Development of Fourier Transform Infrared Spectroscopy for Drug Response Analysis

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2011

Caryn Siân Hughes

School of Chemical Engineering and Analytical Science

Contents

List of Abbreviations

APS	Ammonium persulfate
CaF_2	Calcium fluoride
dd	Double-distilled
DFT	Discrete Fourier transform
DF(s)	Discriminate function(s)
DMEM	Dulbecco's Modified Eagle's Medium
EMSC	Extended multiplicative signal correction
FACS	Fluorescent activated cell sorting
FCS	Foetal calf serum
FSD	Fourier self-deconvolution
FTIR	Fourier transform infrared (spectroscopy)
FTIR-MS	Fourier transform infrared-microspectroscopy
FTIR-PAS	Fourier transform infrared- photoacoustic spectroscopy
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
H_2O	Water
IR	Infrared
LDA	Linear discriminate analysis
LCMS	Liquid chromatography mass spectrometry
MCT	Mercury cadmium telluride
MDR	Multi drug resistance
MMTS	Methyl methanethiosulphonate
MS	Mass spectrometry
PAS	Photoacoustic spectroscopy
PBS	Phosphate buffered saline (solution)
PC	Principal component
PCA	Principal component analysis
PC-LDA	Principal component- linear discriminate analysis
PCs	Principal components
RCC	Renal cell carcinoma
SNR	Signal-to-noise ratio

SDS	Sodium dodecyl sulphate
TBST	Tris-Buffered Saline and Tween 20 (Polysorbate 20)
TCEP	Tris(2-carboxyethyl)phosphine)
TEAB	Triethyl ammonium bicarbonate
TEMED	Tetramethylethylenediamine
TRIS	Tris(hydroxymethyl)aminomethane

Publications

T. J. Harvey, <u>C. Hughes</u>, E. Correia Faria, E. Gazi, A. Ward, N. W. Clarke, M. D. Brown, R. D. Snook and P. Gardner, "**Raman tweezer analysis of prostate and bladder cell lines**", *ICORS XII conference proceedings* 2008, 1115-1117.

T. J. Harvey, <u>C. Hughes</u>, E. Correia Faria, E. Gazi, A. Ward, A. Henderson, N.W. Clarke, M. D Brown, R. D. Snook and P. Gardner, "**Classification of urological cell lines using Raman optical tweezers**", accepted for publication in the *Journal of Biophotonics* 2009, 2(1-2):47-69.

<u>C. Hughes</u>, M. Liew, A. Sachdeva, P. Bassan, P. Dumas, C.A. Hart, M.D. Brown, N.W. Clarke and P. Gardner, "**SR-FTIR spectroscopy of renal epithelial carcinoma side population cells displaying stem cell-like characteristics**", accepted for publication in *The Analyst* 2010, 135, 3133-3141.

Manuscripts in preparation:
<u>C. Hughes</u>, M. D. Brown, D. Smith, S. Perkins, M. Liew , E. Jackson, P. Bassan, R. D. Snook, N.W. Clarke and Peter Gardner "**A Proteomics-Based Validation of FTIR Bio-Spectroscopy for Discrimination of Renal Carcinoma Cell Lines**"

P. Bassan, A. Sachdeva A. Kohler, <u>C. Hughes</u>, A. Henderson, J. Boyle, M. Brown, N. Clarke and P. Gardner **"FTIR Microscopy of biological cells and tissue: data analysis using RMieS-EMSC correction"**

The University of Manchester

Faculty of Engineering and Physical Sciences

ABSTRACT OF THESIS submitted by Caryn Siân Hughes

For the degree of Doctor of Philosophy and entitled 'Development of Fourier Transform Infrared Spectroscopy for Drug Response Analysis'

Date of submission: 27/09/2011

The feasibility of FTIR-based spectroscopy as a tool to measure cellular response to therapeutics was investigated. Fourier transform mid-infrared spectroscopy has been used in conjunction with multivariate analysis (MVA) to assess the chemistry of many clinically relevant biological materials; however, the technique has not yet found its place in a clinical setting. One issue that has held the technique back is due to the spectral distortions caused by resonant Mie scattering (RMieS), which affects the ability to confidently assign molecular assignments to the spectral signals from biomaterials. In the light of recently improved understanding of RMieS, resulting in a novel correction algorithm, the analytical robustness of corrected FTIR spectra was validated against multi-discipline methods to characterise a set of renal cell lines which were selected for their difference in morphology.

After validation of the FTIR methodology by discriminating different cell lines, the second stage of analyses tested the sensitivity of FTIR technique by determining if discrete chemical differences could be highlighted within a cell population of the same origin. The renal carcinoma cell line 2245R contains a sub-population to contain a sub-population of cells displaying 'stem-cell like' properties. These stem-like cells, however, are difficult to isolate and characterise by conventional '-omic' means.
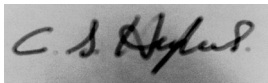
Finally, cellular response to chemotherapeutics was investigated using the established renal cell lines CAKI-2 and A-498. For the model, 5-fluorouracil (5FU), an established chemotherapeutic agent with known mechanisms of action was used.

Novel gold-based therapeutic compounds were also assessed in parallel to determine their efficacy against renal cell carcinoma. The novel compounds displayed initial activity, as the FTIR evidence suggested compounds were able to enter the cells in the first instance, evoking a cellular response. The long-term performance, tracked with standard proliferation assays and FTIR spectroscopy in the renal cancer cell model, however, was poor. Rather than dismissing the compounds as in-active, the compounds may simply be more effective in cancer cell types of a different nature. The FTIR-based evidence provided the means to suggest such a conclusion.

Overall, the initial results suggest that the combination of FTIR and MVA, in the presence of the novel RMieS-EMSC algorithm can detect differences in cellular response to chemotherapeutics. The results were also in-line with complimentary biological-based techniques, demonstrating the powerful potential of the technique as a promising drug screening tool.

Declaration

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

C. S. Hughes 27/09/2011

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

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

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations and in The University's policy on Presentation of Theses.

Acknowledgements

During my PhD I have received the help and support of many people. First and foremost I would like to thank all the people whom I have worked with at the Multidisciplinary Biocentre and the Paterson Institute for Cancer Research, at the University of Manchester. I would also like to thank all of those whom have helped me whilst working at the Synchrotrons in Daresbury, Paris, Trieste and Wisconsin.

In particular I give thanks to my official and un-official supervisors; Dr Peter Gardner, Professor Richard Snook, Dr Mick Brown and Professor Noel Clarke for their endless support, guidance and expertise.

I give thanks to all those who have patiently aided me in training over the course of my PhD such as 'navigating Matlab' and 'cell culturing for non-biologists!'; for this, I thank my fellow colleagues Claire, Matthew, Paul, Thomas, Sarah, Duncan, Alex, Elsa, Sue, Tim and Ehsan. I would also like to thank my other colleagues Frank, Geri, Konrad, Ash, Melody, Adriana and Graeme for their support, advice and collaboration and for having a brilliant 'team-ethic'.

I thank Engineering and Physical Sciences council and the Royal Society of Chemistry for their generous funding of my PhD.

Finally I thank my family for persuading me to take up the rather daunting challenge of studying for a PhD in the first place, as well as providing endless support, