $xcorr$ function. and the corresponding coefficient values that are recorded can be seen in Table 7.1, where 1.000 is the optimum result defining a perfect match between spectra, values above 0.900 are a good match, and lower values represent less correlation between spectra.

The second metric, normalised covariance, has a similar range of values between 0 and 1. Normalised covariance is defined as follows:

$$\text{Covariance} = \frac{(X \cdot Y)^2}{(X \cdot X)(Y \cdot Y)} \tag{7.2}$$

where $(X \cdot Y)$ represents the dot product of X and Y. This metric has previously been applied to compare the similarity between Raman spectra for cosmic ray removal. (165) The corresponding values obtained using this method are available in Table 7.2, where 1.000 is the optimum result, and lower values represent less covariance between spectra.

The third metric, MSE, does not have ranges between 0 and 1, but rather between 0, indicating an identical likeness, and some arbitrary maximum value that is dependent on the differences in the values of the two signals. MSE is commonly used in signal processing to compare the likeness of two signals and is defined as follows:

$$\text{MSE} = \text{mean}((X - Y)^2) \tag{7.3}$$

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF$_2$	0.8741	1.000	0.9189	0.8286	0.6939
IR polished CaF$_2$	0.8726	0.9817	0.8671	0.7882	0.5420
Fused silica	0.8391	0.9422	0.9168	0.7128	0.6218
Aluminium (100 nm)	0.8488	0.9766	0.9211	0.8532	0.7674
Aluminium (1500 nm)	0.8435	0.9854	0.9233	0.8302	0.7837
Glass	0.8473	0.9822	0.9221	0.7469	0.4754
Magnesium fluoride	0.8648	0.5968	0.9403	0.7561	0.7120
Potassium bromide	0.6067	0.9819	0.9112	0.7860	0.6668
Sodium chloride	0.6393	0.9839	0.9354	0.7324	0.7210
Zinc selenide	0.7044	0.3035	0.7643	0.8305	0.4284

Table 7.1: *Comparison of cross correlation ($xcorr$) results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, where all spectra are recorded with an acquisition time of 30 s (x2).*

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF$_2$	0.9429	1.000	0.8814	0.6767	0.1641
IR polished CaF$_2$	0.9191	0.9059	0.7117	0.3558	0.0066
Fused silica	0.9004	0.7682	0.9119	0.2423	0.0318
Aluminium (100 nm)	0.8668	0.8868	0.8457	0.6694	0.4999
Aluminium (1500 nm)	0.7390	0.9245	0.9179	0.5582	0.5167
Glass	0.8817	0.9125	0.8856	0.1905	0.0020
Magnesium fluoride	0.9317	0.5953	0.9343	0.3487	0.2117
Potassium bromide	0.5901	0.9235	0.8927	0.4848	0.1249
Sodium chloride	0.5216	0.9148	0.9512	0.2973	0.2646
Zinc selenide	0.2005	0.5840	0.7944	0.6929	0.2173

Table 7.2: *Comparison of normalised covariance coefficient results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, where all spectra are recorded with an acquisition time of 30 s (x2).*

where mean() denotes the process of taking the average value. For comparative purposes, the MSE values obtained have been presented relative to the maximum value (i.e. all values are compared to the worst spectrum):

$$1 - \left(\frac{MSE}{MSE_{max}} \right) \tag{7.4}$$

This results in values ranging between 0 and 1, which are presented in Table 7.3, where similar to the other metrics, 1.000 represents the optimum result, and lower values represent a greater MSE between both spectra.

Substrate	473 nm	532 nm	660 nm	785 nm	830 nm
Raman-grade CaF_2	0.9634	1.0000	0.9817	0.7973	0.4800
IR polished CaF_2	0.9417	0.9434	0.7986	0.6139	0.1820
Fused silica	0.9360	0.8517	0.9437	0.5428	0.3187
Aluminium (100 nm)	0.9185	0.9273	0.9073	0.6726	0.7001
Aluminium (1500 nm)	0.8399	0.9530	0.9443	0.6151	0.7089
Glass	0.9269	0.9473	0.9293	0.5079	0.0000
Magnesium fluoride	0.9584	0.6758	0.9605	0.6083	0.5197
Potassium bromide	0.6067	0.9536	0.9221	0.6756	0.4489
Sodium chloride	0.7313	0.9540	0.9703	0.5790	0.5595
Zinc selenide	0.3527	0.6398	0.8457	0.7849	0.4377

Table 7.3: *Comparison of mean square error results for each substrate at each wavelength when compared to a cell spectrum recorded on a Raman-grade calcium fluoride substrate using a 532 nm laser, and then compared to the value from glass at 830 nm (i.e. the cell spectrum least similar to that recorded on CaF_2 at 532 nm); where all spectra are recorded with an acquisition time of 30 s (x2).*

7.5 Discussion of results

From a sample substrate perspective, the most consistent Raman spectra, regardless of source wavelength, are obtained from Raman-grade calcium fluoride and both aluminium coated substrates. This has been verified both visually from the recorded spectra (in the ability to identify key biomolecular peaks, and from the SNR seen in the spectra, see Figure 7.3, Figure 7.6, and Figure 7.7), and from all three of the metrics employed in this study, as shown in Tables 7.1–7.3. Good results are also obtained from magnesium fluoride (Figure 7.9), potassium bromide (Figure 7.10), and sodium chloride (Figure 7.11).

Focusing on the NIR source wavelengths used here (785 nm and 830 nm); while it has been shown that a number of substrates provided good results, the qualitatively best spectra (under similar conditions of exposure time and CCD quantum efficiency) are obtained using the aluminium coated substrates. It is believed that the reason for this is twofold. Firstly, the aluminium coating effectively blocks the underlying glass, resulting in the lowest background signal of all the substrates measured. Secondly, it is believed that the reflective substrates provide approximately four times more Raman scattered photons than transparent substrates; this results in the source laser passing through the sample twice, effectively doubling the source power, as well as reflecting all forward scattered photons back towards the MO, which would otherwise be lost for a transparent substrate. This effect is true for a transparent sample, such as an epithelial cell, but may not be valid for a thicker tissue sample that appears to be opaque. For specific source wavelengths, other substrates perform as well as Raman-grade calcium fluoride and aluminium. In the lower wavelength regions (473 nm, 532 nm, and 660 nm), IR polished calcium fluoride, fused silica, and glass produce good biological Raman spectra, as shown in Figure 7.4, Figure 7.5, and Figure 7.8. However, the background signals produced by these

substrates in the higher wavelength regions swamp the Raman peaks generated from the cells. On the other hand, zinc selenide performed poorly in the low wavelength region, but generally improved with increasing wavelength, and even proved to be the best substrate for measurements with a 785 nm laser, as seen in Figure 7.12 and Tables 7.1–7.3, thus making it a good choice of substrate for the red and NIR region.

In terms of source wavelength, there is a greater SNR in the lower wavelength regions, with spectra in the NIR being noisy for the integration time of 30 s used in this study, as shown in Figure 7.3 to Figure 7.12. However, this noise can be reduced by using a longer integration time, or using a CCD detector with a higher quantum efficiency in this region. Spectra in the NIR region have a higher resolution than those from lower wavelength lasers, but resolution for the lower laser wavelengths can easily be increased using a grating with higher line numbers than that used here. NIR lasers are also better suited for biological analysis due to the optical window in biological samples, but larger background signals, from the sample substrate and the optical elements in the system, are present in this region. Therefore, there must be a compromise when deciding upon a laser for Raman spectroscopic analysis of biological specimen; one must choose between a laser that results in lower background signals, but is more likely to cause biomolecular damage to the sample, or a laser that is biologically-friendly but which is more likely to produce large background signals.

Other factors must also be considered when choosing a sample substrate such as cost, reusability, and biocompatibility. Calcium fluoride, and in particular Raman-grade calcium fluoride, is the most expensive of all of the substrates in this study at approximately 150-200 Euro per slide, but it is biocompatible and produces good biological Raman spectra throughout the full range of laser wavelengths applied in this study. Potassium bromide and sodium chloride are a much more cost effective solution to calcium fluoride (with associated spectra shown in Figure 7.10 and Figure 7.11), however both of these substrates are soluble in water, making it more difficult to clean and re-use these substrates. Magnesium fluoride (Figure 7.9) is more expensive than potassium bromide and sodium chloride, yet it is cheaper than calcium fluoride, it has the ability to produce good Raman spectra throughout the range of source wavelengths, and it is biocompatible and reusable. Glass is the most widely available and cheapest substrate, but unfortunately it is not ideal for the recording of spectra in the higher wavelength region where the biological window exists. However, for studies in the lower source wavelength region, glass has the ability to produce results comparable to that from calcium fluoride, as shown in Tables 1–3. Similarly, IR grade calcium fluoride and fused silica work better in the lower wavelength region, as shown in Figure 7.4 and Figure 7.5, but they are more expensive than glass. Finally, zinc selenide is a significantly cheaper alternative to the other substrates for Raman micro-spectroscopy in the NIR region. However, it is not biocompatible, although there may be potential to functionalise this substrate with a 'Raman-friendly' biofunctional layer that would improve biocompatibility without compromising the spectral quality. (160; 161)

It should be noted that some of the substrates presented in this chapter can provide better cell

spectra for 785 nm and 830 nm despite the relatively poor results shown here in Figure 7.3 to Figure 7.12, particularly for Raman-grade calcium fluoride (Figure 7.3). For example, previous studies published by Stone et al. (142), Grimbergen et al. (145), and de Jong et al. (144) have shown good spectra for samples recorded at 830 nm/845 nm using calcium fluoride substrates, which could be achieved here by simply increasing the laser power, enlarging the confocal aperture, or increasing the exposure time. The results presented within this chapter may also be accounted for by considering the significantly lower number of scattered photons for higher laser wavelengths (see Section 3.4.1 on spectral resolution) and the quantum efficiency of the CCD detector used for recording spectra, for example, the use of a CCD detector with a higher quantum efficiency in the NIR region would improve the spectral quality in this region. However, the purpose of this study is to compare the different substrates under identical experimental conditions and to discover the relative performances of these parameters in a controlled experiment. It is accepted that the use of two different experimental systems reduces this level of control, but identical conditions across all experiments were applied insofar as possible.

7.6 Summary

Overall, it has been shown that there exists a high degree of variability across sample substrates and laser wavelengths for the analysis of biological specimen using Raman micro-spectroscopy. This high degree of variability means that there is not one single substrate and one source laser to suit all applications of Raman micro-spectroscopy in the analysis of biological specimen. This chapter discusses which substrates provide the optimum Raman spectra for each source wavelength region. However, to find the true optimum combination, it is necessary to consider the cost effectiveness and biocompatibility of the substrates, in combination with a source laser within the biological window to prevent photodegradation of the samples, whilst having a low background contribution.

In this study, a 'benchmark' spectrum is chosen, against which the quality of all other spectra were compared, to have come from a Raman-grade calcium fluoride substrate with a source wavelength of 532 nm. However, these substrates are very expensive and source lasers in this region are known to result in biomolecular damage within the sample. Therefore, with these considerations, it can be concluded that the use of aluminium coated glass substrates with an NIR laser provides the optimum Raman spectra for everyday clinical studies, being significantly cheaper than most substrates, biocompatible, and resulting in good quality spectra.

When integrating a new modality into a clinical environment for routine screening or diagnostic purposes, it is important to maintain compatibility with existing protocols if possible. With this consideration, the application of glass slides in conjunction with a 532 nm Raman system provides a good compromise in terms of spectral quality and cost effectiveness. Glass slides are cheap, readily available in the clinic, and they allow for biological samples to be stained and analysed by a cytopathology post-Raman acquisition if desired. Unfortunately, glass slides pro-

duce a relatively large background signal, as shown in Figure 7.8(a), which would need to be accurately removed in order to produce reliable and reproducible spectra; Chapter 8 investigates three different background correction algorithms for this purpose.

Chapter 8

Algorithm for the removal of the glass signal from Raman cytology spectra

This chapter has been published as part of the following paper:

- L.T. Kerr, B.M. Hennelly. *A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides.* Chemometr. Intell. Lab., 158, 61-68, 2016.

8.1   Introduction

Traditional preparation methods for cytology samples pose a significant problem for Raman micro-spectroscopy, with long-established clinical techniques depositing cells on glass slides. The results from the previous chapter showed the ability to combine glass slides with a 532 nm Raman system in order to produce cell spectra that are similar to those achieved with CaF_2 substrates. Furthermore, glass slides are inexpensive, biocompatible, and readily available within the cytopathology clinic. However, the signal from the glass slide can sometimes obscure the Raman spectrum. The intensity of the glass signal varies from cell to cell depending on morphology, and although smooth, the signal is more complex within the fingerprint region than the inherent baseline, and cannot easily be removed from the Raman spectrum using polynomial fitting techniques. It is difficult to accurately remove both the glass signal and the baseline signal simultaneously. Therefore, the use of standard glass slides compromises the capability of pre-processing methods to extract reliable and reproducible spectra from biological cells. To avoid this signal, Raman spectra are often recorded from cells on expensive substrates, such as CaF_2, aluminium, or quartz, as discussed in Chapter 7, but this practice is impractical for large scale applications of Raman cytology for diagnostics or screening purposes. This chapter investigates the potential of a number of background subtraction algorithms to remove both the glass signal and the baseline, and investigates the effect of these algorithms on subsequent multivariate analysis for the purpose of cell classification.

It is necessary to remove background signals insofar as possible from Raman spectra in order to facilitate an accurate comparison of related cell spectra, and in particular for the application of multivariate classification for the purpose of disease diagnostics or screening. In general, Raman spectra of biological samples contain a broad baseline signal that often varies randomly from one recording to the next, as discussed in Section 4.4. The signal is most often ascribed to an auto-fluorescence from the sample itself. (74) Although it has been suggested by some authors (75) that it may originate from sample morphology and Mie scattering of the source laser wavelength. Regardless of its origins, various algorithms have been developed to identify and remove this baseline signal from Raman cell spectra, with polynomial fitting techniques being the most common technique used today. (92; 166; 167)

The preparation of cytology samples poses a significant problem for Raman analysis, with current clinical techniques, such as the ThinPrep or SurePath methods, producing cell samples on glass slides for pathological evaluation. Glass is often a necessary consumable in clinical cytopathology due to its low cost. Unfortunately, the background signal from glass adds to the inherent broad baseline signal to further obscure the weak Raman cell spectrum, thus compromising the ability of Raman micro-spectroscopy to produce reliable and reproducible spectra from biological cells. (159) This is particularly evident in the 1050–1150 cm^{-1} region, where the glass signal is often strongest when recording Raman spectra with a 532 nm excitation source. The spectral profile, location, and intensity of the glass signal is dependent on the source wavelength, as previously shown in Figure 7.8, with the ability to recover Raman spectral peaks decreasing as the source wavelength moves from the visible to the NIR region. This study is only focused on 532 nm laser sources, which produce a relatively low, but still problematic glass background signal. In this study, a number of different background subtraction algorithms are investigated in order to accurately remove the glass contamination present in Raman cell spectra, as well as the baseline signal.

Algorithms have been previously developed to remove spectral contaminants from Raman spectra, based on a variety of techniques. Beier et al. (77) proposed an algorithm that simultaneously removed the baseline as well as the background signal from a known contaminant based on an iterative polynomial subtraction method. (73) This algorithm (with minor adjustments) has produced good results in the removal of the glass signal from Raman spectra of epithelial cheek cells, as previously shown in Chapter 7; however, following extensive testing, it is reported that for certain cell lines of particular morphology, this algorithm can result in over-fitting and alteration of key spectral information, which has a negative impact on resultant multivariate classification algorithms.

EMSC algorithms are gaining interest in recent times for the removal of spectral interferents from vibrational spectroscopic data. (80) EMSC can be applied to vibrational spectra to separate between different physical effects based on an OLS fitting approach (168), as discussed in Section 4.4.3. This technique has been applied extensively to FTIR spectroscopic data to correct for Mie scattering effects. (82; 83; 32) Liland et al. (79) recently applied EMSC to fit

whole datasets to reference spectra, resulting in the removal of an interfering signal from Raman spectra of adipose tissue. This interferent was due to an optical effect resulting from the Raman system design, presenting in various intensities from spectrum to spectrum, and could not be completely removed using traditional background correction algorithms, such as a modified polynomial in combination with the standard normal variate. A similar approach is applied here for the removal of the glass signal from Raman cytology spectra. It is believed that this algorithm could help with the advancement of Raman cytology into a clinical setting, allowing for the use of current clinical pathology techniques, such as glass slides. Additionally, the results of the EMSC algorithm, and subsequent PCA results, are compared with two other background algorithms.

8.2 The EMSC algorithm for removal of glass signal

The EMSC algorithm, as mathematically described in Section 4.4.3, can be applied for the removal of the glass signal from Raman cytology spectra. This algorithm can be implemented via an adapted version of Matlab's *polyfit* function, as shown below. This code uses the same notation as in Section 4.4.3. It is assumed that the inputs r and b have been smoothed using a Savitzky Golay filter (or another similar smoothing technique), and that X_0, r, and b are all 1D vectors (i.e. spectral irradiance values only, no wavenumber axis, in $[m \times 1]$ format).

```
function [X_final, background, c_r, c_b, B_N] = EMSC(X_0, r, b, N)

% Create dummy axis
axis = 1:length(X_0);
mu = [mean(axis); std(axis)];
axis = (axis - mu(1)) / mu(2);
axis = axis';

% Construct Vandermonde matrix
V(:,N+1) = ones(length(axis),1,class(axis));
for j = N:-1:1
    V(:,j) = axis.*V(:,j+1);
end

V(:,N+2) = b;
V(:,N+3) = r;

% Solve least squares problem
[Q,R] = qr(V,0);
```

```
p = R\(Q'*S);


% Calculate outputs
P_N = polyval(p(1:N+1), x);
c_b = p(N+2);
c_r = p(N+3);
background = P_N + (c_b*b);
X_final = (X_0 - background) / c_r;
```

The choice of the reference spectrum, r, is a subject of particular interest. It is common to set r to be equal to the mean spectrum for a given dataset of interest. (79; 80) In this study, however, in order to omit the glass signal from the reference spectrum, spectra are recorded from similar cells on CaF$_2$ slides, and r is taken to be equal to that mean spectrum. This substrate produces a relatively weak background signal, as discussed in Chapter 7, and it can therefore be assumed that $r \approx r_{cell} + P_{ref}$, where r_{cell} denotes the true Raman spectral irradiance of the cell on the CaF$_2$ substrate, and P_{ref} represents a baseline signal that is inherent in the reference spectrum. All corrected spectra therefore will be fit to a reference that includes this baseline signal. The presence of this constant baseline is only a matter of aesthetics since the qualitative and quantitative data within a dataset will be unaffected so long as all of the spectra in the dataset are processed using the same reference spectrum (80); it has been shown that this constant baseline, therefore, has no effect on multivariate statistical analysis such as PCA that follows after processing using EMSC.

The purpose of this study is to investigate the application of the EMSC algorithm to pre-process Raman datasets recorded from cells on glass slides, in advance of PCA based classification, with a view to understanding the potential of the method for improving the sensitivity and specificity of cytology. For such an application, it is important to fully evaluate the effect of different reference spectra on the multivariate classification of pre-processed data, in order to assess any possible biasing of the results.

8.3 Methods

8.3.1 Raman spectral acquisition

Spectra are recorded with the custom built confocal Raman micro-spectrometer, as described in Section 3.6.1. Spectra are recorded from each cell nucleus with an acquisition time of 5 s each; two spectra are recorded from the same location within the nucleus, and are averaged together using an algorithm that simultaneously removes cosmic rays. (72) Four datasets of Raman spectra are recorded using this system, with 50 cell spectra present in each dataset; (i) T24 high grade

bladder cancer cells recorded on CaF_2 substrates (Raman grade CaF_2, Crystran, UK); (ii) MDA-MB-231 triple negative breast cancer cells recorded on CaF_2 substrates; (iii) T24 bladder cancer cells recorded on glass slides; and (iv) RT112 low grade bladder cancer cells recorded on glass slides. The latter two datasets contain the spectra selected for pre-processing using the EMSC algorithm, while the other datasets are used to generate the reference spectra used in the algorithm. Further details on the growth and preparation of these cells are available in Section 9.2.1 and Section 11.2.1 later in this thesis.

8.3.2   Background subtraction algorithms

For the sake of comparison, two other well known background subtraction algorithms are first applied to the T24 and RT112 datasets recorded on glass slides.

The first algorithm is the modified polynomial baseline correction method, as proposed by Lieber et al. (73), and discussed in Section 4.4.1. This is an automated approach in which the spectrum is first fitted with an N order polynomial using OLS. Those values of the spectrum that lie above the polynomial are set equal to the value of the polynomial; the resultant signal is again fitted with an N order polynomial, and the process is iteratively repeated, until a point is reached such that the polynomial lies directly underneath the Raman spectral peaks. Here, this method is applied based on a fifth order polynomial and 200 iterations. This algorithm makes no attempt to remove the glass signal, but it is one of the most common approaches used for baseline correction in the literature.

The second algorithm applied is the exact method proposed by Beier et al. (77), similar to that discussed in Section 4.4.2. This method involves the subtraction of a weighted glass signal and an N order polynomial in an iterative manner, until a point is reached such that the modelled baseline lies directly below the Raman peaks. The $fminsearch$ function in Matlab (Matlab, Mathworks Inc., USA) can be applied to determine which glass weight/polynomial combination results in the smallest residual spectrum, as described in (77). In the results presented here, N is chosen to be 5, and the number of iterations is 200.

A glass signal is recorded by focusing the laser onto the surface of a glass slide over a 5 s integration time. The glass signal was then smoothed using a Savitzky Golay filter ($w = 3$, $k = 41$). In this study, only a first order polynomial is chosen for all cases (i.e. a straight line); however, in general, the order of the polynomial is dependent upon the dataset being analysed and the associated baseline intensity present.

The choice of reference spectrum is an important consideration as discussed in Section 8.2. EMSC has previously been shown to work well for Raman spectroscopy; in previous examples the mean spectrum of a dataset has been shown to be a good choice for the reference spectrum. However, in the present study, two datasets are analysed for the purpose of cell classification, therefore, it is essential to use the same reference for both bladder cell datasets on glass slides. The reason for this lies in the constant baseline that is inherent in the reference. Therefore, all

spectra must be fit to the same reference for the sake of fair comparison. In order to understand if the choice of a particular reference spectrum can introduce any bias into subsequent multivariate analysis, two different reference spectra are applied, and PCA analysis of the resultant dataset pairs are compared. In order to generate a reference spectrum, 50 cell spectra are recorded from cells on CaF_2, and averaged together to form a single spectrum. No pre-processing or baseline correction algorithms are applied to the reference spectrum, unless desired for aesthetic purposes. Here, the first reference spectrum is generated from T24 cells that are recorded on a CaF_2 substrate, and the second reference spectrum is taken from MDA-MB-231 cells that are also recorded on a CaF_2 substrate. These two references have been chosen because they are both from epithelial cells, but have different spectral shapes and intensities to each other. Furthermore, one of the reference spectra is related to one of the two bladder cell lines under investigation, while the other is unrelated to the two cell lines under investigation. In both cases, a Savitzky Golay smoothing filter ($w = 3$, $k = 7$) is applied to the mean CaF_2 spectra, to generate the final reference spectra for use in the EMSC algorithm.

PCA is applied to both datasets following each background subtraction method, and the first three PC coefficients are analysed for residual glass signals. Additionally, the standard deviation across an entire dataset is monitored and compared for each background algorithm.

8.4 Results

This section begins with the results of the two well known background subtraction algorithms discussed previously; the modified polynomial (73), and the Beier method consisting of a modified polynomial with the glass signal (77). Figure 8.1(a) shows the mean spectra for T24 and RT112 cells recorded on glass, averaged over 50 cell spectra in each group respectively, following the application of the modified polynomial baseline correction method and normalisation. The shaded region highlights the presence of varying levels of glass signal across both datasets. This varying signal is due to the different cell morphology of the two cell lines; the RT112 cells appear to contain relatively smaller nuclei sizes, and therefore their Raman spectra contain a larger proportion of the glass substrate signal. It is expected that the glass contribution will remain in the processed spectra since the modified polynomial method is designed only to subtract the baseline signal.

Figure 8.1(b) shows the PC score plot, and the first three PC coefficients/loadings obtained when the data in Figure 8.1(a) are subject to PCA. Here, it can be seen that the first PC has a significant glass contribution (see Figure 8.1(c)). It is interesting to note that the spectra are separating mainly along the first PC; the physical interpretation of this is that the cells are effectively separation according to differing morphology across the two cell lines, which is manifesting through the varying power of the glass signal component. Although interesting, this is not a desirable result. It cannot be expected that a similar morphological difference will occur in all cell classification applications. It is far more reliable in general to base classification

on the biochemical variation across the datasets.
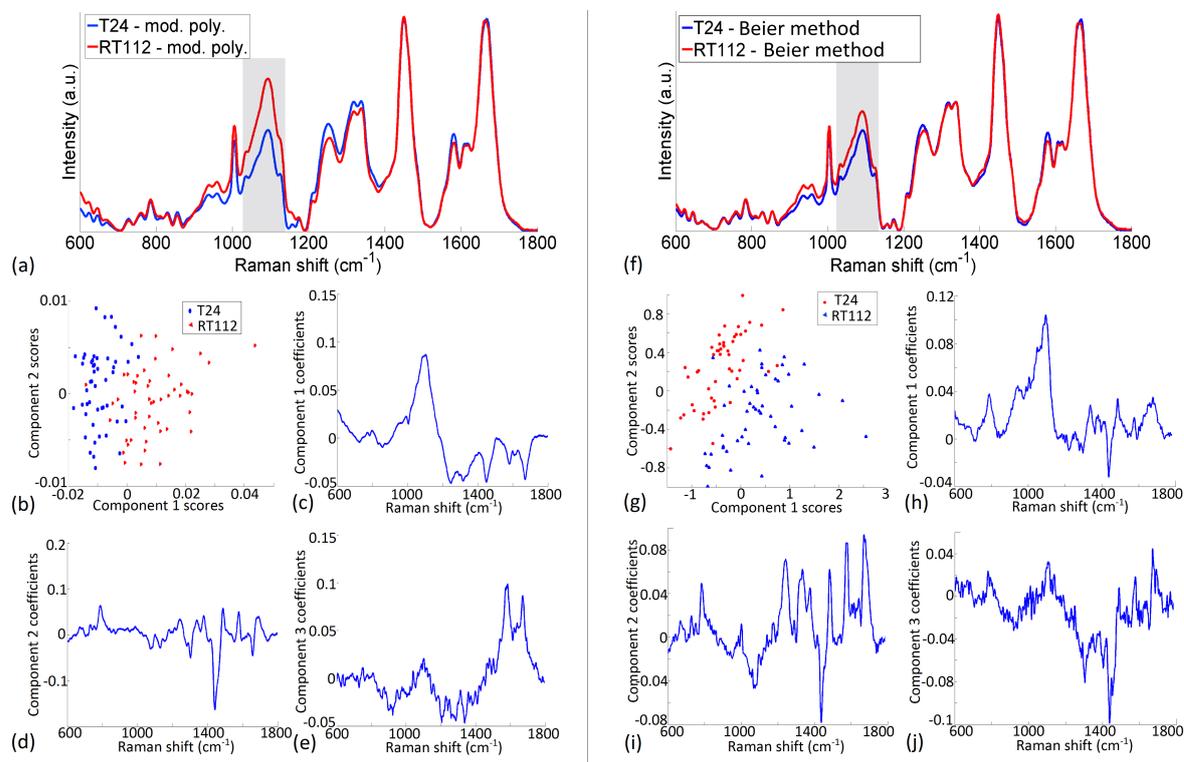
Figure 8.1: *(a) Mean of T24 and RT112 datasets following a modified polynomial background subtraction method; (b) PC scores, and the first three PC coefficients [(c), (d), and (e), respectively] for the data shown in (a); (f) Mean of T24 and RT112 datasets following the background subtraction method proposed by Beier et al., involving a modified polynomial and glass signal subtraction (77); (g) PC scores, and the first three PC coefficients [(h), (i), and (j), respectively] for the data shown in (f). The shaded areas highlight the region where the glass signal is most present within Raman cell spectra.*

The second algorithm that is investigated is the method proposed by Beier et al. (77) which, as described in Section 4.4.2, is designed to simultaneously remove both the glass signal and the baseline based on the combination of a modified polynomial and a weighted glass signal. Figure 8.1(f) displays the mean spectra for T24 and RT112 cells recorded on glass, averaged over 50 cell spectra in each group respectively, following application of this algorithm and normalisation. This figure demonstrates a significant reduction in the amount of glass signal present in the spectra, particularly for RT112 cells, when compared to spectra that have been baseline corrected with a modified polynomial alone.

Figure 8.1(g) shows the PC score plot, and the first three PC coefficients obtained when the processed data is subject to PCA. Similar to the results presented in Figure 8.1(c), the first PC contains a signal within the 1050–1150 cm^{-1} region, associated with glass (see Figure 8.1(h)). Therefore, whilst the glass signal has been reduced, it is still a significant component in both spectral datasets, and will remain the dominant factor in any PCA based classification.

Figure 8.2 illustrates the removal of the glass signal contribution from a sample T24 cell

spectrum recorded on a glass slide based on a reference spectrum generated from T24 cells on CaF$_2$ using EMSC. The red line represents the raw spectrum recorded from a T24 cell on glass, and the black line represents the reference spectrum to which all other spectra are fitted. The green line is the modelled background consisting of a combination of the glass signal and a first order polynomial, determined using the EMSC algorithm, and the blue line is the corrected spectrum, which has had the glass signal subtracted.
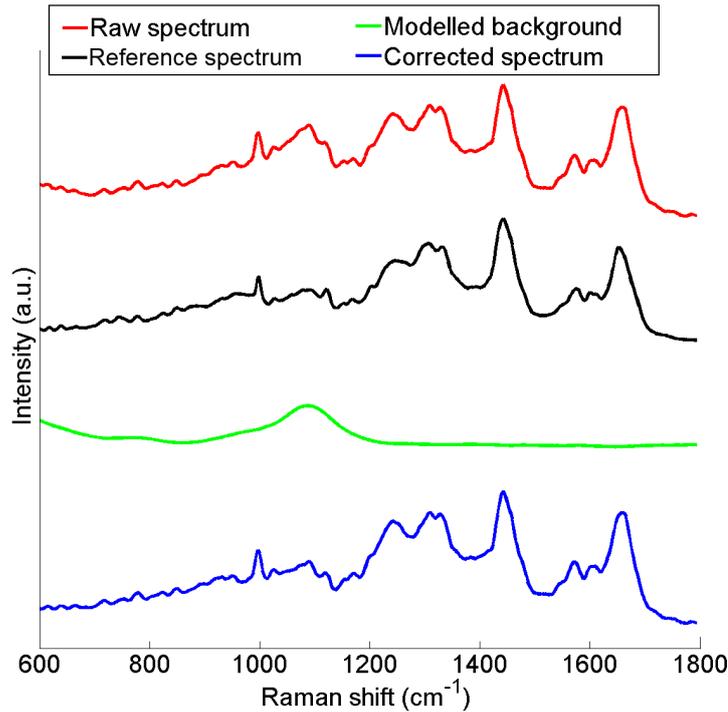
Figure 8.2: *Raw spectrum of T24 cell (red), recorded on a glass slide, which has been fit to a reference spectrum recorded from T24 cells on CaF$_2$ (black) using EMSC, resulting in a modelled background signal (green) consisting of a first order polynomial and a weighted glass signal, and the final background corrected spectrum (blue).*

Figure 8.3(a) shows the equivalent mean spectra for the same cell lines as shown in Figure 8.1, where pre-processing is implemented using the EMSC algorithm; input to this algorithm is the glass signal, a reference spectrum based on T24 cells on a CaF$_2$ substrate, as well as the chosen polynomial order N. Here, the glass signal has been effectively removed from the Raman cell spectra. The remaining peaks within the 1050–1150 cm^{-1} region are Raman cell peaks, seen in urothelial cell spectra recorded on CaF$_2$ and other similar substrates that produce low background signals. (143) The standard deviation is also shown for each cell line following EMSC processing. The amplitude of the standard deviation is amplified by a factor of 10 with respect to the mean spectra shown in the same figure.

Figure 8.3(b) displays the PC score plot obtained when these datasets are subject to PCA, with Figure 8.3(c), (d), and (e) illustrating the first three PC coefficients obtained. In contrast to the PC coefficients using the previously discussed processing methods, the PC coefficients
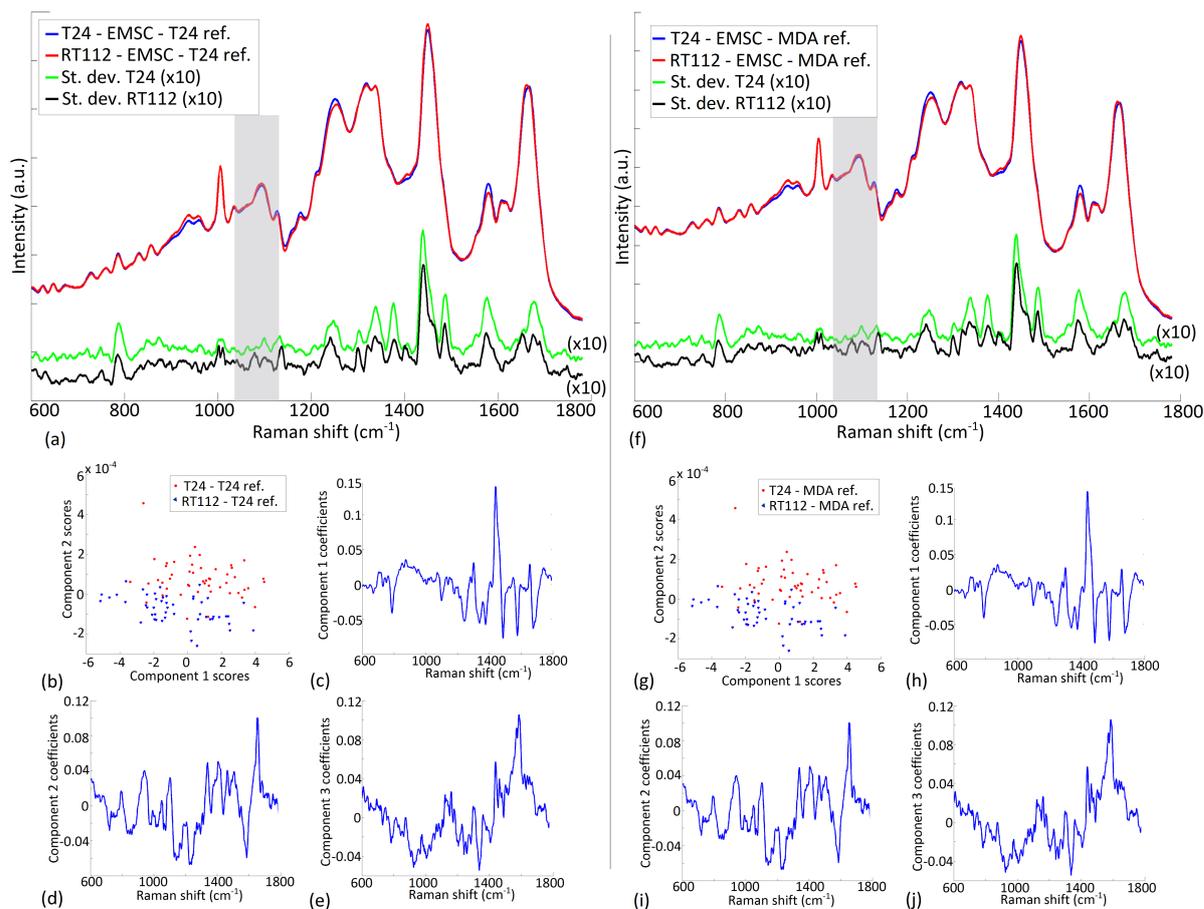
Figure 8.3: *(a) Mean of T24 and RT112 datasets following EMSC based on a T24 reference spectrum; (b) PC scores, and the first three PC coefficients [(c), (d), and (e), respectively] for the data shown in (a); (f) Mean of T24 and RT112 datasets following EMSC based on a MDA-MB-231 reference spectrum; (g) PC scores, and the first three PC coefficients [(h), (i), and (j), respectively] for the data shown in (f). The shaded areas highlight the region where the glass signal is strongest within Raman cell spectra.*

presented in Figure 8.3 do not appear to contain a glass signal component. This is an important result as it is essential for reliable classification algorithms to be based on biomolecular differences across cell groups, and not to be based on the presence, or absence, of a glass signal.

It is important to note that $N = 1$ was used for the EMSC algorithm; in this case, therefore, the EMSC algorithm ultimately amounts to subtracting only a straight line, and a weighted glass signal from each cell spectrum, followed by normalisation. The use of higher orders were also investigated, but the use of $N = 3$, 5, or 7 appeared to offer no improvement over the results presented here. For this reason, and since it may help to invalidate any suggestion of over-fitting, $N = 1$ is chosen. For some cases, which are not presented here, where strong baseline signals are present within a dataset, higher values of N are needed for accurate modelling of the background signal. It has been shown elsewhere that the use of high values of N (e.g. up to 7th order) with EMSC does not result in over-fitting. (80)

Although these results are positive, and it is clear that the glass signal is removed, and will

no longer be a factor in any subsequent classification applied to the PCA results, it could be argued that the EMSC algorithm might inadvertently influence the results of any subsequent multivariate analysis, particularly when the reference is based on the mean spectrum of T24 cells on CaF_2 for both the T24 and RT112 glass datasets. To investigate such effects, a second reference spectrum is used that is unrelated to the two cell lines under investigation. The reference used in this case is based on MDA-MB-231 cells on CaF_2. All other parameters (i.e. glass signal, N) remained the same. The mean spectra of the two processed datasets are shown in Figure 8.3(f), where it can be seen that the overall shape of these mean spectra are moderately different from the corresponding result shown in Figure 8.3(a) for the T24 reference spectrum. This results from the two references containing differing baselines. This difference is merely a question of aesthetics, and has no impact on any multivariate statistical analysis that is to follow EMSC processing. Indeed, it can be seen that the peak differences between the two spectra is effectively the same for both references. An analysis of the standard deviations of the two datasets also shows a very similar trend to that found for the previous T24 reference.

Figure 8.3(g) shows the PC score plot obtained when the two processed datasets are subject to PCA. Remarkably, the PCA score plot appears to be identical to that obtained for the previous reference. Figure 8.3(h), (i), and (j) illustrate the first three PC coefficients obtained. The PC coefficients also appear to be identical to those presented in Figure 8.3(c), (d), and (e), thus showing that the EMSC algorithm appears to produce results in the subsequent multivariate analysis that are independent of the reference spectrum used, and that it has no impact on the relative Raman peaks.
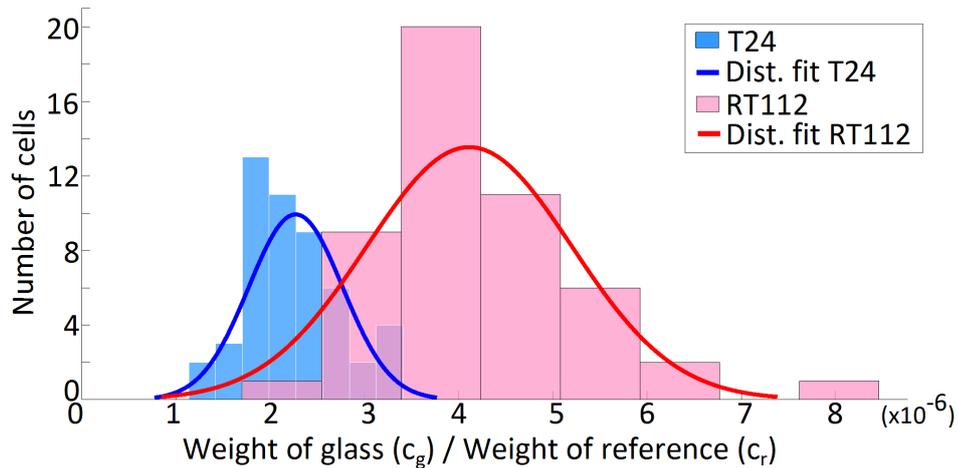
Figure 8.4: *Histograms, and associated distribution fits, of the weight of the glass signal (c_g) divided by the weight of the reference signal (c_r), obtained following EMSC with a T24 reference for T24 and RT112 cells.*

The ratio of c_g to c_r corresponds to the weight of the glass signal relative to the weight of the reference that is present in a given spectrum. These ratios are shown in Figure 8.4 for both the T24 and RT112 datasets following application of the EMSC algorithm with the T24 reference spectrum. By applying a Gaussian distribution fit to these values, it is possible to estimate both

the mean and standard deviation of the glass to reference ratio. It is clear that in general, the RT112 cells contain a stronger glass signal and a larger standard deviation, likely resulting from their smaller morphology. It is possible to choose a threshold ratio value, in between the two mean values, that largely separates the two datasets. By choosing the mid-point as a simple threshold, it is possible to achieve a sensitivity of 97% and specificity of 82% for T24 and RT112 cells. It is believed that this classification is based purely on the morphology of the cells being analysed, with the ratio of c_g to c_r being inversely proportional to the cell thickness. It is interesting to note that in the case of the two algorithms investigated in Figure 8.1, classification was possible using only the first PC, which was primarily composed of the glass signal, rather than the subtle variation in the Raman spectra owing to varying biochemical concentrations. In the case of EMSC, the glass signal can be extracted, and two approaches exist for classification: (i) analysing the ratio of c_g to c_r to perform an approximate classification, or (ii) analysing the Raman spectra after EMSC for a more accurate classification based on biomolecular variation.

In order to compare the three algorithms investigated here, the standard deviation and confidence intervals, across the various processed datasets, were analysed. Figure 8.5 shows the mean, and associated 95% confidence interval, for RT112 cells following each of the background correction algorithms. Following a modified polynomial correction, there remains a significant deviation across the fingerprint region, with the largest differences seen in the region where the glass signal is present. The Beier method reduces this confidence interval; however, there remains a considerable amount of variance seen across the 1050–1150 cm^{-1} region. The third method, based on EMSC, shows a reduced confidence interval across the entire spectrum, including the 1050–1150 cm^{-1} region. This indicates that the glass signal has been effectively removed from all of the spectra in the dataset.

8.5 Summary

For many years, the advancement of Raman micro-spectroscopy into the clinic has been impeded by its incompatibility with standard clinical protocols, particularly the use of inexpensive glass slides, as previously discussed in Chapter 7. In this chapter, the ability to remove the glass signal present within Raman spectra, as well as the baseline signal, has been demonstrated, resulting in spectra that are free from glass contamination. The EMSC algorithm takes as input (i) the signal generated from glass slides upon illumination with a 532 nm laser, (ii) a reference spectrum based on a similar cell type recorded on CaF$_2$ substrates, and (iii) a chosen baseline polynomial order N; based on these input parameters the algorithm estimates a background consisting of a weighted glass signal and a slowly varying baseline curve, which can be subtracted to produce spectra that are equivalent to those recorded on expensive "Raman-friendly" substrates. The output of the EMSC algorithm is the raw spectrum with this background subtracted, followed by multiplication with a normalisation factor that is related to the reference spectrum.
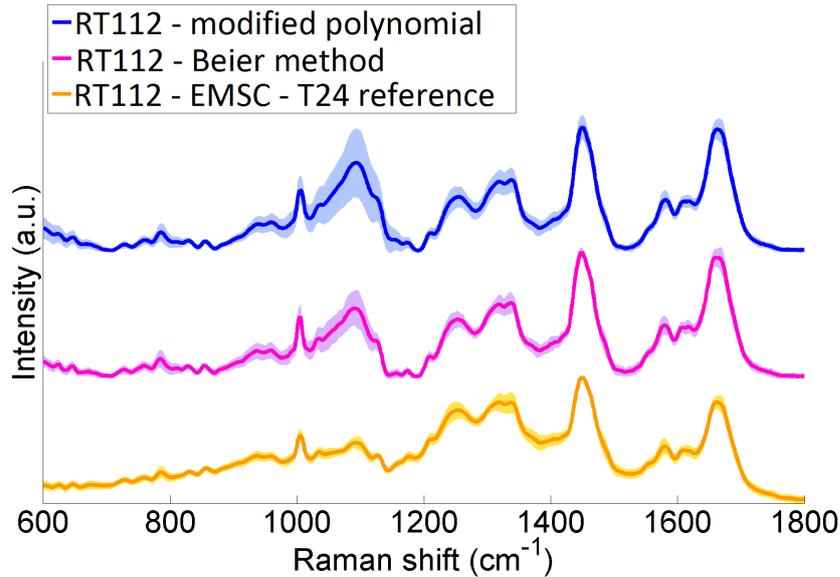
Figure 8.5: *Mean, and associated 95% confidence intervals, of RT112 cells, recorded on glass slides, following three different background algorithms: (i) modified polynomial, (ii) Beier method based on a modified polynomial plus glass signal, and (iii) EMSC algorithm, based on a reference spectrum of T24 cells on CaF_2.*

Additionally, application of PCA to the background corrected spectra indicates that EMSC produces reliable and reproducible results that are independent of the reference spectrum used. This is an important result since objectivity and reproducibility are crucial for providing good diagnostic classification, and it demonstrates that the reference spectrum does not introduce any biasing of PCA based classification. EMSC is also compared with two other well known background subtraction algorithms, for which it can be seen that the glass signal remained a significant component within their first PCs, thus reducing the reliability of these algorithms for Raman based cytological diagnostics on glass slides.

At present, it is not possible to provide a universal reference spectrum that can be applied to Raman datasets recorded on glass slides from any Raman micro-spectrometer. The reasoning for this is due to the lack of accurate system calibration protocols for Raman micro-spectrometers. Such rigorous calibration tools involve wavelength calibration, using a Neon source, intensity calibration, with a NIST calibrated white light source, followed by wavenumber calibration (61; 67; 66), as discussed in Section 4.2. In order to utilise a universal Raman reference spectrum, similar to the Matrigel spectrum applied for FTIR resonant Mie correction (83), every Raman micro-spectrometer would need to be calibrated with a similar calibration tool. Furthermore, even with the application of such calibration tools, variable results are still often recorded across different systems. (63) Therefore, it is advised for the reference spectrum to be recorded from any epithelial cell type on CaF_2 with the user's own Raman micro-spectrometer.

As demonstrated in Figure 8.4, the ratio of the weight of the glass signal to the weight of the reference can be applied as a single classification metric for the case of the two cell

lines investigated within this paper; using this approach, it is possible to separate low grade and high grade bladder cancer cell lines with 97% sensitivity and 82% specificity. This technique could be used to quickly identify large or abnormal cells on a slide. However, as these results are most likely based on cell thickness, it may not be possible to separate cell groups that are more similar in size. This is an interesting application for Raman micro-spectroscopy, but it should be noted that there are alternative modalities that can provide better information about cell morphology, such as digital holographic imaging (169), scanning near-field microscopy (SNOM) (170), scanning electron microscopy (SEM) (171), or LiPhos (living photonics microfluidic platform) (172). Raman micro-spectroscopy identifies the biomolecular differences between different cell groups, and by removing the glass signal from Raman spectra with EMSC, it is possible to classify cells based on biochemistry rather than cell morphology, which produces higher classification results and can be applied across all cell types. Therefore, Raman micro-spectroscopy remains the prefered modality for the identification of cancerous or diseased cells. However, it should be noted that it may be possible to include the c_g/c_r metric as an additional variable, together with the processed cell spectrum, for enhanced classification; the benefit, as well as the manner, of such an approach may be the subject of future work.

There are many further advantages of the EMSC algorithm when compared to commonly used baseline correction methods. It is computationally less intensive, which is an important factor when considering the high patient through-put present in cytopathological laboratories worldwide, and produces spectral datasets with significantly smaller standard deviation, which improves the reproducibility of results. It is hoped that this investigation will help with the advancement of Raman based cytology into a clinical setting, allowing for the use of current clinical techniques, such as the ThinPrep or SurePath methods, and glass slides. Further investigations into the ThinPrep method, and various other protocols employed in the cytopathology clinic, are discussed in Chapter 9.

Chapter 9

Protocols for Raman based urine cytology

This chapter has been published as part of the following paper:

9.1   Introduction

Urine cytology is often used as an adjunct to cystoscopy for the diagnosis of bladder cancer; however, it has a low sensitivity in detecting low grade bladder cancers. Chapter 6 reviewed some recent investigations into the application of Raman micro-spectroscopy for the detection of bladder cancer via urine cytology, and it has been demonstrated to significantly improve the diagnostic sensitivity of urine cytology for low grade bladder cancer under ideal experimental conditions (Section 6.5). This chapter attempts to move Raman micro-spectroscopy a step closer to the clinic by systematically examining the potential of this technology to classify between low and high grade bladder cancer cell lines under the stringent clinical conditions that can be expected in the standard pathology laboratory, in terms of consumables, protocols, and instrumentation. The use of glass slides, traditional fixing agents, lengthy exposure to urine, red blood cell lysing agents, as well as common cell deposition methods, are investigated in terms of their impact on the diagnostic potential of Raman based urine cytology. This study investigates urine samples that have been prepared with the ThinPrep UroCyte method, and analysed with micro-Raman spectroscopy, in order to determine if they could provide a useful alternative to cystoscopy for long term bladder cancer surveillance. Additionally, cytology slides are stained post-Raman acquisition, and analysed using a photo-microscope, to determine if it is possible to integrate Raman micro-spectroscopy into current clinical practices.

9.1.1 Bladder cancer surveillance

Bladder cancer, or UC, has the highest recurrence rate of any cancer, and it has been reported as the most expensive malignancy from diagnosis to death for health care systems, costing the NHS in the UK approximately £55 million a year. (2; 173; 174) As discussed in Section 6.2, urine cytology coupled with cystoscopic examination is the standard for bladder cancer diagnostics. However, cystoscopy is invasive, expensive, and may miss some flat lesions. Approximately 75% of patients present with superficial disease (Ta or T1), known as non-muscle-invasive bladder cancer (NMIBC).

Of particular concern for public health systems internationally is the approach to surveillance of patients previously diagnosed and treated for NMIBC. This particular subset of UC has a high rate of recurrence, and therefore, these patients remain on costly invasive cystoscopic surveillance programmes for the remainder of their lives. Previous research has been carried out into the usefulness of urinary cytology versus cystoscopy in the follow-up of NMIBC. (175; 176) To reduce the number of cystoscopy procedures for both healthcare systems and patients, urinary markers could be used to detect recurrent tumours before they become large and numerous. However, urinary cytology is limited by its low sensitivity for low grade tumours, and whilst urinary biomarkers are better at detecting these tumours, they are still unable to detect half of the low grade tumours that have been identified by cystoscopy. (175; 177) According to current knowledge, no urinary marker can replace cystoscopy during follow-up, or help to lower cystoscopic frequency in a routine fashion. (178)

9.1.2 Standard urine cytology

Urinary cytology is usually requested for patients with unexplained hematuria, irritative voiding symptoms, patients suspected of bladder cancer, and patients monitored for bladder tumour recurrence. The predominant cellular component of urine cytology is normal urothelial cells, which can vary greatly in numbers, sizes, and shapes. These cells may appear as mononuclear, and may be columnar, cuboidal, parabasal-like, or polyhydral. Other normal cellular components found in urine cytology include urothelial clusters, papillary fragments, squamous and glandular cells, renal tubular cells and casts, and prostatic and seminal vesicle cells. (179) Inflammation, infective pathogens, and blood components are also found across many samples. Therefore, the process of isolating and identifying urothelial tumour cells within a voided urine sample can be challenging, particularly for low grade UC.

Cytologically, high grade UC is relatively easy to diagnose with a high degree of sensitivity and specificity due to the presence of anaplastic cells. These cells have high nucleus to cytoplasm (N/C) ratios; the nuclei are often eccentric with large irregular nucleoli. On the other hand, low grade UC has a low diagnostic sensitivity and specificity; these tumours are cytologically characterised by increased cellularity, and the presence of an increased number of urothelial clusters that may or may not be papillary. The cells in these clusters have high N/C

ratios with nuclei bulging out of the cytoplasm. CIS is a flat lesion and is the precursor to most invasive urothelial cancers. Cytology plays an important role in the detection of urothelial CIS since these lesions may be multifocal and not visualised on cystoscopy. These cells are large with high N/C ratios. (179) Other types of urothelial malignancies that may be present include squamous carcinoma, adenocarcinoma, small cell carcinoma, lymphoma, or metastatic tumour cells from different neoplastic primaries.

The specificity of cytology is greater than 90% and the sensitivity is ∼80% for high grade UC; however, the sensitivity is 20–50% for low grade UC. (154) The main reason for this is due to the fact that low grade tumour cells have a similar cytomorphology to normal urothelial cells, and the observation of increased cellularity or papillary fragments, which are associated with UC, may instead be related to lithiasis, infection, or urinary tract instrumentation. The cohesive nature of low grade tumours may also result in less cells being shed into the urine. (173)

9.1.3   Experimental motivation

The objective of this study is to investigate if Raman micro-spectroscopy can successfully classify low and high grade cell lines under the stringent conditions imposed by a typical pathological laboratory. In order to achieve this, four experiments that systematically examine the performance of Raman based urine cytology for cell classification across a range of parameters are proposed.

A comparison between air dried, formalin fixed, and PreservCyt fixed UC cells is the basis of our first experiment. The impact of formalin fixation has previously been investigated with Raman spectroscopy (180), however, to the best of our knowledge, PreservCyt has never been previously compared with air dried and formalin fixed cells for Raman micro-spectroscopic analysis. This initial study allows for the identification of biomolecular differences that are introduced to the UC cell samples by these fixation methods, and provides information on the impact these changes have on the capacity to classify between low and high grade UC cell lines. In order to minimise the number of variables in this initial experiment, these cells were not exposed to urine.

An important consideration is the length of time that cells can remain within a urine solution before Raman based classification becomes unreliable. It has previously been shown that it is possible to distinguish between different cell types after exposure to urine using Raman micro-spectroscopy (151; 91; 55), as previously discussed in Section 6.5. However, it should be noted that these previous studies dealt with cells that were exposed to urine for relatively short periods of time. In order to analyse urine samples from a clinic, it is impractical to design a Raman study based around fresh, unfixed samples; this is due to the rapid deterioration of cells in urine, and the fact that recording all cells on a slide using Raman micro-spectroscopy is typically a time-consuming process. To overcome this problem, standard urine cytology involves the addition of a preserving agent into the urine collection vial, with cells being fixed again after the urine

solution has been decanted. (173; 158; 181; 182) However, the impact of these agents on the recorded spectra, and the capability of multivariate statistical algorithms to classify different cell types after fixing, must be considered. Therefore, the second experiment deals with the impact of urine on cells over 72 h following PreservCyt fixation, with 72 h representing the maximum expected time frame for the transfer of a urine sample from the clinic to the lab for analysis.

Urine cytology is usually performed with a liquid based processing technique such as SurePath or ThinPrep, as discussed in Section 5.5; the third investigation in this study involves preparing urine/cell solutions with the ThinPrep UroCyte method for Raman micro-spectroscopy. There are two subgroups within this experiment: (i) application of only PreservCyt, and (ii) application of both PreservCyt and CytoLyt; this enables a make direct comparisons of both methods, and the impact of CytoLyt on UC cells can be identified in isolation.

Hematuria is the most common symptom present in patients diagnosed with UC, occurring in approximately 90% of cases. (183) Whilst hematuria may be intermittent for patients, it is important to consider the impact of red blood cells in urine when analysing these samples using Raman micro-spectroscopy. Red blood cells are known to produce a large signal that can often swamp the weak Raman spectrum obtained from epithelial cells. Urine solutions that contain blood can be treated with red blood cell lysing agents such as the combination of H_2O_2, ethanol, and industrial methylated spirits as reported by Bonnier et al. (78), or CytoLyt, which is part of the standard ThinPrep cytology method. In the final experiment, the impact of both scant and frank hematuria on Raman spectra of UC cell lines is investigated. Samples are processed using the ThinPrep UroCyte method, in conjunction with CytoLyt in order to lyse any red blood cells that may be present.

9.2 Methods

9.2.1 Sample preparation

Cell culture

* Note: the work outlined in the following cell culture subsection was completed in collaboration with Dr. Therese Lynn (Institute of Immunology, Maynooth University, Ireland).

Bladder cell lines T24 (high grade UC) and RT112 (low grade UC) are cultured in 1:1 mixture of DMEM and Hams-F12 medium supplemented with 5% fetal bovine serum and 2 mM L-Glutamine. Flasks are maintained in a humidified atmosphere with 5% CO_2 at 37°C. When the cell lines reach between 80–100% confluency, the culture medium is removed, and the cells are rinsed with sterile PBS. Trypsin-EDTA (0.5%) is added to the flask, which is incubated at 37°C until the cells completely detach (not exceeding 15 min). An equal volume of 5% serum-containing medium is added to the flask to neutralise the trypsin enzyme. The contents of the

flask are transferred into a sterilin container and centrifuged (1200 rpm, 5 min). The supernatant is removed, and the cell pellet is resuspended in fresh medium. This solution is centrifuged, the medium decanted, and the cells are resuspended in 1 ml PBS. This step is repeated, and the cell pellets are processed as explained in the following subsections.

Fixing agents

In order to compare between air dried, formalin fixed, and PreservCyt fixed cells, the samples are prepared as follows:

(A) Air dried: the cell pellet is suspended in 2 ml PBS, followed by centrifugation. The PBS is decanted, and the cell pellet is resuspended in 1 ml PBS. 30 μl of this cell suspension is dropped onto a glass substrate and left to air dry at room temperature for 2-3 hours.

(B) Formalin fixation: 10% neutral buffered formalin (HT501128, Sigma Aldrich, US) is passed through a 0.2 μm filter (Minisart filters, Sigma Aldrich, US) in order to remove any large salt crystals present in the solution. The cell pellet is suspended in 2 ml of formalin and left for 10 min at room temperature. The solution is centrifuged, decanted, and the cells are resuspended in 2 ml PBS. The resulting solution is again centrifuged, decanted, and the cells are resuspended in 1 ml PBS. 30 μl of this cell suspension is dropped onto a glass substrate and left to air dry at room temperature for 2-3 hours.

(C) PreservCyt fixation: the cell pellet is suspended in 20 ml of PreservCyt and left for 15 min at room temperature. The solution is centrifuged, decanted, and the cells are resuspended in 2 ml PBS. This solution is centrifuged, decanted, and the cells are resuspended in 1 ml PBS. 30 μl of this cell suspension is dropped on to a glass substrate, and left to air dry at room temperature for 2-3 hours.

Urine exposure times

In order to prepare urine samples containing UC cells, the cell pellet is suspended in 10 ml artificial urine (AU-001, Biopanda Diagnostics, UK) and 5 ml PreservCyt. These samples are stored in the fridge for 5 h, 24 h, and 72 h, respectively. Samples are then centrifuged, decanted, and the cell pellet is resuspended in 20 ml PreservCyt and left at room temperature for 15 min. The resulting solution is centrifuged, decanted, and the cell pellet is resuspended in 1 ml PBS. A final centrifugation step is carried out, followed by decantation, and resuspension in 50 μl PBS. This entire volume is then dropped onto a glass substrate and left to air dry at room temperature for 2-3 hours.

ThinPrep UroCyte

In order to prepare urine samples containing UC cells with the ThinPrep UroCyte method, the cell pellet is suspended in 10 ml artificial urine and 5 ml PreservCyt, and stored in a fridge for 24 h. Samples are then centrifuged, and the supernatant decanted. In order to compare the

differences between cells exposed only to PreservCyt, and those exposed to both PreservCyt and CytoLyt, the samples are prepared as follows:

(A) Two or three drops of the cell pellet are suspended in a 20 ml PreservCyt vial and left at room temperature for 15 min. The vial is inserted into the T2 and the cells are transferred onto a ThinPrep glass slide.

(B) The cell pellet is vortexed and resuspended in 30 ml CytoLyt. The solution is centrifuged and decanted, and two or three drops of the cell pellet are resuspended in a 20 ml PreservCyt vial and left at room temperature for 15 min. The vial is inserted in the T2 system and the cells are transferred onto a ThinPrep glass slide.

Hematuria

In order to compare the impact of scant hematuria and frank hematuria present in urine samples, UC cells are suspended in 10 ml artificial urine, 5 ml PreservCyt, and either 0.5 ml whole blood (creating a pink solution to represent scant hematuria), or 3 ml whole blood (producing a deep red solution representing frank hematuria). All samples are centrifuged and decanted. The cell pellet is vortexed and resuspended in 30 ml CytoLyt. The resulting solution is centrifuged, decanted, and two or three drops of the cell pellet are resuspended in a 20 ml PreservCyt vial and left at room temperature for 15 min. The vial is inserted in the T2 system, and the cells are transferred onto a ThinPrep glass slide.

9.2.2 Raman spectral acquisition

Our custom-built Raman micro-spectroscopy system is employed for all measurements in this study, as discussed in Section 3.6.1. Spectra are recorded within the 600–1800 cm^{-1} range with an acquisition time of 5 s each. Two spectra are recorded from the same location within the nucleus of 50 cells from each slide. The recording of two spectra facilitates the removal of cosmic rays as described by James et al. (72) Following this, the EMSC algorithm is applied to remove the glass signal and the slowly varying baseline from each Raman spectrum (as described in Chapter 8). For this study, a 1st order polynomial is used in the EMSC subtraction algorithm for all datasets, except for the case of T24 and RT112 formalin fixed cells on glass, which had a particularly strong baseline signal due to experimental parameters on that given day; in this case N was chosen to be 7. Additionally, all spectra are smoothed with a Savitzky-Golay filter ($k = 5$; $w = 7$). In order to compare the spectra recorded across the various parameters discussed in this study, and in order to determine the impact of these parameters on the classification of both cell lines, a combination of PCA and LDA is applied to the two groups in each of the four experiments. The sensitivity and specificity values are determined based on a leave-one-out cross validation method.

9.2.3 Haematoxylin & Eosin (H&E) staining

Following Raman spectral acquisition, all samples prepared with the ThinPrep UroCyte method are H&E stained. (184) This is achieved by placing the slides into 95% alcohol for 10 min, and then washing with water. Slides are placed into Harris haematoxylin (Sigma Aldrich, US) for 5 min, and then transferred into water, before being submerged into 1% acid-alcohol for 1 s, washed with water, and placed into water for 5 min to 'blue' the haematoxylin. Slides are placed into 1% eosin (Sigma Aldrich, US) for 3 min in order to counterstain the cytoplasm, following which, the slides are submerged in water for 1 s, 95% alcohol for 1 s, and 100% alcohol for 1 s. The slides are placed into 100% alcohol for 5 min, and are then transferred into a bath of xylene (Sigma Aldrich, US) for 3 min, and a second bath of xylene for a further 3 min. Finally, the slides are removed and a coverslip applied with DPX mounting medium (Thermo Fisher Scientific, Ireland), and are left to dry. Following staining, cells are examined under a photo-microscope, with a 40x MO, for changes relating to different stages of cancer such as abnormal cellular shape and size, abnormal nuclei, and cellular arrangement/distribution. The amount of residual blood cells present on the slide from hematuria contaminated samples is also monitored. Alternatively, slides could have been stained with Papanicolaou (Pap) if this method was prefered.

9.3 Results

9.3.1 Fixing agents

The spectra shown in Figure 9.1 represent the mean Raman spectra recorded from T24 UC cells for the cases of air dried, formalin fixed, and PreservCyt fixed, respectively. Raman spectra of formalin fixed cells are found to be most similar to air dried cell spectra with no significant differences observed. PreservCyt fixed cells are found to have differences in the intensity of some peaks within the fingerprint region; an increase can be seen in the peaks at 1093 cm^{-1} and 1250 cm^{-1}, which relate to PO$_2$ stretching (DNA/RNA), and a decrease in peaks at 750 cm^{-1}, 1310 cm^{-1}, 1340 cm^{-1}, and 1580 cm^{-1}, corresponding to tryptophan, guanine, adenine, and bending modes of phenylalanine. (185) Each individual T24 and RT112 spectrum is shown in Figure 9.2, corresponding to all three fixation methods.

For each fixation method, a PC-LDA training model is implemented in order to classify T24 and RT112 UC cells, with associated PCA scores and coefficients shown in Figure 9.3. The resulting sensitivities and specificities are 100% and 98.0% for air dried cells, 98.0% and 100% for formalin fixed cells, and 100% and 100% for PreservCyt fixed cells, respectively. These classification values indicate that the three fixing agents measured here have no significant impact on the ability of PC-LDA to classify spectra across both cell types. This demonstrates that PreservCyt fixation of cells, a common preservation method used across clinics today, can be integrated into Raman micro-spectroscopy. Interestingly, as shown in Figure 9.3, the PC
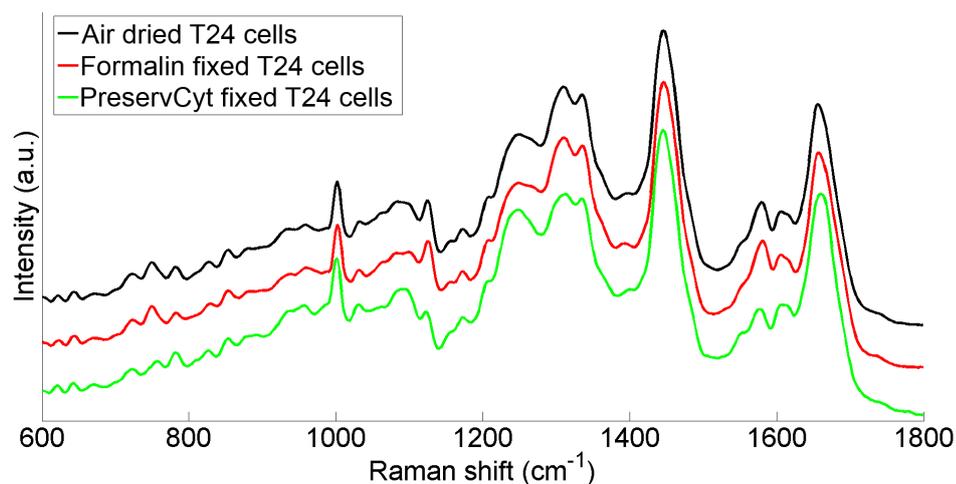
Figure 9.1: *Mean spectrum of air dried T24 cells (black), formalin fixed T24 cells (red), and PreservCyt fixed T24 cells (green). All cells were drop-dry deposited onto glass slides.*
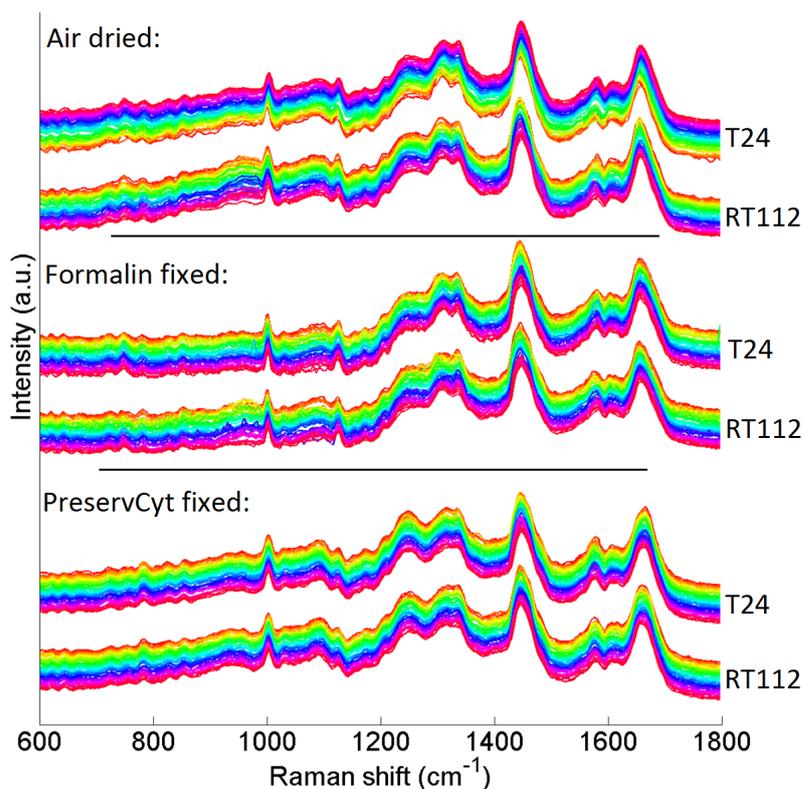
Figure 9.2: *50 spectra from each of the following datasets: (i) air dried T24 and RT112 cells; (ii) formalin fixed T24 and RT112; and (iii) PreservCyt fixed T24 and RT112, corresponding to the mean spectra shown in Figure 9.1.*
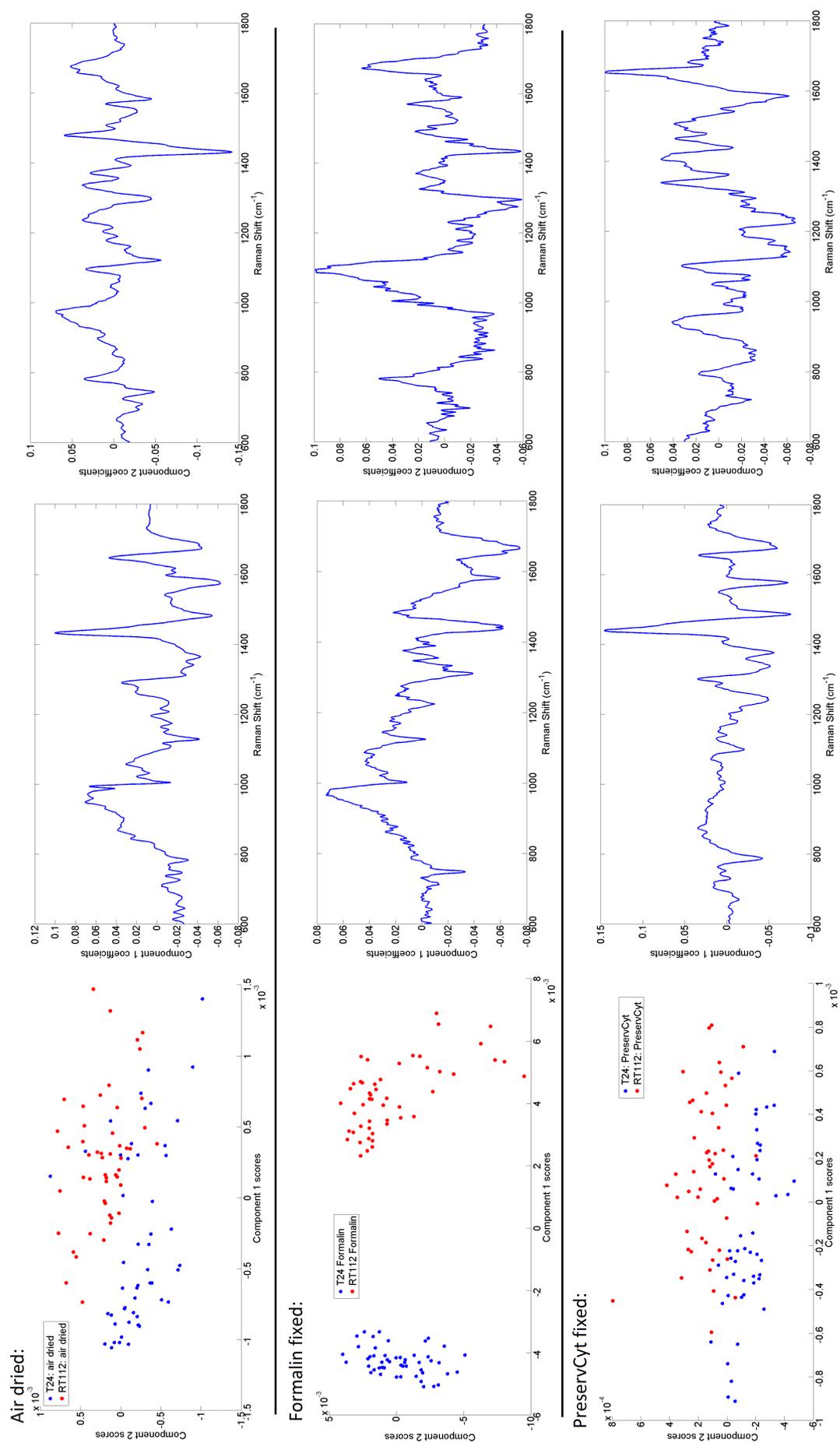
Figure 9.3: *PCA scores and the first two PC loadings shown for (i) air dried T24 and RT112 cells; (ii) formalin fixed T24 and RT112 cells; and (iii) PreservCyt fixed T24 and RT112 cells.*

scores for PreservCyt cells are most similar to air dried cells; this shows that PreservCyt fixation doesn't alter the spectral differences that are seen between T24 and RT112 after air drying.

9.3.2 Urine exposure times

The mean spectra for T24 and RT112 cell lines following urine exposure for 5 h, 24 h, and 72 h, with PreservCyt fixation, are shown in Figure 9.4, with individual cell spectra for each group presented in Figure 9.5.
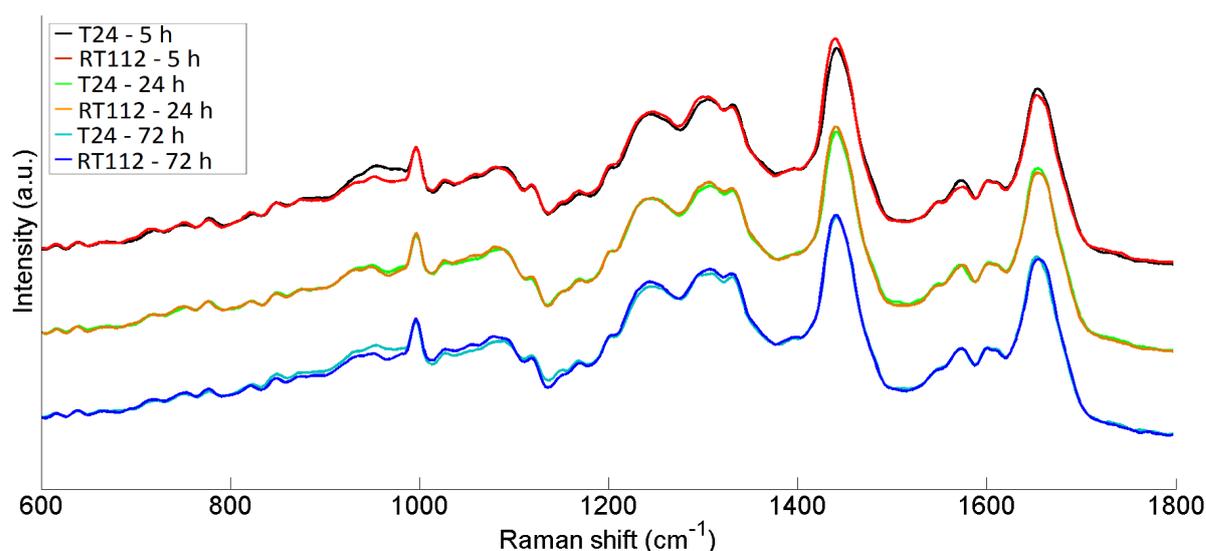
Figure 9.4: *Mean spectrum of T24 and RT112 UC cells after exposure to urine and PreservCyt for 5 h, 24 h, and 72 h; all cells were drop-dry deposited onto glass slides.*

For each exposure time, a PC-LDA training model is implemented, and the corresponding sensitivities and specificities are listed in Table 9.1, where it can be seen that it is still possible to accurately differentiate between both cell lines after 72 h; corresponding PCA scores and coefficients are shown in Figure 9.6. Interestingly, when the data from all three time durations are combined, the classification accuracies remain above 90%, which indicates that PreservCyt can preserve cells in urine for up to 3 days without significantly altering the cell biochemistry or the diagnostic potential of the method.
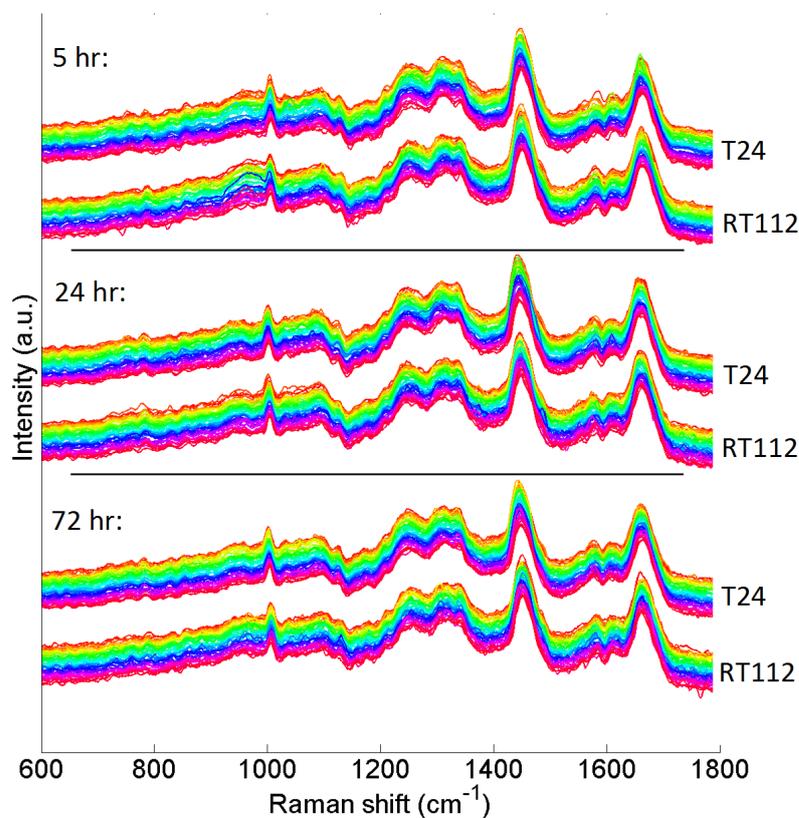
Figure 9.5: *50 T24 and RT112 spectra for each of the following time durations of urine exposure: (i) 5 hours; (ii) 24 hours; and (iii) 72 hours.*

Drop-dry method:	Sens.	Spec.
T24 - 5 h urine exposure	98%	94%
RT112 - 5 h urine exposure	94%	98%
T24 - 24 h urine exposure	98%	96%
RT112 - 24 h urine exposure	96%	98%
T24 - 72 h urine exposure	100%	100%
RT112 - 72 h urine exposure	100%	100%
T24 - Combined urine exposure times	92%	93%
RT112 - Combined urine exposure times	93%	92%

Table 9.1: *Sensitivity and specificity values for T24 and RT112 cell lines after urine exposure based on a PC-LDA model with leave-one-out cross validation (all cells were drop-dry deposited).*
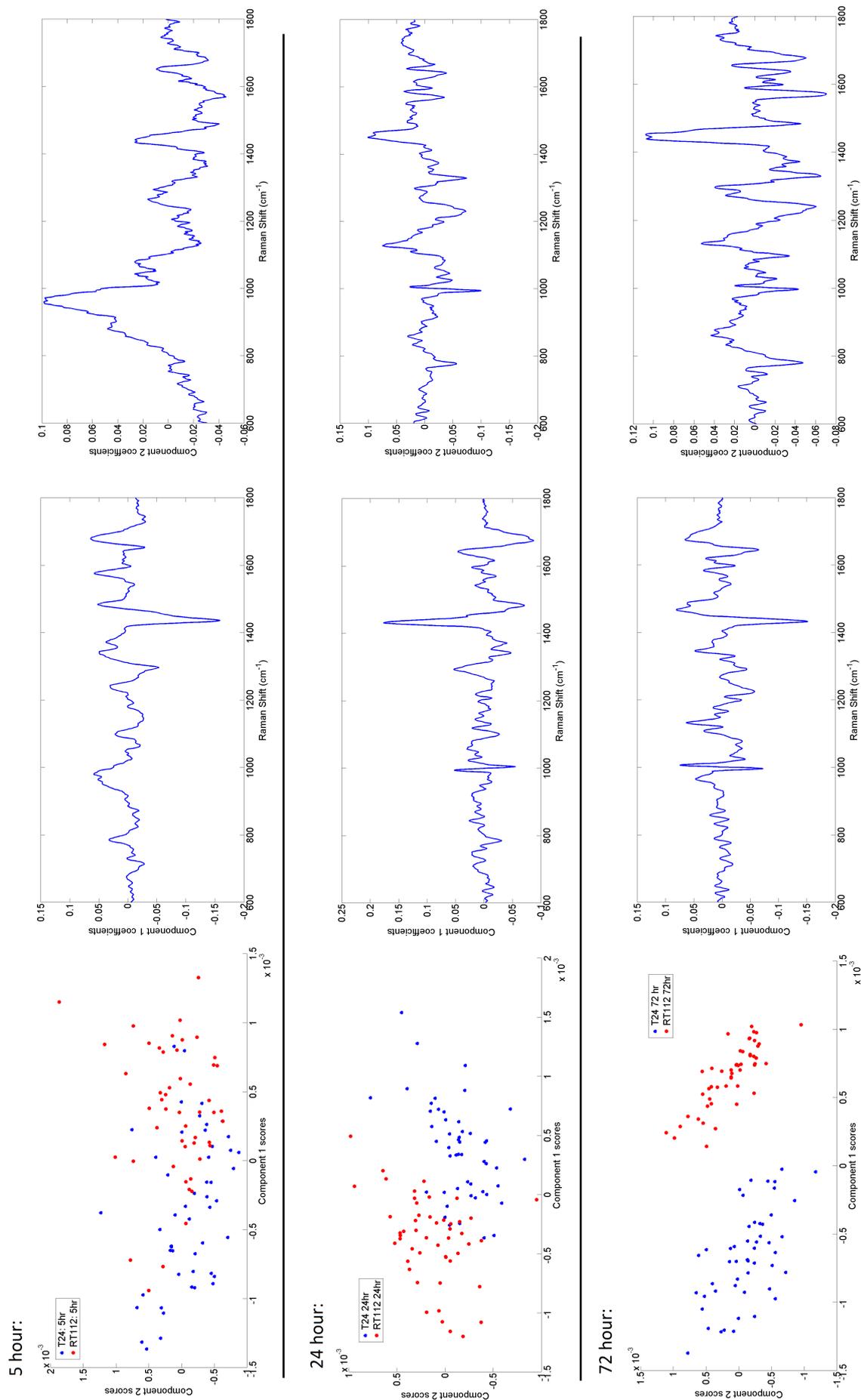
Figure 9.6: *PCA scores and the first two PC loadings shown for T24 and RT112 cells exposed to urine for the following time durations: (i) 5 hours; (ii) 24 hours; and (iii) 72 hours.*

9.3.3 ThinPrep UroCyte

Figure 9.7 shows the mean spectra recorded from T24 and RT112 UC cells after exposure to urine and PreservCyt for 24 h, with and without the addition of the red blood cell lysing agent CytoLyt, followed by ThinPrep processing. Here, it can be seen that CytoLyt has no significant impact on UC cells, and no spectral contributions are observed from cells that are exposed to CytoLyt compared with those exposed to PreservCyt alone. The spectra recorded from ThinPrep processed slides are more consistent than those measured with the drop-dry method, i.e. a smaller standard deviation is measured across the spectra from both T24 and RT112, as shown in Figure 9.8. It is possible that this is due to the uniform monolayer deposition method employed by the T2 machine compared to the random distribution of cells that can accumulate on the slide after the drop-dry method.
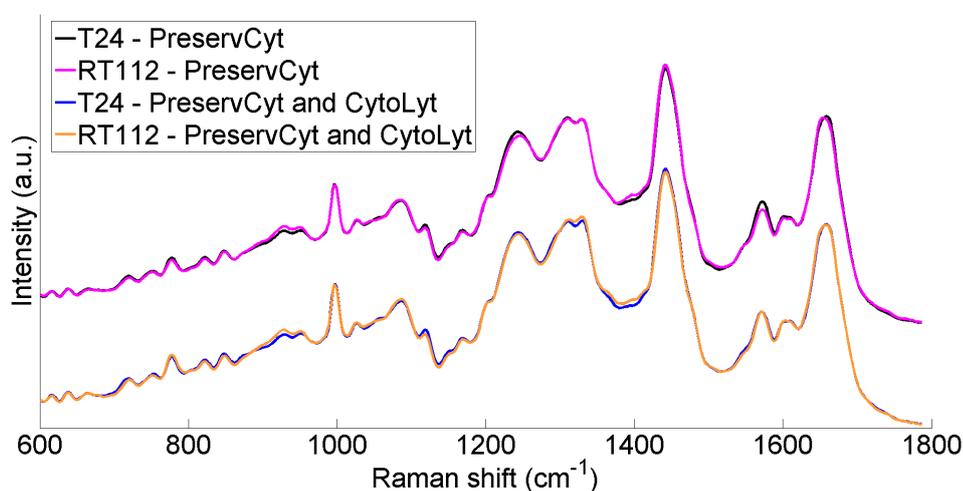
Figure 9.7: *Mean spectra of T24 and RT112 UC cells after exposure to urine and PreservCyt for 24 h, followed by ThinPrep processing with and without the addition of the red blood cell lysing agent CytoLyt.*

Figure 9.9 shows the images obtained from these slides after H&E staining (after Raman spectral acquisition); distinct differences can be seen between both cell lines in terms of cellular shape and size, along with more abnormal nuclei present in T24 UC cells, as expected. The cells do not appear to have been photodamaged by the laser, which indicates that a cytopathologist should still be able to perform standard urine cytology diagnostics on slides that have been stained following Raman spectral acquisition, thus allowing for both Raman and standard urine cytology to be performed on the same slide.

A PC-LDA model is generated based on a leave-one-out cross validation method for both of these methods individually and combined, and the results are displayed in Table 9.2, with PCA scores and loadings available in Figure 9.10. The sensitivities and specificities found are greater than 88%.
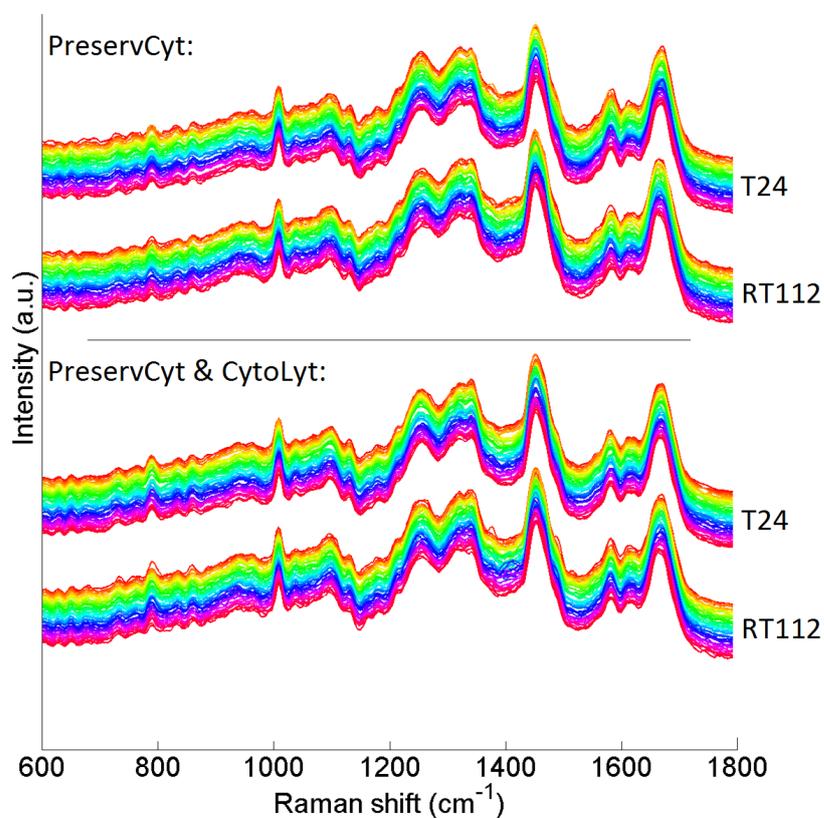
Figure 9.8: *T24 and RT112 cell spectra recorded following cell deposition using the ThinPrep UroCyte method, with (i) only PreservCyt, or (ii) with PreservCyt and CytoLyt.*

ThinPrep method:	Sens.	Spec.
T24 - 24 h urine - PreservCyt only	94%	100%
RT112 - 24 h urine - PreservCyt only	100%	94%
T24 - 24 h urine - PreservCyt and CytoLyt	96%	96%
RT112 - 24 h urine - PreservCyt and CytoLyt	96%	96%
T24 - Both ThinPrep samples	88%	99%
RT112 - Both ThinPrep samples	99%	88%

Table 9.2: *Sensitivity and specificity values for T24 and RT112 cell lines after urine exposure based on a PC-LDA model with leave-one-out cross validation, following ThinPrep deposition.*
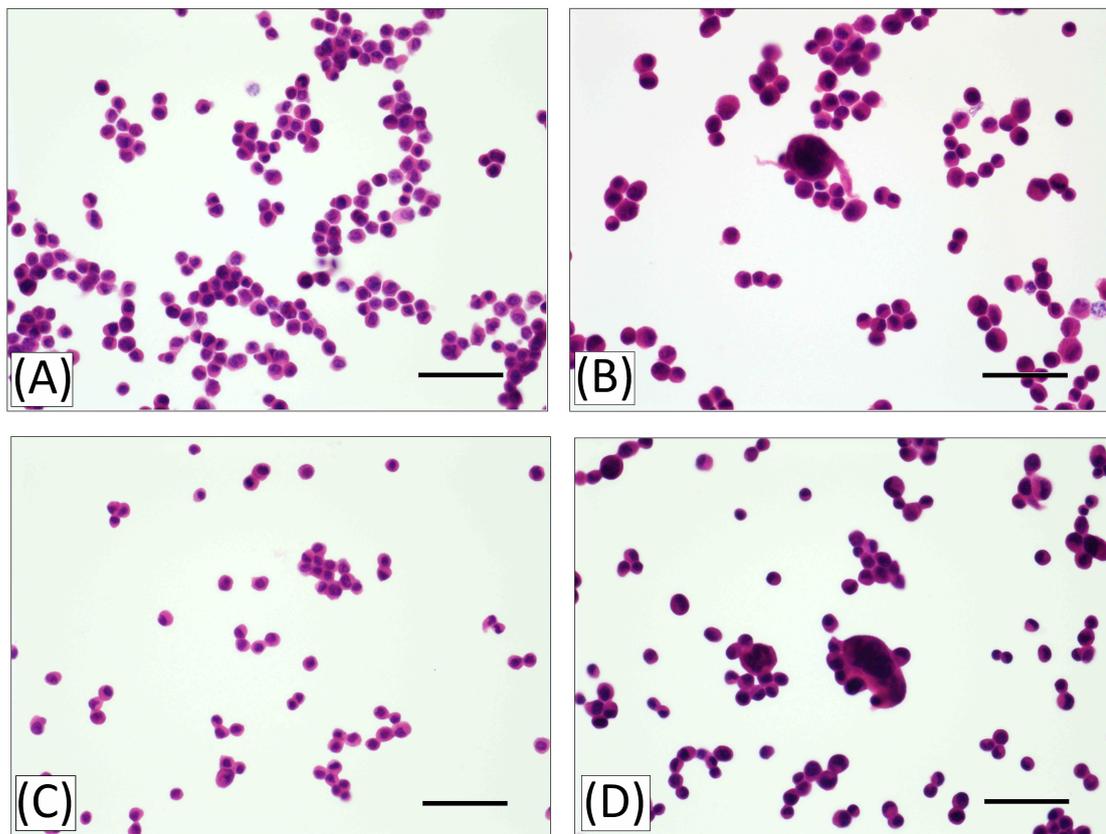
Figure 9.9: *Images obtained from H&E stained UC cells after Raman spectral analysis, prepared with the ThinPrep UroCyte method. (A) RT112 cells, and (B) T24 cells without CytoLyt; (C) RT112, and (D) T24 cells with CytoLyt. Scale bar = 100 µm.*
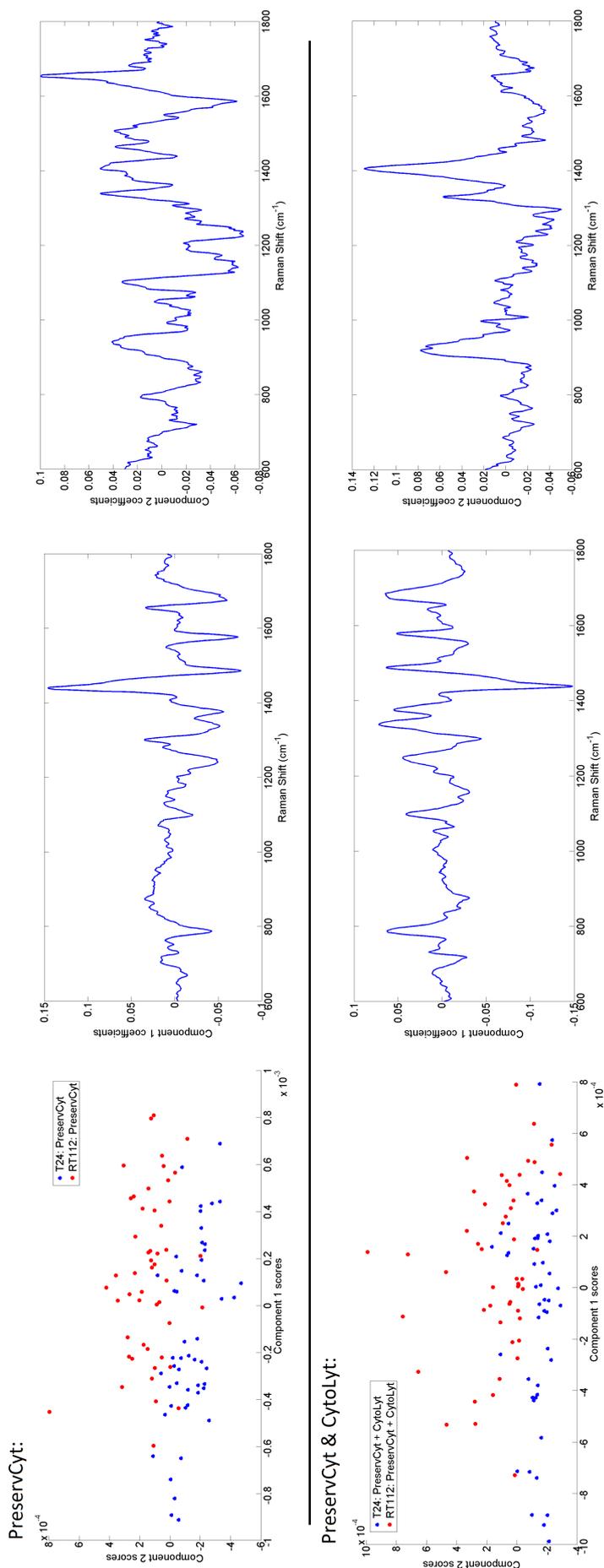
Figure 9.10: *PCA scores and the first two PC loadings shown for T24 and RT112 cells following cell deposition using the ThinPrep UroCyte method, with (i) only PreservCyt, or (ii) with PreservCyt and CytoLyt.*

9.3.4 Hematuria

Spectra recorded from cells obtained from ThinPrep processed urine samples replicating scant and frank hematuria are shown in Figure 9.11. Here, it can be seen that cells obtained from a urine sample with scant hematuria appear to be free from any contamination; this is due to the effective removal of many of the red blood cells present in the sample using CytoLyt. However, for the frank hematuria sample, several additional peaks have been observed that correspond to the presence of residual blood cells, a further breakdown of the peaks associated with blood cells can be found elsewhere. (186) This result indicates that samples that have relatively high blood cell concentrations are not suitable for diagnostics with Raman micro-spectroscopy, although further investigation into additional CytoLyt washes, or the use of H_2O_2 as suggested by Bonnier et al. (78), could possibly help to further remove residual blood from these samples.

Figure 9.12 shows the H&E stained images obtained from both samples containing scant and frank hematuria, respectively. Here, small regions of blood remain visible within the scant hematuria sample, but the majority of UC cells in this sample appear to be isolated from the blood regions. The frank hematuria sample, however, has a low UC cell yield, and contains large regions of blood contamination. It is possible that this is due to large amounts of blood cells gathering on the ThinPrep filter, resulting in less space for the UC cells. This demonstrates the importance of removing blood cells with a lysing agent, as well as applying additional washes, to further remove this contamination.
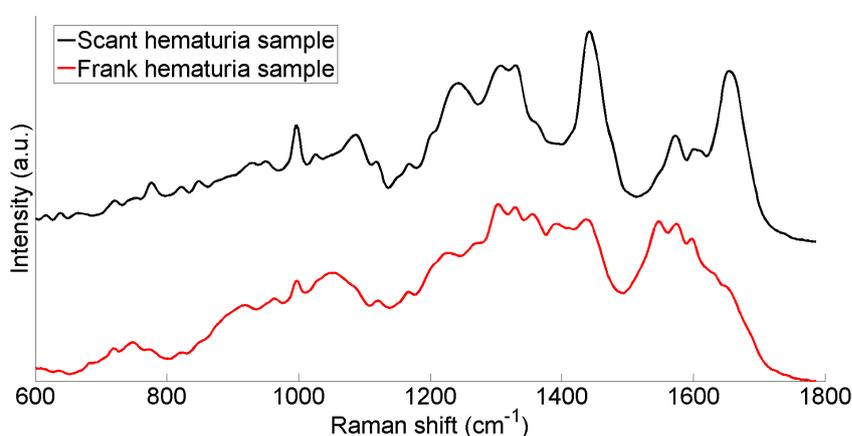
Figure 9.11: *Mean spectra of T24 and RT112 UC cells after exposure to urine and PreservCyt for 24 h with the addition of blood to replicate scant and frank hematuria. All samples were ThinPrep processed with the red blood cell lysing agent CytoLyt.*

9.4 Discussion of results

In this chapter the ability of Raman micro-spectroscopy to discriminate between low and high grade bladder cancer cell lines with a high degree of accuracy has been demonstrated, even within the practical constraints of a typical pathology laboratory. Four separate experiments
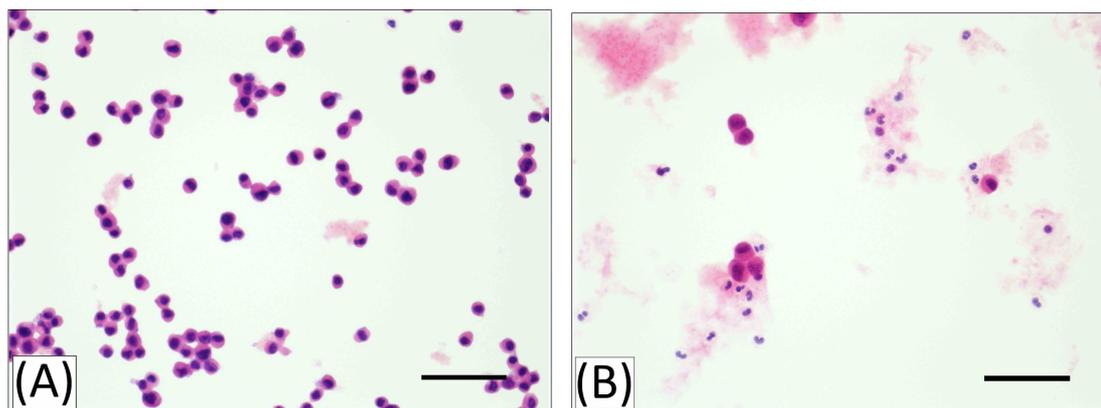
Figure 9.12: *H&E stained images of (A) RT112 cells from a scant hematuria urine sample, and (B) T24 cells from a frank hematuria urine sample. Both samples were processed with CytoLyt and the ThinPrep method. Scale bar = 100 μm.*

are conducted that validate the performance of Raman based classification when applied as an adjunct to the standard practice of urine cytology.

In the first experiment, it is concluded that the Raman spectra obtained from cells following fixation, using the commercial methanol based fixative, PreservCyt, are not significantly different to those recorded from air dried cells and formalin fixed cells, with negligible differences in statistical classification accuracies observed. This is an important result; not only does it demonstrate that Raman micro-spectroscopy can easily be adopted into the standard urine cytology protocol within a clinical setting, but it also provides an alternative to the use of formalin, a common fixative of choice in Raman based cell studies, (158; 180; 182) which is a known carcinogen. (187) While it has been shown by Harvey et al. (55) that the alcohol based fixative SurePath could be used as part of the BD SurePath liquid-based Pap test, SurePath produced a large signal within the fingerprint region, which makes PreservCyt a preferable fixative for urine cytology samples.

In the second experiment, it is shown that UC cells can be exposed to urine for up to 3 days in the presence of PreservCyt without significantly altering the cellular biochemistry. The ability to accurately classify between two cell lines after exposure to urine for 5 h, 24 h, and 72 h has not been significantly affected (see Table 9.1). Interestingly, the results in Table 9.1 indicate that the sensitivities and specificities increase slightly with longer urine exposure times; however, this is more likely to be related to minute changes in the system's alignment or sample preparation methods rather than an improved accuracy that corresponds directly to urine exposure times. Therefore, the combined sensitivities and specificities (91.5% and 93.2% respectively) are a better representation of this method. This result indicates that it may be possible to transport urine samples from a clinic into a Raman lab within a realistic time frame. It is impractical to consider analysing urine samples that have not been exposed to a preservative unless the samples will be processed within a few hours of urine collection due to the rapid deterioration of cells in urine. (188)

In the third experiment, it is demonstrated that the ThinPrep UroCyte method can be applied in combination with Raman micro-spectroscopy resulting in spectra that have a smaller standard deviation than that obtained from cells deposited onto a slide via the drop-dry method. This is most likely due to the uniform monolayer deposition method employed by ThinPrep, as seen in Figure 9.9. ThinPrep also provides a cleaner background, increased cellularity, better preservation, and facilitates easier and faster preparation. (123)

The two chemicals employed by the ThinPrep UroCyte method have little impact on the Raman spectra obtained, and no significant impact in diagnostic classification results are observed. Uniform cellular distribution results in smaller deviations in the background signal when compared to the spectra obtained from cells deposited onto a slide via the drop-dry method. Cells do not require any washing steps after fixation in PreservCyt with ThinPrep due to its filter based properties, whereas drop-dried cells need further washing with PBS in order to avoid spectral contamination from the PreservCyt solution. It should be noted that additional washing steps could result in a reversal of the fixation process of PreservCyt.

The sensitivity and specificity values calculated for ThinPrep prepared samples, as shown in Table 9.2, are greater than 88%. Whilst these results are slightly lower than those found for the drop-dry method, ThinPrep offers many other advantages that make it a preferrable deposition method for Raman micro-spectroscopy. These advantages include ease of preparation, which is particularly important in a busy clinic, the inclusion of red blood cell lysing agents, and a monolayer distribution of cells across the slide, where one focal depth is sufficient to analyse most of the cells on the slide.

In the final experiment, the impact of hematuria on the spectra recorded from UC cells obtained from a urine sample is investigated. Approximately 90% of patients diagnosed with bladder cancer will present with hematuria. Therefore, it is inevitable that Raman spectra will have to be recorded from patients with blood in their urine. The amount of blood will vary from patient to patient, resulting in urine of a mild pink colour (scant hematuria) to deep red (frank hematuria), with the most severe cases resembling the colour of coffee. In this experiment, it is shown that urothelial cells can be easily separated from blood in scant hematuria samples using ThinPrep with CytoLyt, as can be seen in Figure 9.12(A). However, it is more difficult to fully remove red blood cells from frank hematuria samples, resulting in urothelial cells being deposited onto the slide in very close proximity to regions of blood residue (Figure 9.12(B)). The consequence of this is that the Raman spectra obtained from UC cells are contaminated by spectral peaks associated with blood. Additional CytoLyt washes should help to further eliminate this problem, but it may be necessary to wait for the patient to provide a sample at a later date when their levels of hematuria have lowered, e.g. after an antibiotic for infection, or medication for kidney stones, etc.

There are additional factors associated with urine cytology samples that have not been considered here, such as the ability to identify UC cells that are randomly distributed across a slide containing healthy urothelial cells, squamous and glandular cells, renal tubular cells and casts,

inflammation, or infection. It must also be noted that in this chapter, and in previous literature on Raman based urine cytology described in Section 6.5, the number of cells to be recorded from each slide have not been considered in great detail. Shapiro et al. reported recording only five cells per slide; however, since it is possible for the majority of urothelial cells in a typical sample to be non-cancerous, it is unknown how these five cells are chosen or identified. (151) An important problem with standard urine cytology is the inability to differentiate between healthy cells and low grade cancer cells under a microscope, and this task becomes more difficult when the cells are unstained, as required by Raman micro-spectroscopy. Therefore, one reasonable approach may be to record spectra from every available urothelial cell on the slide, and to monitor if any of these cells are classified as cancerous by the trained statistical algorithm. If a cell is identified as such, the corresponding patient could then be flagged for further investigation. Alternatively, it may be possible to train image processing algorithms to identify cells of interest.

Similarly, the number of patients that should be used to accurately train a classifier needs to be further examined. Beleites et al. reported that a sample size of 75–100 is typically needed to verify the accuracy of a classifier. (189) However, it is difficult to train a classifier from urine cytology samples due the low sensitivities associated with the standard urine cytology method; thus, training of algorithms based on biopsy and tissue samples with a known pathological status may be preferable.

9.5 Summary

In summary, the combination of Raman micro-spectroscopy and the ThinPrep UroCyte method provides an ideal platform to replace cystoscopy for bladder cancer surveillance, particularly for high-risk patients who otherwise would require frequent repeat cystoscopic procedures. A key component of this involves the use of inexpensive glass slides, along with the EMSC algorithm that was investigated in Chapter 8. Raman based urine cytology provides all of the advantages associated with standard urine cytology, i.e. ease of procurement, non-invasive, and low cost, as well as the advantages of high sensitivity and specificity that are typically associated with cystoscopy. Additionally, Raman based urine cytology could help to significantly lower the financial burden associated with bladder cancer surveillance for health care systems worldwide, as well as improving the patient's quality of life.

Chapter 10 investigates the potential of integrating the ThinPrep UroCyte method into FTIR spectroscopy, as well as investigating the potential of Raman and FTIR for classifying prostate cancer cells that are present in urine cytology samples.

Chapter 10

Detection of prostate cancer cells with FTIR and Raman micro-spectroscopy

10.1   Introduction

FTIR and Raman micro-spectroscopy are both vibrational spectroscopic techniques, with the physics involved previously discussed in Chapter 2. Both of these methods have been applied extensively for the diagnosis of cancer of the genitourinary system (30; 56; 190; 191; 192; 193; 194); however, these techniques often involve preparing samples using non-traditional clinical protocols, which has hindered their advancement into the cytopathology laboratory. Traditional clinical protocols have been investigated in association with Raman micro-spectroscopy for bladder cancer diagnostics in Chapter 9. This chapter expands upon this concept to include prostate cancer cells as well as FTIR spectroscopy.

Prostate cancer affects one in seven men every year in the USA; however, there remains a lack of modalities available for the accurate diagnosis of this disease. Monitoring of prostate specific antigen (PSA) levels and/or digital rectal examination (DRE) form the first step of the prostate diagnostics, but these techniques present with a high number of false positive results. This study investigates the potential application of FTIR and Raman micro-spectroscopy to identify prostate cancer cells that may be present in a urine sample. In a similar manner to that described in the previous chapter, a simulated study is undertaken, in this case prostate and bladder cell lines are both added to an artificial urine solution, replicating a urine sample obtained following prostatic massage. (195) The synthetic urine samples are prepared using the Thin-Prep UroCyte method, as discussed in further detail in Section 5.5.1. This simulation includes the cell lines PC3, PNT1A, and T24. PC3 represents an aggressive prostate cancer, PNT1A is normal prostate epithelium, and T24 is high grade bladder cancer; these cell lines are chosen as part of this proof of concept study in order to determine if it is possible to identify prostate cancer cells from healthy prostate, and from bladder cancer cells which are morphologically very similar to prostate cancer cells. (196) These three cell lines are exposed to urine and Pre-

servCyt for 24 h, and deposited onto both glass and CaF$_2$ substrates using the T2 machine, as described in Chapter 5. Both methods are compared in terms of classification results, as well as in terms of practical implementation and cost. Cells deposited on CaF$_2$ substrates are analysed with FTIR and Raman micro-spectroscopy, and cells on glass are analysed only with Raman micro-spectroscopy due to the large glass signal present within the fingerprint region of FTIR spectra. (37)

10.1.1 Prostate cancer diagnostics

There are over 230,000 new cases of prostate cancer every year across the United States. (2) This disease has a high rate of recurrence, with 15-52% of prostate cancers progressing to a higher stage. (197; 198) It is evident that the diagnosis, management, and surveillance of these patients results is a significant burden for healthcare systems worldwide, with Sangar and colleagues reporting annual estimated costs of £93 million for prostate cancer on the NHS in the UK. (2)

Detection of prostate cancer is primarily based on an abnormal DRE and/or increased PSA levels; however, prostate biopsy is currently the only way to confirm the presence of clinically localised prostate cancer. Despite its widespread use, 65-70% of males with elevated PSA levels present with negative biopsy results. Importantly, when a patient presents with abnormal PSA levels or DRE, their likelihood of having cancer is 20-25%, and with normal PSA levels and DRE, the possibility of having missed cancer is 10%. (199) Furthermore, biopsies of the prostate are not without significant risks, including infection, systemic sepsis, and hemorrhage, and are known to inherently undersample the prostate, resulting in an estimated 34% false negative rate. (200) Approximately 20-35% of patients that return for a repeat biopsy are diagnosed with prostate cancer that was not detected with the initial biopsy. (200) Therefore, it is generally recommended for patients with negative biopsy results to undergo repeat biopsies with the aim of identifying patients at high risk of concealing prostate cancer. In order to avoid unnecessary biopsies, and the associated risks involved for the patient, and the cost implications for the clinic, it is imperative to develop a non-invasive, accurate test for the diagnosis of prostate cancer.

It is known that prostate cancer cells are shed into biological fluids, particularly after prostatic massage, allowing for the non-invasive detection of these cells in urine or expressed prostatic fluid. (196) Traditional cytology diagnostic methods (i.e. Pap stained slides analysed by a cytopathologist) result in unacceptably low sensitivities for identifying prostate cancer cells. These low sensitivities are a result of low numbers of prostate cancer cells found in urine samples, as well as the difficulty in differentiating between malignant prostate cells and high grade bladder cancer cells, which are morphologically similar to each other. (196)

Cytology coupled with vibrational spectroscopy may prove useful in solving the clinical dilemma of how best to proceed with men who are suspected of harboring prostate cancer. This includes men whose initial biopsy was negative despite an abnormal DRE or PSA, for primary cancer detection after DRE, or for monitoring disease progression in patients who opt

for expectant management (e.g. "watchful waiting" or "active surveillance").

Many studies have previously reported the application of Raman spectroscopy and/or FTIR to classify bladder and prostate cancer cells, resulting in high sensitivities and specificities for classification. (154; 193; 201; 202) However, to translate vibrational spectroscopy into a cytopathology lab, it is important to integrate these technologies with current clinical protocols.

Applications of Raman spectroscopy for classification of bladder and prostate cancer cells from urine samples are reviewed in Section 6.5. FTIR has been extensively applied to both bladder and prostate cancer tissue samples (30; 192; 203; 204), resulting in sensitivities and specificities greater than 90%. However, applications of FTIR to cytology samples are less common. In 2007, Harvey et al. (205) applied FTIR-photoacoustic spectroscopy to differentiate between four formalin fixed prostate cancer cell lines. Bird and colleagues (206) (2008) applied FTIR to discriminate between different cell groups found in voided urine samples. In 2009, Harvey et al. (194) applied FTIR micro-spectroscopy, in transflection mode, to prostate cell lines and primary cells grown on MirrIR slides, showing that classification was independent of the culture medium used or the N/C ratio, indicating that FTIR based classification is explained mainly by the cellular biochemical differences. However, recent studies have identified the presence of electric field standing waves above the reflective MirrIR slides, resulting in unwanted spectral perturbations, making these slides unreliable for FTIR based diagnostics. (37; 207) In 1998, Cohenford et al (125) deposited cervical smear samples onto zinc selenide slides using the ThinPrep method, and applied FTIR to identify the presence of abnormal cervical cells. To the best of our knowledge, FTIR spectroscopy has not previously been applied for prostate or bladder cancer diagnostics on cells obtained from a urine sample, prepared with a T2 machine, and deposited onto CaF_2 slides.

10.2 Methods

10.2.1 Sample preparation

* Note: the cell culturing methods discussed in this section were completed in collaboration with Dr. Marion Butler (Institute of Immunology, Maynooth University, Ireland).

All cell lines are cultured in 1:1 mixture of DMEM and Hams-F12 medium supplemented with 5% fetal bovine serum and 2 mM L-Glutamine. Flasks are maintained in a humidified atmosphere with 5% CO_2 at 37^oC. When the cell lines reach 80% confluency, the culture medium is removed, and the cells are rinsed with sterile PBS. Trypsin-EDTA (0.5%) is added to the flask, which is incubated at 37^oC until the cells have completely detached (not exceeding 15 min). An equal volume of 5% serum-containing medium is added to the flask in order to neutralise the trypsin enzyme. The entire contents of the flask is transferred into a sterilin container, and centrifuged. The supernatant is removed, and the cell pellet is resuspended in fresh medium.

This solution is centrifuged, following which the medium is decanted, and resuspended in 1 ml PBS. This step is repeated, the cell pellets are resuspended in 10 ml artifical urine and 5 ml PreservCyt, and stored at room temperature for 24 h. The vial is then centrifuged, decanted, and the cell pellet is resuspended in a 20 ml PreservCyt vial, and left at room temperature for 15 min. The vial is inserted into the T2 machine, and the cells are transferred onto either a glass or Raman-grade CaF_2 slide.

When operating the T2 machine with CaF_2 substrates, it should be noted that the T2 only accepts slides that are 1" x 3" x 1 mm, with all four corners removed, similar to that shown in Figure 10.1. Raman-grade CaF_2 slides of these dimensions are expensive to purchase, and due to their fragile nature, the removal of the corners is a delicate and time consuming process. Furthermore, due to the underlying cell deposition methods employed by the T2 machine, the cell adherence numbers are lower for CaF_2 substrates when compared to glass slides. It is believed that this is due to the lack of impurities found in CaF_2 substrates, resulting in fewer attractive forces, which are required for good cell adhesion. Nonetheless, adequate numbers of cells are observed on the CaF_2 substrates following ThinPrep deposition, allowing for vibrational spectroscopic data to be recorded.
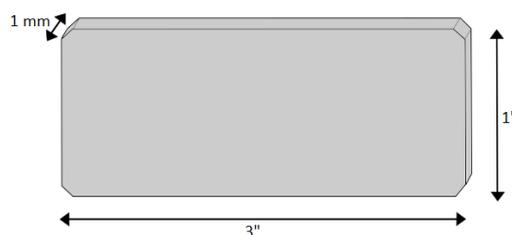
Figure 10.1: *Substrate dimensions that are required in order to fit into the T2 machine.*

10.2.2 Raman data acquisition and processing

Our custom-built Raman micro-spectroscopy system is employed for all measurements in this study, as described in more detail in Section 3.6.1. Spectra are recorded within the 400–1800 cm^{-1} region, with an acquisition time of 5 s each. Two spectra are recorded from the same location within the nucleus of 50 cells from each slide, allowing for the subsequent removal of cosmic rays. (72) All spectra are background corrected using the EMSC algorithm, as described in Chapter 8. The chosen reference spectrum is based on the mean of T24 cells recorded on CaF_2; however, a number of different reference spectra were tested and found to produce similar results. (79; 80; 85) For all data recorded from cells on CaF_2, EMSC is applied using only the reference spectrum, and a first order polynomial (i.e. a straight line). For spectra recorded on glass slides, the glass signal contribution is removed as part of the EMSC process, as well as a first order polynomial. All spectra are classified based on PC-LDA, with sensitivity and specificity values determined using a leave-one-out cross validation, as described in

Section 4.6.3.

10.2.3   FTIR data acquisition and processing

FTIR spectra are recorded using a Perkin Elmer Spotlight 400 FTIR imaging system (Perkin Elmer Inc., USA; located in the Focas Institute, Dublin Institute of Technology, Ireland), with collection of data in transmission mode. Spectra are recorded using the SpectrumIMAGE Spotlight 400 software. Spectral measurements are acquired with a pixel size of 6.25 μm $\times$ 6.25 μm, for 4 scans per pixel, with a spectral resolution of 4 cm^{-1}. Background measurements are recorded from a region on the slide free from any cells or specimen.

EMSC is applied to all FTIR spectra to correct for resonant Mie scattering contributions based on the Matrigel spectrum, as developed by Bassan et al. (81; 83; 82; 32) This algorithm is operated over 5 iterations per spectrum. Following this, spectra are smoothed with a Savitzky-Golay filter ($w = 3$, $k = 41$), and area normalised. Spectra are classified based on PC-LDA, with sensitivity and specificity values determined using a k-fold cross validation approach ($k = 10$).

10.3   Results

10.3.1   Raman results
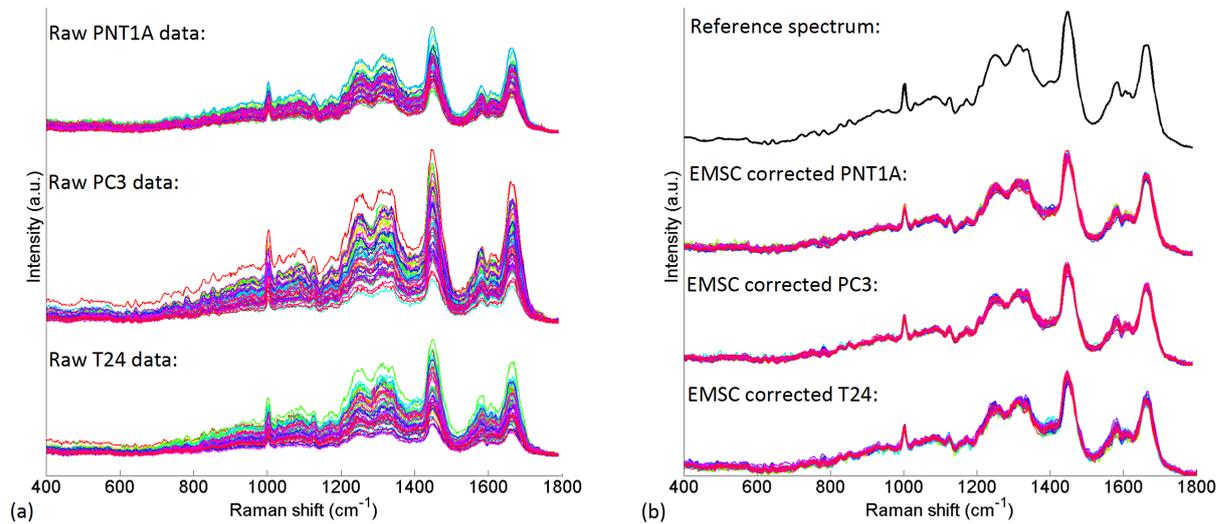
Figure 10.2: *(a) Raw spectra of PNT1A, PC3, and T24 cells, recorded on CaF$_2$ slides; (b) reference spectrum and EMSC corrected PNT1A, PC3, and T24 datasets.*

Figure 10.2(a) shows the raw Raman spectra recorded from PC3, PNT1A, and T24 cells on CaF$_2$ slides, and Figure 10.2(b) shows the reference spectrum, which is input to the EMSC algorithm, as well as the corresponding Raman spectral datasets after background correction

with EMSC. This figure demonstrates a significant reduction in the variance both within and across datasets.
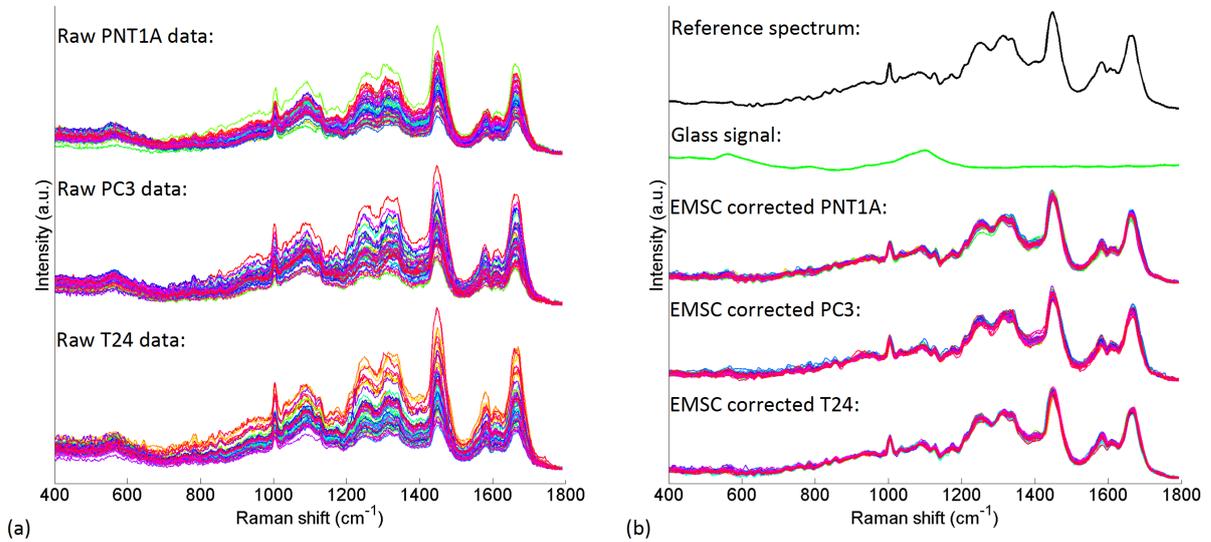
Figure 10.3: *(a) Raw spectra of PNT1A, PC3, and T24 cells, recorded on glass slides; (b) reference spectrum and glass signal input to EMSC, and corrected PNT1A, PC3, and T24 datasets.*

Figure 10.3(a) shows the raw Raman spectra recorded from PC3, PNT1A, and T24 cells on glass slides. The glass signal is evident in the raw spectra within the 500–600 cm^{-1} and 1050–1150 cm^{-1} regions. Figure 10.3(b) displays the reference spectrum and the smoothed glass signal (using a Savitzky-Golay filter; $w = 3$, $k = 41$) both of which are input to the EMSC algorithm. Also shown in the figure are the corresponding Raman spectral datasets after background correction using the EMSC algorithm, as described in Chapter 8. A reduction in the signal intensity can be seen within 500–600 cm^{-1} and 1050–1150 cm^{-1} regions that is associated with the removal of the glass component. Similar to the data shown in Figure 10.2, a significant reduction in variance can be seen across all three datasets following background correction.

PCA and LDA scores are shown in Figure 10.4 for PC3, PNT1A, and T24, with (a) and (b) representing the data recorded on CaF$_2$, (c) and (d) the data recorded on glass, and (e) and (f) representing a combination of the data recorded on both CaF$_2$ and glass. These PCA results show poor separation, although it is possible to separate T24 from both prostate cell lines across the second PC for Raman data recorded on CaF$_2$. Good separation across the three groups can be seen following LDA, as shown in Figure 10.4(b), (d), and (f).

Table 10.1 provides the sensitivities and specificities that are calculated following PC-LDA classification, based on a leave-one-out cross validation for the results shown in Figure 10.4. The results achieved are greater than 92% across both substrates. Interestingly, good classification results can still be achieved when spectra from both substrates are combined. This result shows that it may be possible to record Raman spectra from either substrate, and EMSC background correction can allow for these datasets to be combined when training a multivariate
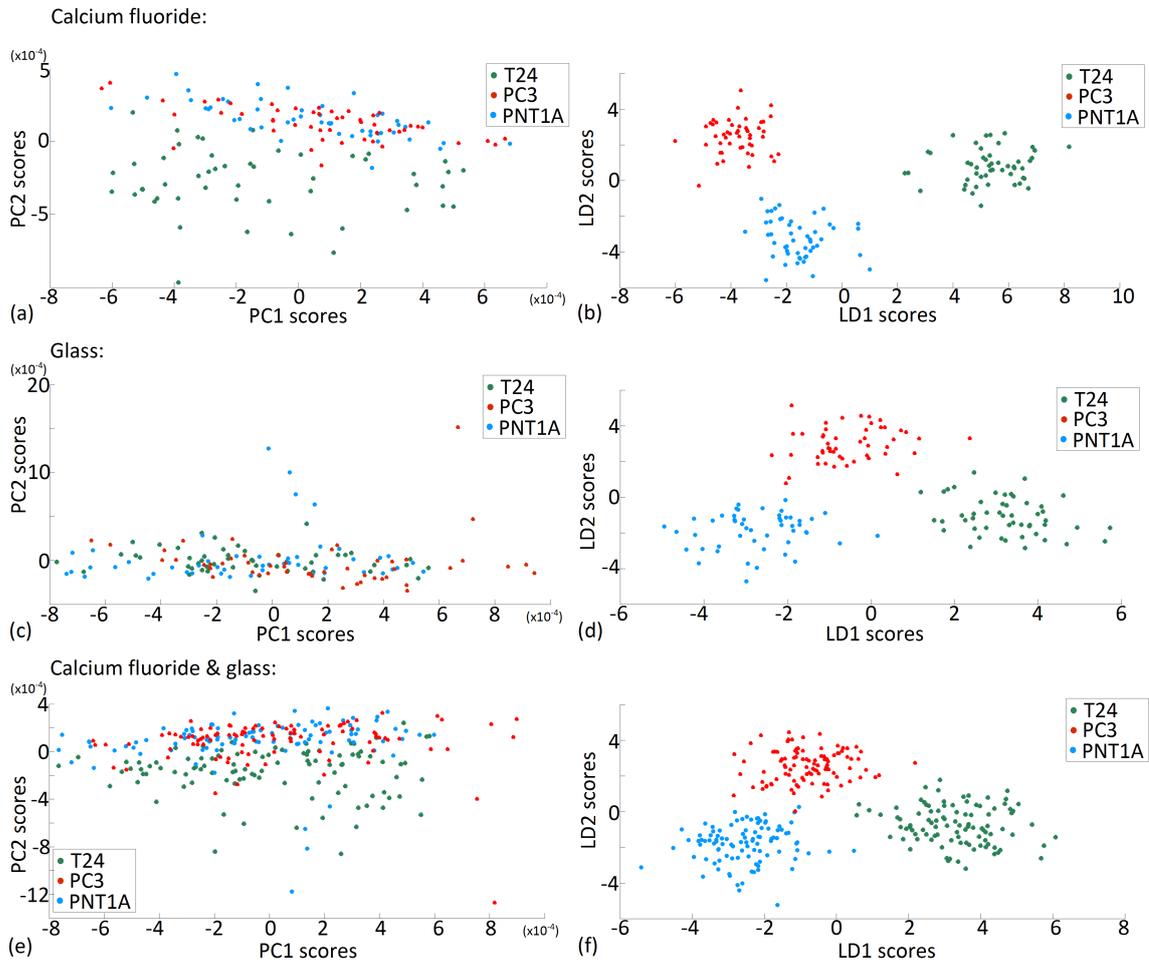
Figure 10.4: *(a) PCA scores and (b) LDA scores for Raman data of PNT1A, PC3, and T24 cells recorded on CaF$_2$; (c) PCA scores and (d) LDA scores for PNT1A, PC3, and T24 Raman cell spectra recorded on glass; (e) PCA scores and (f) LDA scores for PNT1A, PC3, and T24 cells that were recorded on both CaF$_2$ and glass.*

| | Raman micro-spectroscopy: | | | | | | FTIR: | |
| | CaF$_2$: | | Glass: | | CaF$_2$ & glass: | | CaF$_2$: | |
	Sens.:	Spec.:	Sens.:	Spec.:	Sens.:	Spec.:	Sens.:	Spec.:
PC3	96%	97%	96%	98%	92%	99%	93%	99%
PNT1A	96%	97%	96%	97%	95%	95%	98%	99%
T24	94%	97%	92%	96%	95%	98%	97%	96%

Table 10.1: *PC-LDA results of Raman spectra recorded on CaF$_2$ and glass slides, based on a leave-one-out cross validation, and for FTIR spectra recorded on CaF$_2$ based on a k-fold cross validation approach ($K = 10$).*
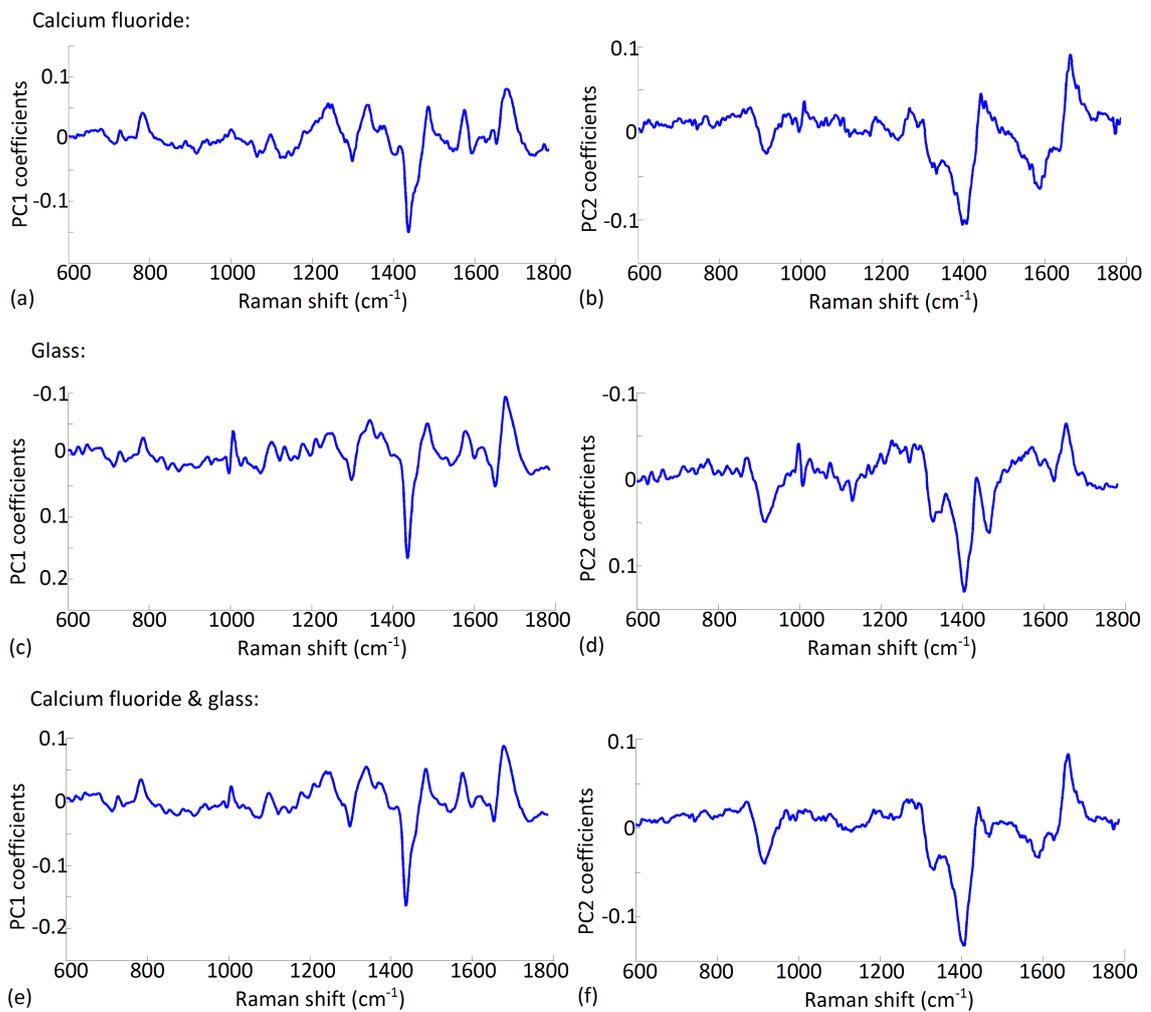
Figure 10.5: *The first and second PC coefficients, (a) and (b) respectively, obtained from Raman spectra of PNT1A, PC3, and T24 cells recorded on CaF$_2$; (c) and (d) represent the first and second PC coefficients obtained from Raman spectra of PNT1A, PC3, and T24 cells recorded on glass; and (e) and (f) are the first and second PC coefficients obtained from PNT1A, PC3, and T24 cells recorded from both on CaF$_2$ and glass. PC coefficients correspond to the PCA scores shown in Figure 10.4(a), (c), and (e), respectively.*

statistical classifier.

The first two PC coefficients from each analysis can be found in Figure 10.5(a) and (b) for CaF$_2$ slides, (c) and (d) for glass slides, and a combination of CaF$_2$ and glass are shown in (e) and (f). Similar features appear across the first two PC coefficients for all three cases, with common peaks observed at 782 (DNA), 1243 (amide III), 1304 (nucleic acids), 1340 (collagen), 1437 (lipids), 1485 (nucleic acids), 1573 (nucleic acids), and 1681 cm^{-1} (amide I) for the first PC, and at 915 (RNA), 1336 (DNA), 1401 (collagen), and 1663 cm^{-1} (DNA) for the second PC. (163; 208) Importantly, there are no prominent features associated with the glass signal appearing in any of these coefficients. This result indicates that spectra recorded on glass slides could be directly compared to Raman spectra recorded on CaF$_2$ slides (assuming adequate system calibration has been achieved, see Section 4.2 for further information). This is an important result for research groups, or clinics, that would like to alter their current protocols to include glass slides, but would like to retain retrospective Raman datasets that have been recorded on CaF$_2$ substrates in their lab for inclusion in ongoing studies.

10.3.2 FTIR results

Figure 10.6 shows the fingerprint region (1000–2000 cm^{-1}) of the FTIR spectra recorded from PNT1A, PC3, and T24 cells on CaF$_2$, with the raw spectra shown in Figure 10.6(a), and spectra following EMSC resonant Mie scatter correction shown in Figure 10.6(b). Similar to Raman data (as shown in Figures 10.2 and 10.3), the variance has been reduced across each FTIR dataset following EMSC, along with a reduction of any spectral shifting that may have been present in the raw spectra due to Mie scattering artifacts. (81)
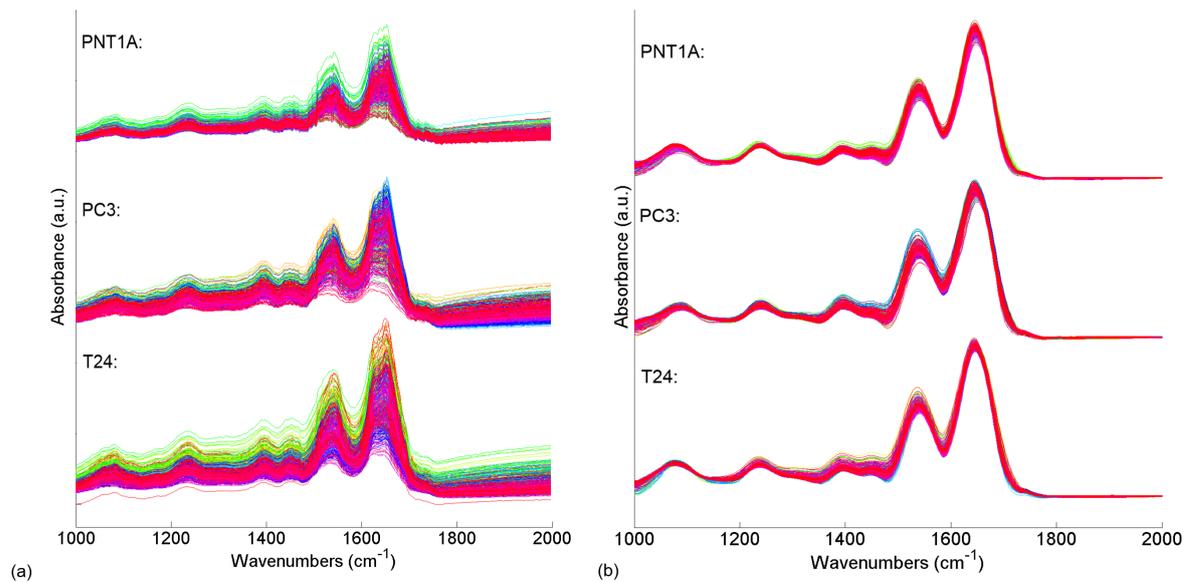
Figure 10.6: *FTIR spectra of PNT1A, PC3, and T24 cells, recorded on CaF$_2$ slides, with (a) representing the raw spectra, and (b) shows the corresponding spectra following EMSC resonant Mie scattering correction.*
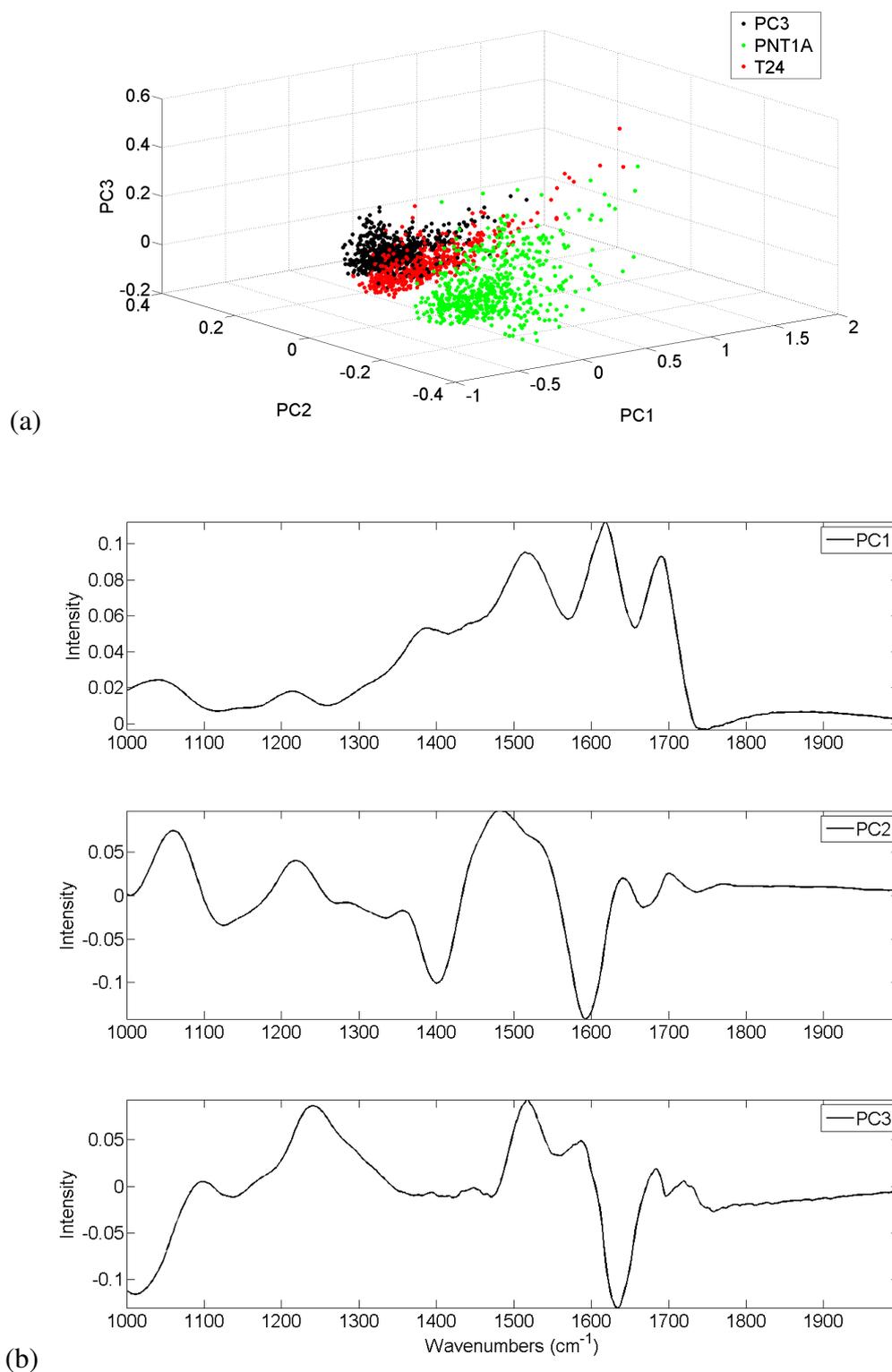
Figure 10.7: *(a) PCA scores for FTIR spectra of PNT1A, PC3, and T24 cells, and (b) shows the first three corresponding PC coefficients.*

Table 10.1 shows the PC-LDA classification results obtained for the data shown in Figure 10.6(b), based on a $k-$fold cross validation ($k = 10$). These results are similar to those obtained with Raman micro-spectroscopy, in terms of the ability to classify across all three cell lines, with sensitivities and specificities obtained of greater than 93%.

Based on the PCA scores shown in Figure 10.7(a), it can be seen that the FTIR spectra are separable based on PCA alone, whereas all three Raman spectral datasets are only separable following PC-LDA. Interestingly, PC3 and T24 cells are grouped closer together relative to PNT1A cells, allowing for complete separation between cancerous and healthy cells, similar to the separation observed between cancerous and healthy cells across the first LD for Raman spectra recorded on CaF_2, as shown in Figure 10.4(b).

As shown in Figure 10.7(a), most of the separation across the FTIR spectral datasets occurs over the second PC. Figure 10.7(b) shows the first three PCs, with the main peak differences seen in the second PC at 1060 (glycogen), 1220 (PO_2 asymmetric stretching, Amide III), 1400 (Amide III), 1470 (CH_3 asymmetric stretching, proteins, lipids), and 1580 cm^{-1} (Amide III). (30; 191; 202)

10.4 Summary

The results presented in this chapter suggest that Raman and FTIR spectroscopy have the potential to identify prostate cancer cells that are present in a urine cytology sample prepared using the ThinPrep UroCyte method, which is commonly used in clinical cytopathology laboratories. The results presented here show that Raman micro-spectroscopy can be applied to both CaF_2 and glass slides, and that these recordings can be carried out somewhat interchangeably, as well as showing that FTIR spectroscopy can be applied to CaF_2 slides that have been processed using the ThinPrep method. These results reinforce previous studies that have suggested that vibrational spectroscopy could be applied to slides obtained directly from the clinic. (78) However, it is important to highlight some of the incompatibility issues that have been observed with CaF_2 and the T2 machine, as discussed in Section 10.2.1, as well as acknowledging the financial burden associated with operating a diagnostic technique using expensive CaF_2 substrates.

In order to overcome any issues with CaF_2 and the T2 machine, it may be possible to incorporate glass slides into FTIR based cytology studies in a similar manner as Raman micro-spectroscopy. Bassan et al (37) recently showed the ability to classify between paraffin embedded breast cancer tissue samples in the higher wavenumber region. However, the combination of glass slides and FTIR have not been considered here as this study focused only on the fingerprint region of the FTIR spectrum. Furthermore, the complex resonant Mie scattering signals that are inherent in FTIR cytology spectra may prove problematic to remove using EMSC due to the varying levels of the glass signal present across each spectrum. Nonetheless, this may prove interesting for future studies, and it may be possible to expand the resonant Mie scattering EMSC algorithm to incorporate the glass signal, similar to that presented in Chapter 8.

Similar classification results have been achieved within this study for both vibrational spectroscopy methods, as shown in Table 10.1. However, it is also important to consider other aspects of both these techniques in terms of the practical implementation in the clinic. For this study, Raman spectra are recorded from 50 cells on each slide, with a total acquisition time of 10 s per cell. With the inclusion of additional time required to locate cellular nuclei and the correct z-focal depth position, the overall time required to analyse one Raman cytology slide is on the order of 1-2 h, with longer times required for inexperienced users. FTIR on the other hand allows for the analysis of large regions of a slide, and therefore larger numbers of cells, within the order of 10-15 min, even when operated by an inexperienced user. Whilst the times required for both techniques could be improved by automating the systems involved, Raman micro-spectroscopy remains a much slower technique than FTIR. However, pre-processing of both groups of spectral datasets shows that Raman algorithms are significantly faster to run than the complex algorithms required to remove Mie scattering artifacts from FTIR spectra, resulting in similar time frames overall for the application of either technique.

Given the degree of accuracy obtained here, which is significantly higher than that achieved with PSA or DRE, there is a good argument for the possible introduction of vibrational spectroscopic techniques into the clinic to assist cytopathologists in accurately identifying the presence of prostate cancer cells. Prior to the widespread application of core biopsy, many prostate cancers were diagnosed with fine needle aspiration (FNA); however, core biopsy samples produced higher accuracies in determining the presence of prostate carcinoma. With the ThinPrep method, it is also possible to prepare FNA samples to produce similar cytology slides as that achieved with urine cytology, with the added benefit of larger numbers of prostate cells. This study has shown that Raman and FTIR micro-spectroscopy may have the potential to improve upon the diagnostic capabilities of FNA samples. FNA carries the additional benefits of being quicker, less expensive to perform, and safer for the patient. (209) FNA samples are routinely obtained from patients suspected of harbouring breast cancer; Chapter 11 investigates the application of Raman micro-spectroscopy for the analysis of such simulated breast cytology samples.

Chapter 11

Monitoring of changes in cellular biochemistry associated with hypoxia and exosomal communication

11.1   Introduction

Exosomes are microvesicles (40–100 nm in diameter) of endocytic origin released into the extracellular environment by most cell types, as briefly discussed in Section 5.4. It is believed that they play a significant role in cell signalling and communication by virtue of their capacity to traffic miRNA, mRNA, and proteins to recipient target cells. (112) Furthermore, it has previously been shown that the levels of exosomes in patients with cancer are significantly higher than those in normal plasma, with hypoxic tumours producing even more exosomes when compared to the levels observed from normoxic tumours. (119) This chapter investigates the application of Raman micro-spectroscopy for (i) identifying the differences between cells growing in normoxia versus cells growing in hypoxia, and (ii) monitoring of any changes in cellular biochemistry upon exposure to exosomes. These measurements are based on MDA-MB-231 triple negative breast cancer (TNBC) cells, although it is believed that similar changes could be induced in cells within the urinary bladder under these conditions. TNBC is a specific subtype of epithelial tumour which lacks protein expressions for the estrogen receptor (ER), the progesterone receptor (PR), or the human EGF receptor 2 (HER2), and is often associated with a high occurrence rate of hypoxia. (210) Approximately 15% of all breast cancers are TNBC, although it accounts for 30% of breast cancers amongst African-American women. (211)

Hypoxia is a key feature of many tumours throughout the body, and whilst this chapter focuses on TNBC, it is believed that similar physiological changes would be observed within bladder and prostate tumours. Hypoxia has previously been observed in bladder (212; 213; 214) and prostate (215; 216) tumours, as well as investigations on the release of their associated exosomes (113; 217; 218). This chapter reports the application of Raman micro-spectroscopy to

identify any biomolecular changes that occur due to either the presence of hypoxia or interactions with exosomes.

11.1.1 Hypoxic TNBC tumours

TNBC is associated with advanced disease stage, BRCA1 gene mutations, a younger age at presentation, and has a poorer prognosis compared to other breast cancer subtypes. TNBC tumours are innately hypoxic, a tumour micro-environment integral to aggressive tumour behaviours such as increased invasion. Tumour hypoxia (reduced levels of O_2 availability) occurs when rapid cancer cell proliferation leads to the growth of cells that are too far away from a blood vessel, and therefore these cells become structurally and functionally abnormal. (211) Hypoxia exerts significant effects on the tumour micro-environment with profound consequences for cancer progression, often leading to patient mortality. Additionally, hypoxia has been linked with invasion, angiogenesis, apoptosis, metastasis, chemoresistance, and radiation resistance. (210; 219) Knowledge of the degree and extent of hypoxia is invaluable for patient treatment planning. Current techniques for measuring hypoxia include polarographic needle electrodes, ^{19}F magnetic resonance imaging (MRI), and phosphorescence imaging. (219) However, there remains a clinical need for a technique to measure hypoxia on a cellular level in a reliable and reproducible manner. The first experiment within this chapter is based on investigating the ability of Raman micro-spectroscopy to classify between normoxic and hypoxic TNBC cells, as well as identifying the main biomolecular differences between both groups.

11.1.2 Exosomes

It has previously been shown that breast cancer cells release greater levels of exosomes upon exposure to hypoxia (220); the impact of such exosomes on the surrounding tumour micro-environment is an ongoing subject of interest, with increasing evidence showing that tumour cell-derived exosomes play an important role in angiogenesis, metastasis, and tumour progression. (220)

Exosomes are defined as extracellular vesicles that are released from cells upon fusion of the MVEs with the plasma membrane, as previously shown in Figure 5.4. Over the last decade, there has been significant interest into the understanding of exosomes, and their role in cell communication, disease transmission, and their potential application for drug delivery. (113; 221; 222; 223) However, despite research efforts, little is known of the basic biology involved in these processes, or how exosomes affect normal cellular physiology. Furthermore, it is difficult to isolate exosomes from other similar sized extracellular vesicles, with many studies using the terms interchangeably, resulting in great scepticism amongst the research community.

The second aim of this chapter is to monitor any changes within the cellular biochemistry of TNBC cells upon exposure to exosomes. To achieve this, exosomes are isolated from normoxic and hypoxic cells respectively, across three different batches, and are introduced to separate

batches of cells for 48 h, following which Raman micro-spectroscopy is applied for cellular analysis.

11.1.3 Fine needle aspiration

All cytology samples within this chapter are prepared for Raman micro-spectroscopic analysis using the ThinPrep method, as previously discussed in Section 5.5.1. Throughout this thesis, the ThinPrep UroCyte method has been applied to urine cytology specimen; however, with this study, the cell pellets are placed in a vial of PreservCyt, representing FNA breast specimen. As discussed in Section 10.5, FNA samples are often prepared with ThinPrep method in the clinic (224), resulting in cytology slides that can be analysed using Raman micro-spectroscopy. Breast FNA is a relatively simple procedure, which removes cells from a breast lesion or cyst with a needle that is thinner than that used to obtain a standard blood sample. FNA is safer for the patient, and relatively inexpensive to process, when compared with core needle biopsies. (225) In contrast to prostate diagnostics, FNA analysis is currently a popular technique employed in the clinic for the analysis of suspicious breast lesions.

11.2 Materials and methods

11.2.1 Sample preparation

Raman micro-spectroscopy is sensitive to minute changes across samples and experimental protocols, batch-to-batch cellular variations, and daily variations of the Raman system. These changes can result in small differences in the Raman spectra that can lead to irreproducible results. (226) Therefore, in order to optimise the reproducibility and reliability of the results presented in this study, the first experiment consists of three batches of cells grown in normoxia (21% O_2) and hypoxia (1% O_2), respectively. Similarly, the second experiment consists of three batches of normoxic cells grown in the presence of exosomes that have been released from normoxic cells (NEx), and three batches grown with exosomes released by hypoxic cells (HEx). Furthermore, Raman spectral measurements are recorded randomly across each cytology sample in order to further reduce any experimental variability.

Cell culture

* Note: The work outlined in the following cell culture subsection was completed by our collaborator Luke Gubbins (Conway Institute, University College Dublin, Ireland).

MDA-MB-231 cells are grown in RPMI 1640 culture medium, in either normoxic conditions (21% O_2) or hypoxic conditions (1% O_2) for 48 h, until 70–80% confluency is reached. For parent cell analysis in the first experiment, the cells are trypsinised, and centrifuged into a cell

pellet, as described in Section 10.2.1. The cell pellet is suspended in a 20 ml vial of PreservCyt, and left at room temperature for 15 min.

Exosome isolation

* Note: The work outlined in the following exosome isolation subsection was completed by our collaborator Luke Gubbins (Conway Institute, University College Dublin, Ireland).

Cells grown to 70-80% confluency are cultivated for a further 48 h in medium supplemented with exosome depleted fetal calf serum. The supernatants are collected and centrifuged at 300g for 10 min, following by 2,000g for 20 min to remove cellular debris. Supernatants are filtered using a 0.22 μm filter (Millex; Merck Millipore, Ireland), transferred into fresh ultracentrifuge tubes, and centrifuged at 120,000g for 80 min at 4^0C in a Beckman Coulter Optima LE-80K ultracentrifuge (Beckman Coulter Inc., USA). The supernatant is completely removed, and the pellet in each tube is resuspended in 1 ml of PBS. The resuspended pellets from the same cell line are pooled in a single ultracentrifuge tube, and PBS is added. The sample is centrifuged again at 120,000g for 80 min at 4^0C. The supernatant is removed, and the exosomal pellet is resuspended in 50-100 μl PBS, and stored at -80^0C. The presence of exosomes is verified with western blotting, NanoSight measurements (Malvern, USA), and mass spectrometry (results not shown here).

Co-culture cells with exosomes

* Note: The work outlined in the following co-culture subsection was completed by our collaborator Luke Gubbins (Conway Institute, University College Dublin, Ireland).

MDA-MB-231 cells are trypsinised and counted, and seeded at an initial density of 180,000 cells per well in six well plates. The cells are allowed to adhere for approximately 12 h. The medium is changed, and replaced with RPMI 1640 supplemented with 10% v/v exosome depleted fetal calf serum, 1% penicillin/streptomysin (Gibco, ThermoFisher Scientific, Ireland). The exosomes (either derived from normoxic or hypoxic cells) are added, and the cells are grown for a further 48 h in normoxic conditions. For co-culture cell analysis in the second experiment, the cells are trypsinised, and centrifuged into a cell pellet, as described in Section 10.2.1. The cell pellet is suspended in a 20 ml vial of PreservCyt, and left at room temperature for 15 min.

ThinPrep preparation

The vials containing PreservCyt and the cell pellets are inserted into the T2 machine, along with ThinPrep glass slides and filter tubes. The cells are transferred onto the glass slide based on the T2 fluid/FNA setting.

11.2.2 Raman spectral acquisition

Our custom-built Raman micro-spectroscopy system is employed for all measurements in this study, as described in further detail in Section 3.6.1. Spectra are recorded within the 600–1800 cm^{-1} range with an acquisition time of 5 s each; two spectra are recorded from the same location within the nucleus, and are averaged together using an algorithm that simultaneously removes cosmic rays (72), with 60 cells recorded from each slide. Spectra are wavenumber calibrated, and background corrected using the EMSC algorithm to remove both the slowly varying baseline signal, and the unwanted glass signal, as described in Chapter 8. The EMSC algorithm is operated with a reference spectrum recorded from MDA-MB-231 cells on CaF$_2$ and a fifth order polynomial.

PC-LDA is applied to the datasets in order to classify between cells that are grown in a normoxic environment compared to those grown in an hypoxic environment. Further PC-LDA classifications are applied to identify cells that have been exposed to either NEx or HEx. All classifications are based on a leave-one-out cross validation approach.

11.3 Results

Figure 11.1 shows all of the Raman spectra recorded from normoxic and hypoxic parent cells, across all three batches, following EMSC background correction. Similar to the spectra shown in previous chapters, the glass signal has been removed from these spectra.

Following PCA analysis, Figure 11.2(a) shows the PCA scores for normoxic and hypoxic parent cells (all batches combined), with most separation occurring across a combination of PC1 and PC2. Further PCA analysis is applied across each batch individually, resulting in the PC coefficients shown in Figure 11.2, with (b) and (c) representing PC1 and PC2 for the first batch of parent cells, respectively, (d) and (e) representing PC1 and PC2 for the second batch of parent cells, and (f) and (g) representing PC1 and PC2 for the third batch of parent cells. Focusing on the first PC coefficient, tentative peak assignments are provided for the compositional differences that can be seen across all three batches relating to DNA (782, 1098, 1336, 1373 cm^{-1}), nucleic acids (1485, 1577 cm^{-1}), lipids (1064, 1128, 1299, 1437, 1744 cm^{-1}), amino acids (869 cm^{-1}), phenylalanine (1002 cm^{-1}), amide III (1243 cm^{-1}), and amide I (1676 cm^{-1}). Biomolecular differences across the second PC coefficient relate to DNA (725, 782, 1098 cm^{-1}), amino acids (928, 1548, 1622 cm^{-1}), lipids (1064, 1294, 1434 cm^{-1}), proteins (1152, 1573 cm^{-1}), ribose (996 cm^{-1}), amide III (1243 cm^{-1}), collagen (1401), nucleic acids (1485 cm^{-1}), and amide I (1652 cm^{-1}), with peak assignments obtained from the following references: (163; 202; 208; 227; 228; 229). Observation of consistent peak differences across all three batches for both PCs allows for the identification of the biochemical changes that occur as cells are exposed to hypoxic environments. Interestingly, the majority of these peak differences relate to DNA, nucleic acids, and lipid content, which is in
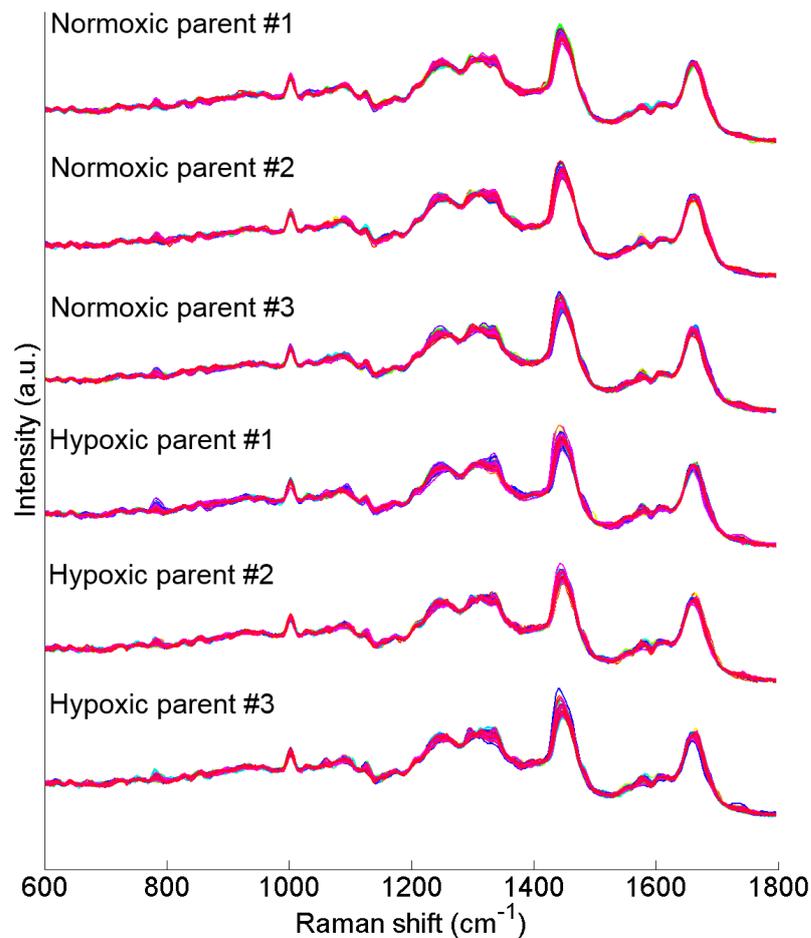
Figure 11.1: *Raman spectra of normoxic and hypoxic MDA-MB-231 cells, recorded from all three batches, following EMSC background correction.*

agreement with previous reports that have linked cell growth dependence on unsaturated fatty acids for survival under O_2 deprivation. (230)

Table 11.1 provides the sensitivities and specificities obtained following PC-LDA classification for each batch, and for all batches combined. These results show a good ability to differentiate between cells growing under normoxic conditions compared to those in hypoxia using Raman micro-spectroscopy. The clinical relevance of such results allows for FNA cytology samples to be identified as originating from a tumour with or without hypoxic regions, which could allow for improved patient treatment planning. This also raises the possibility of applying Raman fiberoptic probes within a surgical procedure for identifying hypoxic regions within a tumour, as well as identifying the borders between normoxic and hypoxic regions. The best results are achieved for the second and third batches, which may relate to batch-to-batch variations, or small differences within the Raman system during the recording of the first batch of cells. This highlights the importance of repeating experiments across multiple batches and recording of spectra across different days in order to combine the results to provide a more realistic classifier.
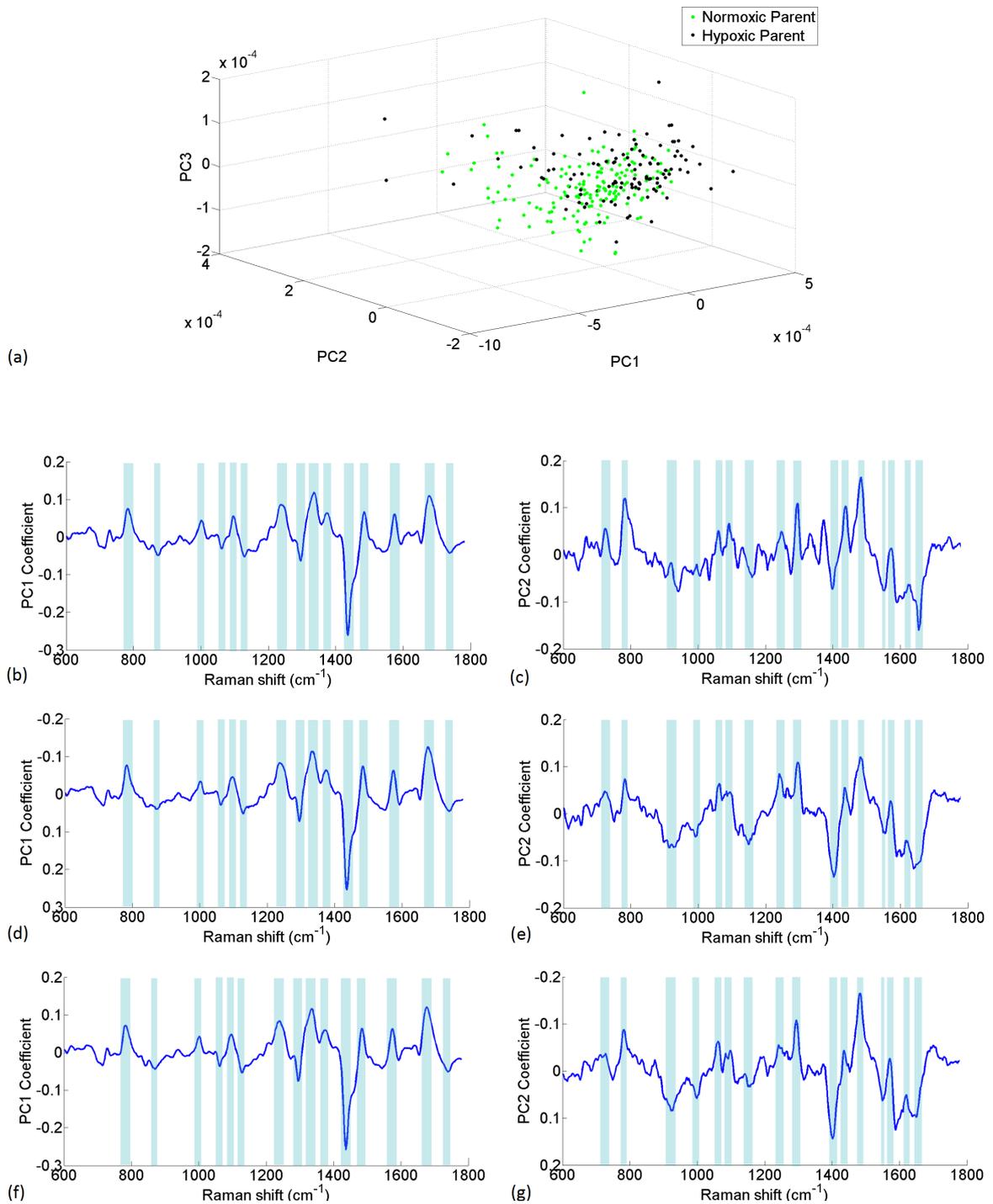
Figure 11.2: *(a) PCA scores for normoxic and hypoxic parent cells (all batches combined), (b) PC1 coefficient and (c) PC2 coefficient for normoxic and hypoxic parent cells (Batch 1), (d) PC1 coefficient and (e) PC2 coefficient for normoxic and hypoxic parent cells (Batch 2), and (f) PC1 coefficient and (g) PC2 coefficient for normoxic and hypoxic parent cells (Batch 3). Consistent peak differences across all three batches are highlighted in blue.*

	Sensitivity:	Specificity:
Batch 1:		
Normoxic parent cells	72%	68%
Hypoxic parent cells	68%	72%
Batch 2:		
Normoxic parent cells	95%	87%
Hypoxic parent cells	87%	95%
Batch 3:		
Normoxic parent cells	98%	97%
Hypoxic parent cells	97%	98%
All batches:		
Normoxic parent cells	84%	88%
Hypoxic parent cells	88%	84%

Table 11.1: *PC-LDA classification results of Raman parent cell spectra recorded across each individual batch, and for all batches combined.*

Figure 11.3 shows the Raman spectra recorded from MDA-MB-231 cells grown in the presence of either NEx or HEx, for each of the three batches, following EMSC background subtraction. Figure 11.4(a) shows the PCA scores for normoxic parent cells and normoxic cells exposed to NEx; poor separation is seen across the first three PCs for both of these groups. Figure 11.4(b) shows the PCA scores for normoxic parent cells and normoxic cells exposed to HEx; both of these cell groups are separable across a combination of PC2 and PC3.

Figure 11.5 shows the first PC coefficients for each batch of normoxic parent cells when compared with normoxic cells exposed to either NEx or HEx, with (a) and (b) representing the results from the first batch, (c) and (d) correspond to the second batch, and (e) and (f) relate to the third batch. The first PC coefficients are similar for all batches, with consistent peak differences observed for DNA (782, 1098, 1334 cm^{-1}), lipids (1297, 1437 cm^{-1}), nucleic acids (1484, 1575 cm^{-1}), phenylalanine (1002 cm^{-1}), amide III (1243 cm^{-1}), and amide I (1676 cm^{-1}), with peak assignments obtained from the following references: (163; 202; 208; 227; 228; 229).. Interestingly, these peak differences are also observed in PC1 coefficients for PCA analysis of the normoxic and hypoxic parent cells (see Figure 11.2(b), (d), and (f)).

Similar to Figure 11.5, Figure 11.6 shows the second PC coefficients for each batch of co-cultured measurements, with (a) and (b) corresponding to the first batch of cells, (c) and (d) corresponding to the second batch, and (e) and (f) representing the third batch. The second PC coefficients show less correlation across each batch, with three major peak differences observed across each batch relating to lipids (1294 cm^{-1}), collagen (1401 cm^{-1}), and amide I (1655 cm^{-1}), as highlighted in yellow in Figure 11.6. Consistent peaks observed across batch one and batch three correspond to DNA (724, 785, 1132 cm^{-1}), and nucleic acids (1180,
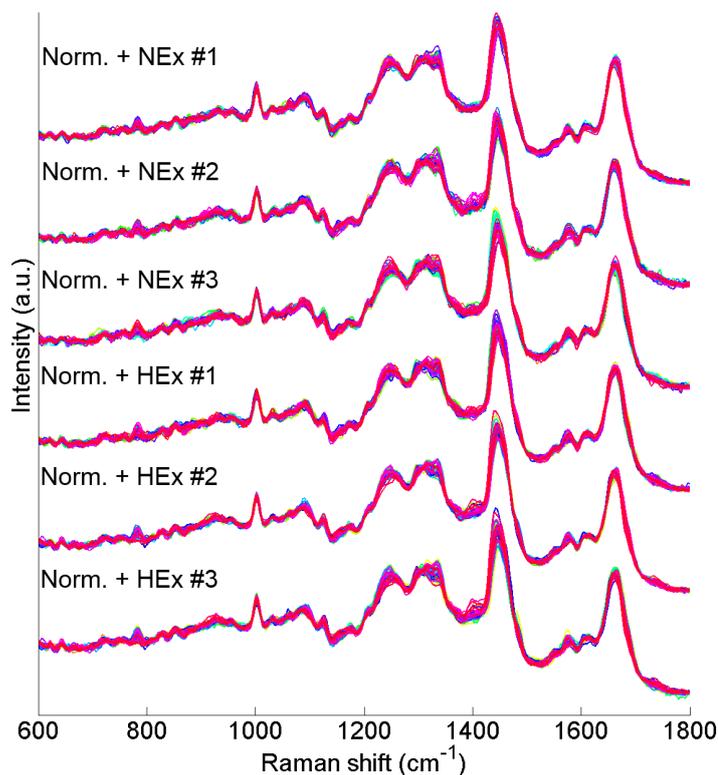
Figure 11.3: *Raman spectra of MDA-MB-231 cells growing in normoxic after exposure to either NEx or HEx, recorded across all three batches, following EMSC background correction.*
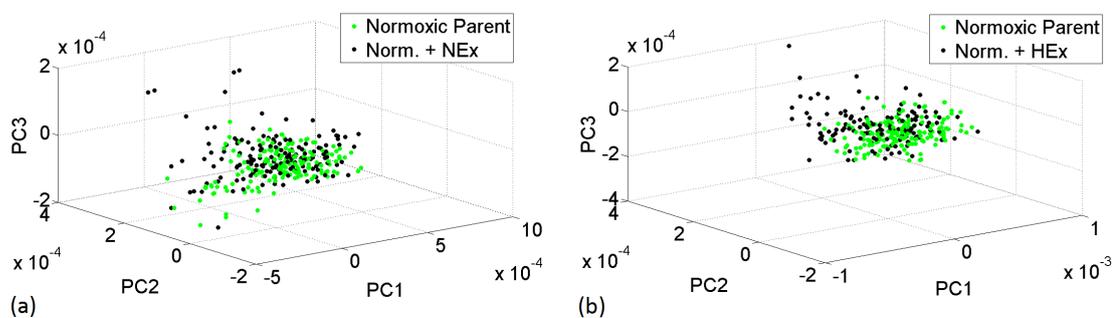
Figure 11.4: *(a) PCA scores for normoxic parent cells and normoxic cells exposed to NEx, (b) PCA scores for normoxic parent cells and normoxic cells exposed to HEx. These scatter plots represent the Raman data recorded across all three batches.*
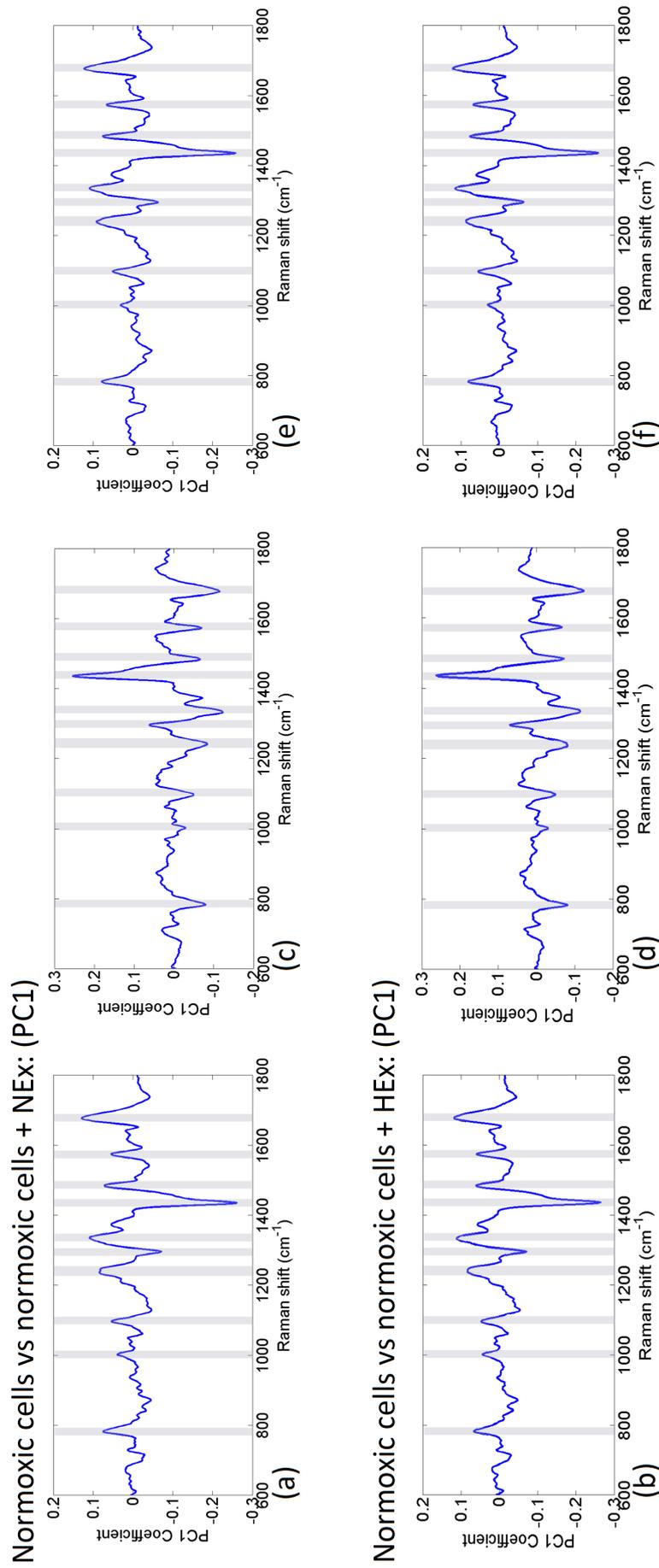
Figure 11.5: *PC1 coefficients for each of the co-culture experiments, where (a) and (b) represent PC1 coefficients for batch one, (c) and (d) represent batch two, and (e) and (f) correspond to batch three. Common peak differences across all six PC1 coefficients are highlighted in grey.*
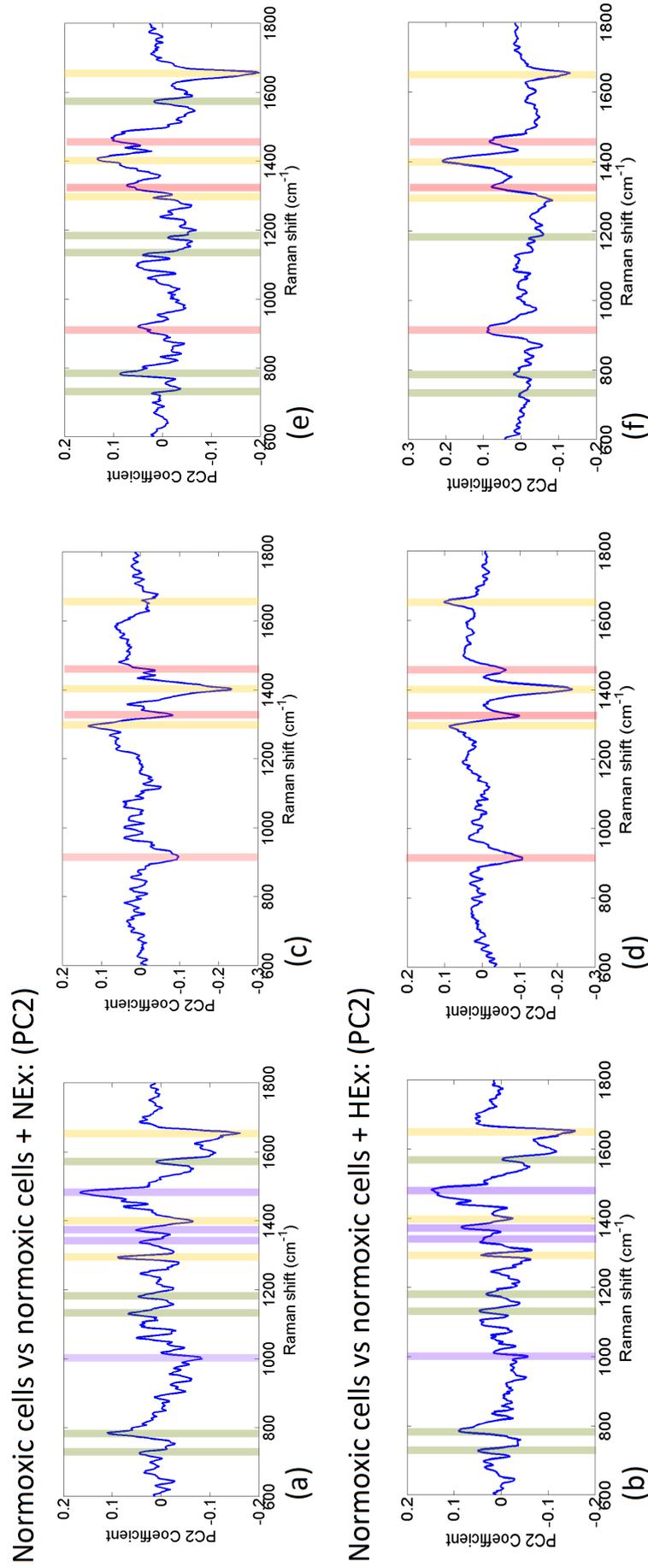
Figure 11.6: PC2 coefficients for each of the co-culture experiments, where (a) and (b) represent PC2 coefficients for batch one, (c) and (d) represent batch two, and (e) and (f) correspond to batch three. Common peak differences across different batches are highlighted in yellow, green, pink, and purple.

	Sensitivity:	Specificity:
Batch 1:		
Normoxic parent cells	92%	88%
Normoxic cells with NEx	88%	92%
Batch 2:		
Normoxic parent cells	95%	75%
Normoxic cells with NEx	75%	95%
Batch 3:		
Normoxic parent cells	95%	73%
Normoxic cells with NEx	73%	95%
All batches:		
Normoxic parent cells	87%	78%
Normoxic cells with NEx	78%	87%
Batch 1:		
Normoxic parent cells	100%	98%
Normoxic cells with HEx	88%	92%
Batch 2:		
Normoxic parent cells	97%	87%
Normoxic cells with HEx	87%	97%
Batch 3:		
Normoxic parent cells	80%	80%
Normoxic cells with HEx	80%	80%
All batches:		
Normoxic parent cells	87%	81%
Normoxic cells with HEx	81%	87%

Table 11.2: *PC-LDA classification results of Raman spectra recorded from normoxic cells and normoxic cells exposed to either NEx or HEx, across each individual batch and for all batches combined.*

1571 cm^{-1}), as highlighted in green, and across batch two and batch three correspond to collagen (915-935 cm^{-1}), nucleic acids (1327 cm^{-1}), and DNA (1457 cm^{-1}), highlighted in pink in Figure 11.6. The PC coefficients are most consistent across batches two and three with respect to each other. Further peak differences that relate to phenylalanine (1002 cm^{-1}), DNA (1338, 1375 cm^{-1}), and nucleic acids (1483 cm^{-1}) are observed within batch one that correspond to cells exposed to either NEx or HEx, as shown in Figure 11.6(a) and (b) respectively, with specific peaks highlighted in purple. Overall, these results show that there is good consistency across batches two and three, however, there is poor consistency with batch one, which might correspond to differences in exosomal composition within each batch, therefore resulting in
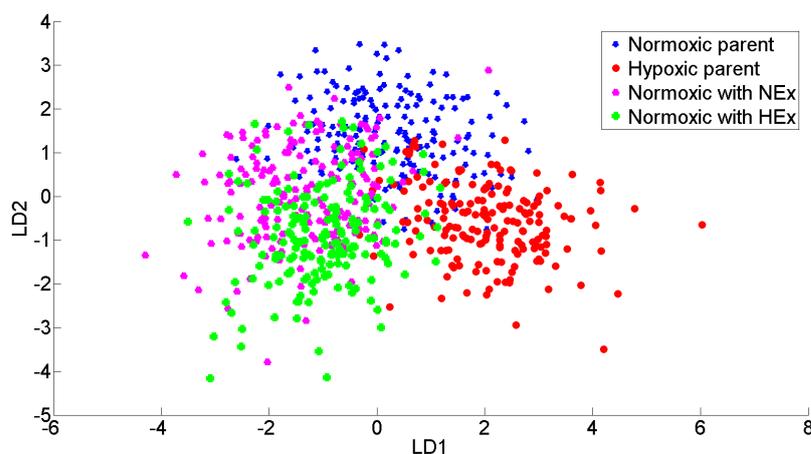
Figure 11.7: *LDA scores for normoxic and hypoxic parent cells, and normoxic cells exposed to either NEx or HEx.*

	Sensitivity:	Specificity:
Batch 1:		
Normoxic parent cells	71%	88%
Hypoxic parent cells	58%	89%
Normoxic cells with NEx	83%	96%
Normoxic cells with HEx	95%	99%
Batch 2:		
Normoxic parent cells	88%	90%
Hypoxic parent cells	88%	96%
Normoxic cells with NEx	55%	93%
Normoxic cells with HEx	68%	88%
Batch 3:		
Normoxic parent cells	75%	86%
Hypoxic parent cells	87%	97%
Normoxic cells with NEx	57%	91%
Normoxic cells with HEx	58%	85%
All batches:		
Normoxic parent cells	72%	87%
Hypoxic parent cells	83%	94%
Normoxic cells with NEx	61%	86%
Normoxic cells with HEx	63%	91%

Table 11.3: *PC-LDA classification results of Raman spectra recorded from normoxic and hypoxic parent cells, and normoxic cells exposed to either NEx or HEx, across each individual batch and for all batches combined.*

slightly different physiological changes. However, these results also show that the changes induced in normoxic cells upon exposure to exosomes is somewhat independent on whether those exosomes were released by either normoxic or hypoxic parent cells. Nonetheless, the majority of peak assignments relate to changes in DNA and nucleic acid content, a trait commonly observed with cancer cell progression.

Table 11.2 presents the sensitivities and specificities obtained following PC-LDA classification for normoxic cells and normoxic cells exposed to either NEx or HEx, across each batch, and for all batches combined. For NEx, it is possible to separate cells that have been exposed to NEx compared to those that have not, with greater than 73% accuracy, and despite the differences seen across the second PC coefficient for each batch, the results show a good ability to identify both cell groups, even when all batches are combined. Greater classification accuracies are obtained for cells exposed to HEx, with results achieved greater than 80%. Similarly, it is possible to identify cells when spectra recorded across all batches are combined for PC-LDA classification.

By combining all of the data recorded across this study, it is possible to apply PC-LDA for whole group classification. Figure 11.7 shows the LDA scatter plot obtained, with good separation observed across normoxic and hypoxic parent cells based on LD2. Hypoxic cells can be separated from normoxic cells that are co-cultured with either NEx or HEx based on LD1. Interestingly, there is a significant amount of overlap between cells exposed to either NEx or HEx, which reinforces the results shown in Figure 11.6, whereby the oxygen conditions of the cells that exosomes are secreted from has little impact on the biomolecular changes that they induce in other cells.

Table 11.3 shows the sensitivities and specificities obtained for classification when all four cell subtypes are combined. The results presented here vary from batch to batch, with sensitivity values ranging from 55-95%. Nonetheless, specificity values remained above 86%.

11.4   Summary

This study consists of two main experiments, with the first focusing on the ability of Raman micro-spectroscopy and multivariate statistical algorithms to classify between TNBC cells that are growing in normal oxygen conditions compared to those growing in hypoxia, and the second aims to identify any differences that are induced in cells upon exposure to exosomes obtained from either normoxic or hypoxic cells. Clinically, TNBC is recognised as an innately hypoxic cancer type, and the ability to identify the presence of hypoxia has significant implications for the patient involved. The results presented here show that consistent biomolecular changes occur within TNBC cells when they are growing in hypoxia, as shown in the PC coefficients presented for all three batches in Figure 11.2. Additionally, most of these changes relate to DNA content, nucleic acids, and lipids, which is in agreement with previous publications that link hypoxic cell growth with changes in lipid content. (230)

The combination of Raman micro-spectroscopy and PC-LDA shows the ability to classify between normoxic and hypoxic cells that are prepared with the ThinPrep method. Clinically, hypoxic tumour cells can become radiation and chemotherapy resistant, requiring specific treatment plans in order to improve patient survival rates. (210; 219) However, current diagnostic modalities do not always provide reliable or reproducible detection of hypoxic regions. Therefore, it may be possible to integrate Raman micro-spectroscopy into current clinical protocols for the analysis of FNA cytology samples, in order to identify patients harbouring hypoxic tumours, such that alternative treatment plans can be implemented to target these cells. Additionally, it may be possible to include Raman based fiberoptic probes, via endoscopes or hypodermic needles, into a surgical procedure to assist in the identification of hypoxic tumour regions. (231; 232)

Focusing on the results presented in the second study, it has been shown that exosomes can induce biomolecular changes within cells growing in normoxic conditions. These changes are independent of the source of the exosomes, i.e. the same changes are induced for NEx and HEx. However, the peak differences observed in Figure 11.6 show that the biomolecular changes are dependent upon the cell batch that the exosomes are isolated from. In order to fully understand the reasoning behind this, further analysis of the exosomal content across each batch is required; such analyses could form part of a future study involving the application of mass spectrometry or SERS (surface enhanced Raman spectroscopy) to investigate if different batches of cells produce exosomes containing different levels of proteins, mRNA, or miRNA, which might be responsible for the communication of different signals with recipient cells. Despite these inconsistencies, this study has shown that exposing cells to exosomes results in a change in cellular biomolecular composition. Further research is required to determine the impact of exosomes released by cancerous cells on neighbouring healthy cells, and whether this can result in the spreading of cancerous tumours.

PC-LDA based classification of normoxic cells and cells exposed to either NEx or HEx shows a good ability to identify both cell groups. This is an interesting result, as despite the biomolecular differences seen across the second PC coefficient, the combination of PCA and LDA can separate both groups, even when all three batches are combined. With further research, this combination of Raman micro-spectroscopy and PC-LDA could provide the clinician with information on the presence of cancer cells within a tumour if an FNA cytology sample shows the presence of healthy cells that have been exposed to cancerous exosomes.

Chapter 12

Conclusion

12.1 Summary of thesis contributions

The primary objective of this thesis is for the development of Raman based urine cytology for bladder cancer detection and surveillance, with a key emphasis on the translation of Raman micro-spectroscopy into the clinic. This PhD project began with the publication of a detailed literature review (*L.T. Kerr et al., Photon. Lasers Med., 3(3), 193-224, 2014*), which identified a variety of source wavelengths, sample substrates, fixation agents, and cell deposition methods that have been applied for Raman based urine cytology to date. Therefore, within this thesis, a range of traditional protocols and consumables have been systematically examined in terms of their compatibility with Raman micro-spectroscopy. Compatibility of some of these methodologies with FTIR spectroscopy has also been examined, along with an extension of the approach in order to detect prostate cancer cells.

The first achievement within this thesis involves the design and building of a confocal Raman micro-spectrometer, as described in Section 3.6.1. In order to design a Raman system, it is necessary to determine the impact of different source wavelengths on the associated Raman spectrum, and to choose the optimum laser for this application. Five source wavelengths ranging from 473-830 nm are investigated, showing that NIR lasers result in less sample damage, but the Raman signal obtained has a significantly lower intensity than that achieved with a source in the visible wavelength region.

Another important consideration is the compatibility of the sample substrate with the chosen source wavelength, as published in the following paper: *L.T. Kerr et al., Anal. Methods, 7(12), 5041-5052, 2015*. This thesis compared ten different substrates for Raman cytology in terms of the associated background signals, cost, the ability to re-use such slides, and their biocompatibility. The best results, in terms of the quality of the Raman spectra obtained, along with the lowest background signals, are shown for Raman-grade CaF_2 and aluminium thin film substrates. However, CaF_2 is expensive, particularly for routine applications within the cytopathology clinic. Aluminium thin film substrates, whilst cheaper than CaF_2, are still more

expensive than glass slides, and the thin film surface can be easily damaged or scratched, potentially altering the Raman spectra due to Mie scattering. Glass slides are cheap, readily available in the clinic, and are biocompatible, making them an ideal substrate for large scale clinical applications; however, they produce a large background signal within the Raman fingerprint region. Additionally, this signal varies, in terms of location, shape, and intensity, depending on the source wavelength.

Following from the conclusion that glass slides are a necessary consumable from a clinical perspective due to their low cost, three background correction algorithms are investigated within this thesis for the removal of the glass signal from Raman spectra recorded from cells with a 532 nm laser (as published in *L.T. Kerr et al., Chemometr. Intell. Lab., 158, 61-68, 2016*). Results show that the EMSC algorithm is the best choice for the removal of the glass signal in conjunction with the inherent slowly varying baseline signal. Furthermore, the application of PCA to spectra corrected using this algorithm shows that the results obtained are independent of the choice of reference spectrum used. This indicates that the reference spectrum does not introduce any biasing of PCA based classification.

In order to move Raman micro-spectroscopy a step closer to the clinic, various protocols and consumables are examined in terms of their impact on Raman cell spectra. In the absence of patient samples, simulated urine samples are created in order to replicate the stringent conditions present within the cytopathology clinic (published in *L.T. Kerr et al., Anal. Methods, 25(8), 4991-5000, 2016*). The use of glass slides, traditional fixation agents, and the ThinPrep cell deposition method are investigated, along with lengthy exposure to urine, and the application of red blood cell lysing agents to remove residual blood present from patients suffering with hematuria. Results show that the combination of the ThinPrep UroCyte method with Raman micro-spectroscopy provides an ideal platform for the identification of bladder cancer cells. It has also been shown that these cytology samples can be stained post-Raman acquisition, and that the cells do not appear to have been photodamaged by the laser, thus indicating that a cytopathologist should still be able to perform standard urine cytology diagnostics on the slides if desired.

The integration of the ThinPrep UroCyte method with CaF_2 substrates is investigated, allowing for the integration of this cell preparation technique with FTIR spectroscopy. However, poorer cell adhesion is observed with CaF_2 samples that are prepared in this manner compared to the use of glass slides; nonetheless, adequate numbers of cells are observed across CaF_2 for vibrational spectroscopic analysis. As a direct comparison between FTIR and Raman micro-spectroscopies, similar classification results are achieved from both methods for the identification of prostate cancer cells within a urine sample that also contains bladder cancer and healthy prostate epithelium. FTIR is a significantly faster, and easier, technique for the acquisition of spectra, however, Raman requires less numerical pre-processing due to the lack of complex resonant Mie scattering artifacts. These results also showed the ability to obtain similar PCA coefficients from cells recorded on CaF_2 and glass substrates following EMSC background sub-

traction. With further development of this algorithm, and acquisition of larger datasets across a variety of cell types, it may be possible to use either substrate interchangeably throughout an experiment, provided a robust calibration tool has been used.

The final component of this thesis focuses on the application of Raman micro-spectroscopy for the identification of cells growing in hypoxic conditions. Hypoxia is a feature of many forms of cancer throughout the body, including both the bladder and the prostate. Analysis of breast cancer cells growing in both normoxia and hypoxia shows that the main biomolecular differences associated with both oxygen conditions relate to DNA, nucleic acids, and lipid content. A triplicate experiment, using a combination of Raman micro-spectroscopy and PC-LDA shows the ability to classify between normoxic and hypoxic cells, which may allow for the application of this technique to identify patients requiring more intensive treatment plans. The second part of this study involves the analysis of cells that are grown in the presence of exosomes (nano-sized vesicles that have been hypothesized to play a role in cellular communication and disease transmission), and whilst the results observed are not reproducible across each batch, biomolecular changes are observed upon exposure to exosomes, indicating that some form of communication may be taking place. In order to fully understand the role of exosomes in cancer communication, further experimentation is required, with the inclusion of a larger number of cancerous and non-cancerous cell lines.

12.2   Clinical translation

When designing a diagnostic device, it is important to acknowledge its place within the market, whether it is needed by the clinician, and how much value it would have if sold commercially. Within this section, the potential for the future commercialisation, and clinical translation, of Raman based urine cytology for bladder cancer surveillance is discussed. In order to bring such a device to the market, it would be necessary to develop a university spin-out company. The company should be focused on the following objectives:

- Develop its technologies to final product and approval stage, and apply for technology patenting,

- Establish a full medical advisory board,

- Conduct clinical trials with a large cohort of patients,

- Apply for regulatory approval (FDA, 510k, CE mark of medical device regulation (National Standards Authority of Ireland), 98/79/EC in vitro diagnostic medical device directive, and S.I. No. 304 of European Communities Regulations 2001),

- Establish corporate identity, brand names, and trademarks.

Another important consideration is to assess the current technologies available on the market, and to determine what advantages Raman based urine cytology can provide. A brief analysis of the urine cytology market shows that the UroVysion test (Abbott Molecular Inc., USA) can be used as an aid to identify bladder cancer cells based on the presence of chromosomal abnormalities. (233) Similarly, the BTA Stat test (Polymedco Inc., USA) can identify specific antigens associated with bladder cancer. (234) A similar Raman based device is currently available for cervical cytology (CervAssist, Raman Diagnostics Ltd., Ireland); however, there are no such devices available for urine cytology. Raman based urine cytology offers three distinct advantages over these technologies:

- As demonstrated within this thesis, there is no additional sample preparation required; urine cytology samples prepared using standard clinical protocols and consumables are compatible with Raman micro-spectroscopy,

- The Raman system can become fully automated (235), allowing for clinical staff to simply place the cytology slide onto the microscope stage; such a system could potentially identify cells, record spectra, and apply statistical algorithms for classification,

- This device could potentially offer clinicians a tool to identify patients with early-stage bladder cancer, which would result in better treatment planning and patient survival rates.

It is believed that there is a pre-existing market for a highly accurate, non-invasive device for the diagnosis of urothelial carcinoma via urine cytology. The target market for a Raman based urine cytology device would initially be hospitals and health care systems that currently spend a large percentage of their annual budget on the surveillance of patients suspected of harbouring bladder cancer; these patients are typically aged between 60 and 85 years, and current cystoscopic surveillance can result in significant anxiety and morbidity for these patients. With further development, it may become possible to develop a Raman device that is suitable for use within a general practitioners office, further easing patient anxiety.

On a broader scale, knowledge of medical regulations, the need for adequate clinic trials, and the patenting process are essential for research groups worldwide who are developing technologies that are suitable for commercialisation. Scientists and engineers within universities often strive to make a significant contribution to human knowledge; however, it is important to remember to connect with businesses, and to aim to translate research from the lab into the commercial world. In recent years, the gap between university research and industry has narrowed, and with government funded grants being aimed towards research with a large potential for commercialisation (e.g. Enterprise Partnership Scheme, Ireland), it is imperative that researchers are aware of the selling potential of their work, as having customers who are willing to invest in your product or ideas is often more significant than the peer review process.

12.3 Future work

It is expected that the application of Raman based urine cytology for bladder cancer diagnostics will continue to expand and develop over the next decade, and it is hoped that the investigations and results presented within this thesis will aid and enhance the work of future researchers in this area. With more time, this project could have been developed further to include:

- More comprehensive cell line/urine simulation studies, with the inclusion of healthy bladder cells, different subtypes/grades of bladder cancer, as well as the presence of other cells found in a typical urine sample (i.e. urethral, ureter, and kidney cells).

- A clinical study, with urine cytology samples obtained from patients currently undergoing cystoscopic procedures for bladder cancer detection or surveillance.

- Development of a tool for a fully calibrated Raman micro-spectroscopy system, allowing for the comparison of spectra recorded across a large time frame, or across different Raman systems.

- Development of larger training and testing datasets for classification would provide a more accurate representation of the diagnostic capabilities of Raman micro-spectroscopy.

- Further optimisation of the EMSC background correction algorithm to include a "universal" Raman reference spectrum that could be used by research groups worldwide, similar to the Matrigel spectrum applied to FTIR spectra.

- A greater understanding of the biochemistry associated with specific peak assignments; further work is required to allow for accurate, and reliable, peak assignments to be made for Raman spectra. Whilst many studies have previously investigated the biochemistry and physical chemistry involved, there are many instances of where peaks have multiple assignments in the literature, related to different biomolecules, resulting in great scepticism amongst the research community.

- Further investigations into the impact of cellular morphology on the associated Raman spectra obtained, and how Mie scattering can impact upon the recorded Raman signal.

- Monitoring of biomolecular changes observed when non-cancerous cells are exposed to exosomes derived from cancerous cells using Raman micro-spectroscopy.

- Analysis of other cell types growing in normoxia, hypoxia, and anoxia (severe hypoxia); these results may prove useful for cancer diagnostics, but also for understanding brain injuries associated with a lack of oxygen.

Bibliography

[1] Raman CV. The colour of the sea. Nature (London). 1921;108(367):10–11.

[2] Sangar VK, Ragavan N, Matanhelia SS, Watson MW, Blades RA. The economic consequences of prostate and bladder cancer in the UK. BJU Int. 2005;95(1):59–63.

[3] Tetu B. Diagnosis of urothelial carcinoma from urine. Mod Pathol. 2009;22(2):S53–59.

[4] Strutt JW. On the light from the sky, its polarization and colour. Philos Mag Series 4. 1871;41(271):107–120.

[5] Compton AH. A quantum theory of the scattering of x-rays by light elements. Phys Rev. 1923;21(5):483–502.

[6] Compton AH. The spectrum of scattered x-rays. Phys Rev. 1923;22(5):409–413.

[7] Raman CV, Krishnan KS. The optical analogue of the Compton effect. Nature (London). 1928;121(711):377–378.

[8] Saleh BEA, Teich MC. Fundamentals of Photonics. 2nd ed. John Wiley & Sons; 2007.

[9] Mazilu M, De Luca AC, Riches A, Herrington CS, Dholakia K. Optimal algorithm for fluorescence suppression of modulated Raman spectroscopy. Opt Express. 2010;18(11):11382–11395.

[10] Atkins P, de Paula J. Atkins' Physical Chemistry. OUP Oxford; 2010.

[11] Wollrab JE. Rotational Spectra and Molecular Structure: Physical Chemistry. Academic Press; 2013.

[12] Ingle JD, Crouch SR. Spectrochemical Analysis. Prentice Hall; 1988.

[13] Smith E, Dent G. Modern Raman Spectroscopy: A Practical Approach. Wiley; 2004.

[14] Ornstein LS, Rekveld J. Intensity measurements in the Raman effect and the distribution law of Maxwell-Botlzmann. Phys Rev. 1929;34:720–725.

[15] Thornton ST, Rex A. Modern Physics for Scientists and Engineers. Cengage Learning; 2012.

[16] Sathyanarayana DN. Vibrational Spectroscopy: Theory and Applications. New Age International; 2015.

[17] Atkins PW, Friedman RS. Molecular Quantum Mechanics. OUP Oxford; 2011.

[18] Ferraro JR, Nakamoto K, Brown CW. Introductory Raman spectroscopy. Academic Press, Inc.; 2002.

[19] Sasi Kumar PR. Photonics: An Introduction. PHI Learning Pvt. Ltd.; 2012.

[20] Mishra RK. Molecular and Biological Physics of Living Systems. Springer Science and Business Media; 2012.

[21] Bernath PF. Spectra of Atoms and Molecules. Oxford University Press; 2005.

[22] Wang D, Guo W, Hu J, Liu F, Chen L, Du S, et al. Estimating atomic sizes with Raman spectroscopy. Sci Rep. 2013;3:1486.

[23] Brooks RL. The Fundamentals of Atomic and Molecular Physics. Springer Science and Business Media; 2014.

[24] Lyng FM, Faolain EO, Conroy J, Meade A, Knief P, Duffy B, et al. Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool. Exp Mol Pathol. 2007;82(2):121–129.

[25] Diem M, Mazur A, Lenau K, Schubert J, Bird B, Miljkovic M, et al. Molecular pathology via IR and Raman spectral imaging. J Biophotonics. 2013;6(11-12):855–886.

[26] Clemens G, Hands JR, Dorling KM, Baker MJ. Vibrational spectroscopic methods for cytology and cellular research. Analyst. 2014;139:4411–4444.

[27] Kerr LT, Byrne HJ, Hennelly BM. Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. Anal Methods. 2015;7(12):5041 – 5052.

[28] Prasad PN. Introduction to Biophotonics. Wiley-Interscience; 2003.

[29] Smith BC. Fundamentals of Fourier Transform Infrared Spectroscopy. CRC Press; 1995.

[30] Baker MJ, Gazi E, Brown MD, Shanks JH, Gardner P, Clarke NW. FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer. BJ Cancer. 2008;99:1859–1866.

[31] Hughes C, Iqbal-Wahid J, Brown M, Shanks JH, Eustace A, Denlet H, et al. FTIR microspectroscopy of selected rare diverse sub-variants of carcinoma of the urinary bladder. J Biophotonics. 2013;6(1):73–87.

[32] Bassan P, Byrne HJ, Lee J, Bonnier F, Clarke C, Dumas P, et al. Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells. Analyst. 2009;134:1171–1175.

[33] El-Tawil SG, Adnan R, Muhamed ZN, Othman NH. Comparative study between Pap smear cytology and FTIR spectroscopy: a new tool for screening for cervical cancer. Pathology. 2008;40(6):600–603.

[34] Old OJ, Fullwood LM, Scott R, Lloyd GR, Almond LM, Shepherd NA, et al. Vibrational spectroscopy for cancer diagnostics. Anal Methods. 2014;6(12):3901–3917.

[35] Ollesch J, Heinze M, Heise HM, Behrens T, Bruning T, Gerwert K. It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy. J Biophotonics. 2014;7(3-4):210–221.

[36] Yoshida S, Toshida M, Yamamoto M, Takeda J. Optical screening of diabetes mellitus using non-invasive Fourier-transform infrared spectroscopy technique for human lip. J Pharm Biomed Anal. 2013;25(76):169–176.

[37] Bassan P, Mellor J, Shapiro J, Williams KJ, Lisanti MP, Gardner P. Transmission FT-IR chemical imaging on glass substrates: applications in infrared spectral histopathology. Anal Chem. 2014;86:1648–1653.

[38] Bonnier F, Mehmood A, Knief P, Meade A, Hornebeck W, Labkin H, et al. In vitro analysis of immersed human tissues by Raman microspectroscopy. J Raman Spectrosc. 2011;42:888–896.

[39] Bonnier F, Ali SM, Knief P, Lambkin H, Flynn K, McDonagh V, et al. Analysis of human skin tissue by Raman microspectroscopy: Dealing with the background. Vib Spectrosc. 2012;61:121–132.

[40] Dieing T, Hollricher O, Toporski J. Confocal Raman microscopy. Springer Science and Business Media; 2011.

[41] Svoboda K, Block SM. Biological applications of optical forces. Annu Rev Biophys Biomol Struct. 1994;23:247–285.

[42] Pitt GD, Batchelder DN, Bennett R, Bormett RW, Hayward IP, Smith BJE, et al. Engineering aspects and applications of the new Raman instrument. IEEE Proceedings - Science, Measurement and Technology. 2005;152(6):241–318.

[43] Li R, Verreault D, Payne A, Hitchcock CL, Povoski SP, Martin EW, et al. Effects of laser excitation wavelength and optical mode on Raman spectra of human fresh colon, pancreas, and prostate tissues. J Raman Spectrosc. 2014;45(9).

[44] Vij DR, Mahesh K. Medical Applications of Lasers. Springer US; 2002.

[45] Welch AJ, Torres JH, Cheong WF. Laser physics and laser-tissue interactions. Tex Heart Inst J. 1989;16(3):141–149.

[46] Waynant RW. Lasers in Medicine. CRC Press; 2002.

[47] Wieboldt D. Understanding Raman spectrometer parameters. Spectroscopy. 2010;Special issue.

[48] Kiselev R, Schie IW, Askrabic S, Krafft C, Popp J. Design and first applications of a flexible Raman micro-spectroscopic system for biological imaging. Biomed Spectrosc Imaging. 2016;5(2):115–127.

[49] Adar F. Considerations of grating selection in optimizing a Raman spectrograph. Spectroscopy. 2013;28(10).

[50] Lerner JM. Imaging spectrometer fundamentals for researchers in the biosciences - a tutorial. Cytometry, Part A. 2006;69(8):712–734.

[51] Farhane Z, Bonnier F, Caset A, Maguire A, O'Neill L, Byrne HJ. Cellular discrimination using in vitro Raman micro spectroscopy; the role of the nucleolus. Analyst. 2015;140:5908–5919.

[52] Ashkin A. Acceleration and trapping of particles by radiation pressure. Phys Rev Lett. 1970;24:156–159.

[53] Neuman KC, Block SM. Optical trapping. Rev Sci Instrum. 2004;75(9):2787–2809.

[54] Fore S, Chan J, Taylor D, Huser T. Raman spectroscopy of individual monocytes reveals that single-beam optical trapping of mononuclear cells occurs by their nucleus. J Opt. 2011;13(4):044021.

[55] Harvey TJ, Hughes C, Ward AD, Faria EC, Henderson A, Clarke NW, et al. Classification of fixed urological cells using Raman tweezers. J Biophotonics. 2009;2(1-2):47–69.

[56] Gazi E, Ward AD, Clarke NW, Harvery TJ, Snook RD, Gardner P, et al. Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers. J Biomed Opt. 2008;13(6):064004.

[57] Harvey TJ, Hughes C, Ward AD, Gazi E, Faria EC, Clarke NW, et al. The use of Raman tweezers and chemometric analysis to discriminate the urological cell lines, PC-3, LNCaP, BPH and MGH-U1. American Institute of Physics AIP Conference Proceedings Volume 1075, Perspectives in Vibrational Spectroscopy (ICOPVS 2008) Eds VK Vaidyan and VS Jayakumar. 2008;p. 138–140.

[58] Harvey TJ, Hughes C, Ward AD, Gazi E, Faria EC, Clarke NW, et al. Analysis of prostate and bladder cells using Raman tweezers. Proceedings of the XXI International Conference on Raman Spectroscopy. 2008;p. 1015–1017.

[59] Utzinger U, Richards-Kortum RR. Fiber optic probes for biomedical optical spectroscopy. J Biomed Opt. 2003;8(1):121–147.

[60] Lewis IR, Edwards H. Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line. CRC Press; 2001.

[61] Hutsebaut D, Vandenabeele P, Moens L. Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy. Analyst. 2005;130:1204–1214.

[62] Choquette SJ, Etz ES, Hurst WS, Blackburn DH, Leigh SD. Relative intensity correction of Raman spectrometers: NIST SRMs 2241 through 2243 for 785 nm, 532 nm, and 888 nm/514.5 nm excitation. Appl Spectrosc. 2007;61(2):117–129.

[63] Isabelle M, Dorney J, Lewis A, Lloyd GR, Old O, Shepherd N, et al. Multi-centre Raman spectral mapping of oesophageal cancer tissues: a study to assess system transferability. Farad Discuss. 2016;187:87–103.

[64] Wolthuis R, Bakker Schut TC, Caspers PJ, Buschman HPJ, Romer TJ, Bruining HA, et al. Raman spectroscopic methods for in vitro and in vivo tissue characterization. In: Fluorescent and luminescent probes for biological activity. San Diego: Academic Press. 1999;p. 433–455.

[65] Bocklitz TW, Dorfer T, Heinke R, Schmitt M, Popp J. Spectrometer calibration protocol for Raman spectra recorded with different excitation wavelengths. Spectrochim Acta A Mol Biomol Spectrosc. 2015;5(149):544–549.

[66] Dorfer T, Bocklitz T, Tarcea N, Schmitt M, Popp J. Checking and improving calibration of Raman spectra using chemometric approaches. Z Phys Chem. 2011;255:753–764.

[67] Rodriguez JD, Westenberger BJ, Buhse LF, Kauffman JF. Standardization of Raman spectra for transfer of spectral libraries across different instruments. Analyst. 2011;136:4232–4240.

[68] Dussault D, Hoess P. Noise performance comparison of ICCD with CCD and EMCCD cameras. Proc SPIE, Image Reconstruction from Incomplete Data III. 2004;5563:195–204.

[69] Irie K, McKinnon AE, Unsworth K, Woodhead IM. A model for measurement of noise in CCD digital-video cameras. Measurement Science and Technology. 2008;19(4).

[70] Barton SJ, Kerr LT, Domijan K, Hennelly BM. On the effect of experimental noise on the classification of biological samples using Raman micro-spectroscopy. Proc SPIE, Biophotonics: Photonics for Better Health Care. 2016;9887.

[71] Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36(8):1627–1639.

[72] James TM, Schlosser M, Lewis RJ, Fischer S, Bornschein B, Telle HH. Automated quantitative spectroscopic analysis combining background subtraction, cosmic ray removal, and peak fitting. Appl Spectrosc. 2013;67(8):949–959.

[73] Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc. 2003;57(11):1363–1367.

[74] Cao A, Pandya AK, Serhatkulu GK, Weber RE, Dai H, Thakur JS, et al. A robust method for automated background subtraction of tissue fluorescence. J Raman Spectrosc. 2007;38:1199–1205.

[75] Byrne HJ, Knief P, Keating ME, Bonnier F. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. Chem Soc Rev. 2016;45(7):1865–1878.

[76] Matthews JH, Fink KD. Numerical Methods Using Matlab: Third Edition. Prentice Hall; 1999.

[77] Beier BD, Berger AJ. Method for automated background subtraction from Raman spectra containing known contaminants. Analyst. 2009;134:1198–1202.

[78] Bonnier F, Traynor D, Kearney P, Clarke C, Knief P, Martin C, et al. Processing Thin-Prep cervial cytological samples for Raman spectroscopic analysis. Anal Methods. 2014;6:7831–7841.

[79] Liland KH, Kohler A, Afseth NK. Model-based pre-processing in Raman spectroscopy of biological samples. J Raman Spectrosc. 2016;47(6):643–650.

[80] Afseth NK, Kohler A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. Chemometr Intell Lab. 2012;117:92–99.

[81] Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, et al. Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst. 2010;135:268–277.

[82] 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.

[83] Bassan P, Byrne HJ, Bonnier F, Lee J, Dumas P, Gardner P. Resonant Mie scattering in infrared spectroscopy of biological materials - understanding the 'dispersion artefact'. Analyst. 2009;134:1586–1593.

[84] Smulko JM, Dingari NC, Soares JS, Barman I. Anatomy of noise in quantitative biological Raman spectroscopy. Bioanalysis. 2014;6(3):411–421.

[85] Kerr LT, Hennelly BM. A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides. Chemometr Intell Lab. 2016;Accepted for publication.

[86] Tfayli A, Gobinet C, Vrabie V, Huez R, Manfait M, Piot O. Digital dewaxing of Raman signals: discrimination between nevi and melanoma spectra obtained from paraffin-embedded skin biopsies. Appl Spectrosc. 2009;63(5):564–570.

[87] Farhane Z, Bonnier F, Casey A, Byrne HJ. Raman micro spectroscopy for in vitro drug screening: subcellular localisation and interactions of doxorubicin. Analyst. 2015;140:4212–4223.

[88] Efeoglu E, Keating M, McIntyre J, Casey A, Byrne HJ. Determination of nanoparticle localisation within subcellular organelles in vitro using Raman spectroscopy. Anal Methods. 2015;7:10000–10017.

[89] Prieto MCH, Matousek P, Towrie M, Parker AW, Wright M, Ritchie AW, et al. Use of picosecond Kerr-gated Raman spectroscopy to suppress signals from both surface and deep layers in bladder and prostate tissues. J Biomed Opt. 2005;10(4):044006.

[90] Matousek P, Towrie M, Parker AW. Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shift excitation Raman difference techniques. J Raman Spectrosc. 2002;33(4):238–242.

[91] Canetta E, Mazilu M, De Luca AC, Carruthers AE, Dholakia K, Neilson S, et al. Modulated Raman spectroscopy for enhanced identification of bladder tumour cells in urine samples. J Biomed Opt. 2011;16(3):037002.

[92] Afseth NK, Segtnan VH, Wold JP. Raman spectra of biological samples: a study of preprocessing methods. Appl Spectrosc. 2006;60(12):1358–1367.

[93] Gautam R, Vanga S, Ariese F, Umapathy S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. EPJ Techniques and Instrumentation. 2015;2(8):1–38.

[94] Jolliffe IT. Principal Component Analysis. Springer Series in Statistics; 2002.

[95] Davies AMC, Fearn T. Back to basics: the principles of principal component analysis. Tony Davies Column, Spectrosc Eur. 2004;p. 20–23.

[96] Pollard TD, Earnshaw WC, Lippincott-Schwartz J. Cell Biology. Elsevier Health Sciences; 2007.

[97] Plopper G. Principles of Cell Biology. Jones and Bartlett Publishers; 2012.

[98] Rosdahl CB, Kowalski MT. Textbook of Basic Nursing. Lippincott Williams and Wilkins; 2008.

[99] Ruddon RW. Cancer Biology. Oxford University Press; 2007.

[100] Hejmadi M. Introduction to Cancer Biology. Bookboon; 2010.

[101] Weingberg RA. The Biology of Cancer. Garland Science; 2007.

[102] King RJB, Robins MW. Cancer Biology: Third Edition. Pearson Education Ltd.; 2006.

[103] Babjuk M, Burger M, Zigeuner R, Shariat SF, van Rhijn BWG, Comperat E, et al. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013. Eur Urol. 2013;64:639–653.

[104] Vakoc BJ, Fukumura D, Jain RK, Bouma BE. Cancer imaging by optical coherence tomography: preclinical progress and clinical potential. Nat Rev Cancer. 2012;12:363–368.

[105] Palmer GM, Ramanujam N. Diagnosis of breast cancer using optical spectroscopy. Med Laser Appl. 2003;18(3):233–248.

[106] Majeed H, Kandel M, Han K, Luo Z, Virgilia M, Tangella K, et al. Breast cancer diagnosis using spatial light interference microscopy. J Biomed Opt. 2015;20(11):111210.

[107] Shannon AM, Bouchier-Hayes DJ, Condron CM, Toomey D. Tumour hypoxia, chemotherapeutic resistance and hypoxia-related therapies. Cancer treat Rev. 2003;29:297–307.

[108] Gray LH, Conger AD, Ebert M, Hornsey S, Scott OC. Concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy. Br J Radiol. 1953;26:638–648.

[109] Teicher BA. Hypoxia and drug-resistance. Cancer Metastasis Rev. 1994;13:139–168.

[110] Watson JA, Watson CJ, McCrohan AM, Woodfine K, Tosetto M, McDaid J, et al. Generation of an epigenetic signature by chronic hypoxia in prostate cells. Hum Mol Gen. 2009;18(19):3594–3604.

[111] Ewald JA, Desotelle JA, Wilding G, Jarrard DF. Therapy-induced senescence in cancer. J Natl Cancer Inst. 2010;102(20):1536–1546.

[112] Carr B, Malloy A, Warren J. Nanoparticle tracking analysis. NTA Review of Applications and Usage, Nanosight; 2013.

[113] Blackwell RH, Franzen CA, Flanigan RC, Kuo PC, Gupta GN. The untapped potential of urine shed bladder cancer exosomes: biomarkers, signaling, and therapeutics. Bladder. 2014;1(1).

[114] Szajnik M, Derbis M, Lach M, Patalas P, Michalak M, Drzewiecka H, et al. Exosomes in plasma of patients with ovarian carcinoma: potential biomarkers of tumour progression and response to therapy. Gynecol Obstet. 2013;S4.

[115] Taylor DD, Gercel-Taylor C. Exosomes/microvesicles: mediators of cancer-associated immunosuppressive microenvirnoments. Sem Immunopathol. 2011;33:441–454.

[116] Filipazzi P, Burdek M, Villa A, Rivoltini L, Huber V. Recent advances on the role of tumour exosomes in immunosuppression and disease progression. Sem Cancer Biol. 2012;22:342–349.

[117] Hsu DH, Paz P, Villaflor G, Rivas A, Mehta-Damani A, Angevin E, et al. Exosomes as a tumour vaccine: enhancing potency through direct loading of antigenic peptides. J Immunother. 2003;26(5):440–450.

[118] Riches A, Campell E, Borger E, Powis S. Regulation of exosome release from mammary epithelial and breast cancer cells - A new regulatory pathway. Eur J Cancer. 2015;50(5):1025–1034.

[119] King HW, Michael MZ, Gleadle JM. Hypoxic enhancement of exosome release by breast cancer cells. BMC Cancer. 2012;12(421).

[120] Thery C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and function. Nat Rev Immunol. 2002;2(8):569–579.

[121] van der Pol E, Hoekstra AG, Sturk A, Otto C, van Leeuwen TG, Nieuwland R. Optical and non-optical methods for detection and characterization of microparticles and exosomes. J Thromb Haemost. 2010;8:2596–2607.

[122] Tatischeff I, Larquet E, Falcon-Perez JM, Turpin PY, Kruglik SG. Fast characterisation of cell-derived extracellular vesicles by nanoparticles tracking analysis, cyro-electron microscopy and Raman tweezers microspectroscopy. J Extracell Vesicles. 2012;1(19179).

[123] Koss LG, Hoda RS. Koss's Cytology of the Urinary Tract with Histopathologic Correlations. Leopold G. Koss; 2012.

[124] Rosenthal DL, Wojcik EM, Kurtycz DFI. The Paris System for Reporting Urinary Cytology. Springer; 2016.

[125] Cohenford MA, Rigas B. Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: Implications for tumor biology. Proc Natl Acad Sci USA. 1998;95:15327–15332.

[126] Koh CM. Preparation of cells for microscopy using cytospin. Methods Enzymol. 2013;533:235–240.

[127] Severcan F, Haris PI. Vibrational Spectroscopy in Diagnosis and Screening. IOS Press; 2012.

[128] Huang Z, Mc Williams A, Lam S, English J, Mc Lean DI, Lui H, et al. Effect of formalin fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial tissues. Int J Oncol. 2003;23:649–655.

[129] Siegal R, Naishadham D, Jemal A. Cancer Statistics 2013. CA Cancer J Clin. 2013;63:11–30.

[130] Koss GL, Dietch D, Ramanathan R, Sherman AB. Diagnostic value of cytology of voided urine. Acta Cytologica. 1985;29:810–816.

[131] Khochikar MV. Rationale for an early detection program for bladder cancer. Indian J Urol. 2011;27(2):218–225.

[132] Wallace DM, Bryan RT, Dunn JA, Begum G, Bathers S. Delay and survival in bladder cancer. BJU Int. 2002;89:868–878.

[133] Shariat SF, Karam JA, Lotan Y, Karakiewizc PI. Critical evaluation of urinary markers for bladder cancer detection and monitoring. Rev Urol. 2008;10(2):120–135.

[134] Mowatt G, N'Dow J, Vale L, Nabi G, Boachie C, Cook JA, et al. Photodynamic diagnosis of bladder cancer compared with white light cystoscopy: systematic review and meta-analysis. Int J Technol Assess Health Care. 2011;27(1):3–10.

[135] Walsh PC, Vaughan ABRE, Wein AJ, Kaboussi LR, Novick AC, Partin AW, et al. Campbell's Urology. 8th ed. Elsevier; 2002.

[136] Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J Clin. 2011;61(4):212–236.

[137] Heney NM. Natural history of superficial bladder cancer. Prognostic features and long-term disease course. Urol Clin North Am. 1992;19:429–433.

[138] MacVicar D. Carcinoma of the Bladder. Cambridge University Press; 2008.

[139] Crow P, Stone N, Persad RA, Wright MPJ. Optical diagnostics in urology: current applications and future prospects. BJU Int. 2003;92:400–407.

[140] O Faolain E, Hunter MB, Byrne JM, Kelehan P, Lambkin HA, Byrne HJ, et al. Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. J Histochem Cytochem. 2005;53(1):121–129.

[141] Fullwood LM, Griffiths D, Ashton K, Dawson T, Lea RW, Davis C, et al. Effect of substrate choice and tissue type on tissue preparation for spectral histopathology by Raman microspectroscopy. Analyst. 446-454;139:2014.

[142] Stone N, Kendall C, Shepherd N, Crow P, Barr H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. J Raman Spectrosc. 2002;33(7):564–573.

[143] Crow P, Uff JS, Farmer JA, Wright MP, Stone N. The use of Raman spectroscopy to identify and characterize transitional cell carcinoma in vitro. BJU Int. 2004;93:1232–1236.

[144] De Jong BWD, Bakker Schut TC, Maquelin K, van der Kwast T, Bangma CH, Kok DJ, et al. Discrimination between nontumor bladder tissue and tumor by Raman spectroscopy. Anal Chem. 2006;78:7761–7769.

[145] Grimbergen MCM, van Swol CRP, van Moorselaar RJA, Uff J, Mahadevan-Jansen A, Stone N. Raman spectroscopy of bladder tissue in the presence of 5-aminolevulinic acid. J Photochem Photobiol. 2009;95:170–176.

[146] Crow P, Molckovsky A, Stone N, Uff J, Wilson B, Wongkeesong LM. Assessment of fiberoptic near-infrared Raman spectroscopy for diagnosis of bladder and prostate cancer. J Urol. 2005;65:1126–1130.

[147] Grimbergen MCM, van Swol CFP, Draga ROP, van Diest P, Verdaasdonk RM, Stone N, et al. Bladder cancer diagnostics during cystoscopy using Raman spectroscopy. Proc SPIE 7161, Photonic Therapeutics and Diagnostics V. 2009;7161:716114.

[148] Draga ROP, Grimbergen MCM, Vijverberg PLM, van Swol CFP, Jonges TGN, Kummer JA, et al. In vivo bladder cancer diagnosis by high-volume Raman spectroscopy. Anal Chem. 2010;82:5993–5999.

[149] Barman I, Dingari NC, Singh GP, Kumar R, Lang S, Nabi G. Selective sampling using confocal Raman spectroscopy provides enhanced specificity for urinary bladder cancer diagnosis. Anal Bioanal Chem. 2012;404:3091–3099.

[150] O'Toole C, Povey S, Hepburn P, Franks LM. Identity of some human bladder cancer cell lines. Nature. 1983;301(5899):429–430.

[151] Shapiro A, Gofrit ON, Pizov G, Cohen JK, Maier J. Raman molecular imaging: a novel spectroscopic technique for diagnosis of bladder cancer in urine specimens. Eur Urol. 2011;59:106–112.

[152] Tosoni I, Wagner U, Sauter G, Egloff M, Knonagel H, Alund A, et al. Clinical significance of interobserver differences in the staging and grading of superficial bladder cancer. BJU Int. 2000;85:48–53.

[153] Lyng F, O Faolain E, Conroy J, Meade A, Knief P, Duffy B, et al. Vibrational spectroscopy for pathology, from biochemical analysis to diagnostic tool. Exp Mol Pathol. 2007;82:121–129.

[154] Kerr LT, Domijan K, Cullen I, Hennelly BM. Applications of Raman spectroscopy to the urinary bladder for cancer diagnostics. Photon Lasers Med. 2014;3(3):193–224.

[155] Byrne HJ, Baranska M, Puppels GJ, Stone N, Wood B, Gough KM, et al. Spectropathology for the next generation: Quo vadis? Analyst. 2015;140:2066–2073.

[156] Stone N, Prieto MCH, Crow P, Uff J, Ritchie AW. The use of Raman spectroscopy to provide an estimation of the gross biochemistry associated with urological pathologies. Anal Bioanal Chem. 2007;387(5):1657–1668.

[157] Cui L, Butler HJ, Martin-Hirsch PL, Martin FL. Aluminium foil as a potential substrate for ATR-FTIR, transflection FTIR or Raman spectrochemical analysis of biological specimens. Anal Methods. 2016;8:481 – 487.

[158] Mariani MM, Lampen P, Popp J, Wood BR, Deckert V. Impact of fixation on in vitro cell culture lines monitored with Raman spectroscopy. Analyst. 2009;134:1154–1161.

[159] Mikoliunaite L, Rodriguez RD, Sheremet E, Kolchuzhin V, Mehner J, Ramanavicius A, et al. The substrate matters in the Raman spectroscopy analysis of cells. Sci Rep. 2015;5:13150.

[160] Bonnier F, Knief P, Meade A, Dorney J, Bhattacharya K, Lyng F, et al. Collagen matrices as an improved model for an in vitro of live cells using Raman microspectroscopy. Clinical and Biomedical Spectroscopy and Imaging II Proc SPIE. 2011;(80870F).

[161] Meade AD, Lyng FM, Knief P, Byrne HJ. Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vitro cultured human keratinocytes. Anal Bioanal Chem. 2006;5(387):1717–1728.

[162] Takeuchi H, Hashimoto S, Harada I. Simple and efficient method to eliminate spike noise from spectra recorded on charge-coupled device detectors. Appl Spectrosc. 1993;47(1):129–131.

[163] Movasaghi Z, Rehman S, Rehman IU. Raman spectroscopy of biological tissues. Applied Spectroscopy Reviews. 2007;42(5):493–541.

[164] De Gelder J, de Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. J Raman Spectrosc. 2007;38(9):1133–1147.

[165] Cappel UB, Bell IM, Pickard LK. Removing cosmic ray features from Raman map data by a refined nearest neighbour comparison method as a precursor for chemometric analysis. Appl Spectrosc. 2010;64(2):195–200.

[166] Liland KH, Almoy T, Mevik BH. Optimal choice of baseline correction for multivariate calibration of spectra. Appl Spectrosc. 2010;64(9):1007–1016.

[167] Lasch P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemometr Intell Lab. 2012;117:100–114.

[168] Martens H, Stark E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. J Pharm Biomed Anal. 1991;9(8):625–635.

[169] Mann CJ, Yu L, Lo CM, Kim MK. High-resolution quantitative phase-contrast microscopy by digital holography. Optical Express. 2005;13(22):8693–8698.

[170] Rieti S, Manni V, Lisi A, Giuliani L, Sacco D, D'Emilia E, et al. SNOM and AFM microscopy techniques to study the effect of non-ionizing radiation on the morphological and biochemical properties of human keratinocytes cell line (HaCaT). J Microsc. 2004;213(1):20–28.

[171] Passey S, Pellegrin S, Mellor H. Scanning electron microscopy of cell surface morphology. Curr Protoc Cell Biol. 2007;4(17):1–13.

[172] King D, Leiva CR, Glynn M, de la Corta BJD, Cindric S, Kernan D, et al. Living photonics (LiPhos) single cell analysis - label free detection methods. Acta Physiol. 2015;215:32–33.

[173] Sullivan PS, Chan JB, Levin MR, Rao J. Urine cytology and adjunct markers for detection and surveillance of bladder cancer. Am J Transl Res. 2010;2(4):412–440.

[174] Sievert KD, Amend B, Nagele U, Schilling D, Bedke J, Horstmann M, et al. Economic aspects of bladder cancer: what are the benefits and costs? World J Urol. 2009;27(3):295–300.

[175] Van Rhijn BWG, van der Poel HG, van der Kwast TH. Cytology and urinary markers for the diagnosis of bladder cancer. Eur Urol Suppl. 2009;8(7):536–541.

[176] Yutkin V, Nisman B, Pode D. Can urinary biomarkers replace cystoscopic examination in bladder cancer surveillance? Expert Rev Anticanc. 2010;10(6):787–790.

[177] Van Rhijn BW, van der Poel HG, van der Kwast TH. Urine markers for bladder cancer surveillance: a systematic review. Eur Urol. 2005;47(6):736–748.

[178] Kamat AM, Karam JA, Grossman HB, Kader AK, Munsell M, Dinney CP. Prospective trial to identify optimal bladder cancer surveillance protocol: reducing costs while maximizing sensitivity. BJU Int. 2011;108(7):1119–1123.

[179] Ashfaq R. ThinPrep Non-Gyn Morphology Reference Atlas: Urinary Cytology. Cytyc; 2005.

[180] Meade AD, Clarke C, Draux F, Sockalingum GD, Manfait M, Lyng FM, et al. Studies of chemical fixation effects in human cell lines using Raman microspectroscopy. Anal Bioanal Chem. 2010;396:1781–1791.

[181] Kiernan JA. Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do. Micros Today. 2000;00-1:8–12.

[182] Draux F, Gobinet C, Sule-Suso J, Trussardi A, Manfait M, Jeannesson P, et al. Raman spectral imaging of single cancer cells: probing the impact of sample fixation methods. Anal Bioanal Chem. 2010;397:2727–2737.

[183] Pasin E, Josephson DY, Mitra A, Cote RJ, Stein JP. Superficial bladder cancer: An update on etiology, molecular development, classification, and natural history. Rev Urol. 2008;10(1):31–43.

[184] Fischer AH, Jacobson KA, Rose J, Zeller R. Hematoxylin and eosin staining of tissue and cell sections. CSH Protocols. 2008;5:4986.

[185] Lyng FM, Traynor D, Ramos IRM, Bonnier F, Byrne HJ. Raman spectroscopy for screening and diagnosis of cervical cancer. Anal Bioanal Chem. 2015;407(27):8279–89.

[186] Bankpur A, Zachariah E, Cihdangil S, Valiathan M, Mathur D. Raman tweezers spectroscopy of live, single red and white blood cells. PLoS One. 2010;5(4):10427.

[187] Blackwell M, Kang H, Thomas A, Infante P. Formaldehyde: evidence of carcinogenicity. Am Ind Hyg Assoc J. 1981;42(7):A34.

[188] Ahmed HG, Tom MAM. The consequence of delayed fixation on subsequent preservation of urine cells. Oman Med J. 2011;26(1):14–18.

[189] Beleites C, Neugebauer U, Bocklitz T, Krafft C, Popp J. Sample size planning for classification models. Anal Chim Acta. 2013;760:25–33.

[190] Crow P, Barrass B, Kendall C, Hart-Prieto M, Wright M, Persad R, et al. The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. Br J Cancer. 2005;92:2166–2170.

[191] Gazi E, Dwyer J, Gardner P, Ghanbari-Siahkali A, Wade AP, Miyan J, et al. Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study. J Pathol. 2003;201:99–108.

[192] Patel II, Trevisan J, Singh PB, Nicholson CM, Krishnan RKG, Matanhelia SS, et al. Segregation of human prostate tissues classified high-risk (UK) versus low-risk (India) for adenocarcinoma using Fourier-transform infrared or Raman microspectroscopy coupled with discriminant analysis. Anal Bioanal Chem. 2011;401:969–982.

[193] Liu Y, Zhu Y, Li Z. Application of Raman spectroscopy in Andrology: non-invasive analysis of tissue and single cell. Transl Androl Urol. 2014;3(1):125–133.

[194] Harvey TJ, Gazi E, Henderson A, Snook RD, Clarke NW, Brown M, et al. Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy. Analyst. 2009;134:1083–1091.

[195] Garret M, Jassie M. Cytologic examination of post prostatic massage specimens as an aid in diagnosis of carcinoma of the prostate. Acta Cytol. 1976;20(2):126–131.

[196] Fujita K, Pavlovich CP, Netto G, Konishi Y, Isaacs WB, De Marzo A, et al. Specific detection of prostate cancer cells in urine by multiplex immunofluorescence cytology. Hum Pathol. 2009;40(7):924–933.

[197] Chodak GW, Thisted RA, Gerber GS, Johansson JE, Jones GW, Chisholm GD, et al. Results of conservative management of clinically localized prostate cancer. N Engl J Med. 1994;330(4):242–248.

[198] Cookson MS, Herr HW, Zhang ZF, Soloway S, Sogani P, Fair WR. The treated natural history of high risk superficial bladder cancer: 15-year outcome. J Urol. 1997;158(1):62–67.

[199] Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. J Am Board Fam Pract. 2003;16(2):95–101.

[200] Djavan B, Zlotta A, Remzi M, Ghawidel K, Basharkhah A, Schulman CC, et al. Optimal predictors of prostate cancer on repeat prostate biopsy: a prospective study of 1,051 men. J Urol. 2000;163(4):1144–1148.

[201] Hanchanale VS, Rao AR, Das S. Raman spectroscopy and its urological applications. Indian J Urol. 2008;24(4):444–450.

[202] Lyng FM, Ramos IRM, Ibrahim O, Byrne HJ. Vibrational microspectroscopy for cancer screening. Appl Sci. 2015;5:23–35.

[203] Hughes C, Iqbal-Wahid J, Brown M, Shanks JH, Eustace A, Denley H, et al. FTIR microspectroscopy of selected rare diverse sub-variants of carcinoma of the urinary bladder. J Biophotonics. 2013;6(1):73–87.

[204] Masilamani V, AlSalhi MS, Devanesan S, Atif M, Rabah D, Farhat K, et al. A parallelism between spectral grading and Gleason grading of malignant prostate tissues. Photodiagnosis Photodyn Ther. 2013;10(2):168–172.

[205] Harvey TJ, Henderson A, Gazi E, Clarke NW, Brown M, Faria EC, et al. Discrimination of prostate cancer cells by reflection mode FTIR photoacoustic spectroscopy. Analyst. 2007;132:292–295.

[206] Bird B, Romeo MJ, Diem M, Bedrossian K, Laver N, Naber S. Cytology by infrared micro-spectroscopy: Automatic distinction of cell types in urinary cytology. Vib Spectrosc. 2008;48:101–106.

[207] Filik J, Frogley MD, Pijanka JK, Wehbe K, Cinque G. Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. Analyst. 2012;137(4):853–861.

[208] Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discuss. 2004;126:141–157.

[209] Narayan P, Jajodia P, stein R, Ranagho EA. A comparison of fine needle aspiration and core biopsy in diagnosis and preoperative grading of prostate cancer. J Urol. 1989;141(3):56–563.

[210] O'Reilly EA, Gubbins L, Sharma S, Tully R, Guang MHZ, Weiner-Gorzel K, et al. The fate of chemoresistance in triple negative breast cancer (TNBC). BBA Clinical. 2015;3:257–275.

[211] Semenza GL. The hypoxic tumour microenvironment: A driving force for breast cancer progression. Biochim Biophys Acta. 2016;1863(3):382–391.

[212] Hunter BA, Eustace A, Irlam JJ, Valentine HR, Denley H, Oguejiofor KK, et al. Expression of hypoxia-inducible factor-1 predicts benefit from hypoxia modification in invasive bladder cancer. Br J Cancer. 2014;111:437–443.

[213] Galvin DJ, Watson RW, O'Neill A, Coffey RN, Taylor C, Gillespie JI, et al. Hypoxia inhibits human bladder smooth muscle cell proliferation: a potential mechanism of bladder dysfunction. Neurourol Urodyn. 2004;23(4):342–348.

[214] Hoskin PJ, Sibtain A, Daley FM, Saunders MI, Wilson GD. The immunohistochemical assessment of hypoxia, vascularity and proliferation in bladder carcinoma. Radiother Oncol. 2004;72(2):159–168.

[215] Marignol L, Rivera-Figueroa K, Lynch T, Hollywood D. Hypoxia, notch signalling, and prostate cancer. Nat Rev Urol. 2013;10:405–413.

[216] Marignol L, Coffey M, Lawler M, Hollywood D. Hypoxia in prostate cancer: a powerful shield against tumour destruction? Cancer Treat Rev. 2008;34(4):313–327.

[217] Gonzales P, Pisitkun T, Knepper MA. Urinary exosomes: is there a future? Nephrol Dial Transplant. 2008;23:1799–1801.

[218] Nilsson J, Skog J, Nordstrand A, Baranov V, Mincheva-Nilsson L, Breakefield XO, et al. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. Br J Cancer. 2009;100:1603–1607.

[219] Sun X, Niu G, Chan N, Shen B, Chen X. Tumour hypoxia imaging. Molecular Imaging and Biology. 2011;13:399–410.

[220] King HW, Michael MZ, Gleadle JM. Hypoxic enhancement of exosome release by breast cancer cells. BMC Cancer. 2012;12:421.

[221] Edgar JR. Q+A: What are exosomes? BMC Biology. 2016;14(46):1–7.

[222] Tan A, De La Pena H, Seifalian AM. The application of exosomes as a nanoscale cancer vaccine. Int J Nanomedicine. 2010;10(5):889–900.

[223] Jella KK, Rani S, O'Driscoll L, McClean B, Byrne HJ, Lyng FM. Exosomes are involved in mediating radiation induced bystander signaling in human keratinocyte cells. Radiat Res. 2014;181(2):138–145.

[224] Bedard YC, Pollett AF. Breast fine-needle aspiration. A comparison of ThinPrep and conventional smears. Am J Clin Pathol. 1999;111(4):523–527.

[225] Rosa M, Mohammadi A, Masood S. The value of fine needle aspiration biopsy in the diagnosis and prognostic assessment of palpable breast lesions. Diagn Cytopathol. 2012;40(1):26–34.

[226] Krafft C, Popp J. The many facets of Raman spectroscopy for biomedical analysis. Anal Bioanal Chem. 2015;407(3):699–717.

[227] Barkur S, Bankapur A, Pradhan M, Chidangil S, Mathur D, Ladiwala. Probing differentiation in cancer cell lines by single-cell micro-Raman spectroscopy. J Biomed Opt. 2015;20(8):085001.

[228] Kong K, Rowlands CJ, Varma S, Perkins W, Leach IH, Koloydenko AA, et al. Diagnosis of tumours during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. Proc Natl Acad Sci USA. 2013;110(38):15189–15194.

[229] Ali SM, Bonnier F, Lambkin H, Flynn K, McDonagh V, Healy C, et al. A comparison of Raman, FTIR and ATR-FTIR micro spectroscopy for imaging human skin tissue sections. Anal Methods. 2013;9:2281–2291.

[230] Ackerman D, Simon MC. Hypoxia, lipids, and cancer: surviving the harsh tumor microenvironment. Trends Cell Biol. 2014;24(8):472–478.

[231] Iping Petterson IE, Day JC, Fullwood L, Gardner B, Stone N. Characterisation of a fiber optic Raman probe within a hypodermic needle. Anal Bioanal Chem. 2015;407(27):8311–8320.

[232] Almond LM, Hutchings J, Barr H, Day J, Stone N, Kendall C. Development of fiber-optic Raman probes for in vivo diagnosis of upper gastointestinal cancers. Spectrosc Eur. 2011;23(3).

[233] Reid-Nicholson MD, Ramalingam P, Adeagbo B, Cheng N, Peiper SC, Terris MK. The use of Urovysion fluorescence in situ hybridization in the diagnosis and surveillance of non-urothelial carcinoma of the bladder. Mod Pathol. 2009;22:119–127.

[234] Nasuti JF, Gomella LG, Ismial M, Bibbo M. Utility of the BTA stat test kit for bladder cancer screening. Diagn Cytopathol. 1999;21(1):27–29.

[235] Wu SY. Automated Raman cytology system for cancer diagnostics. Masters Thesis, National University of Ireland Maynooth; 2014.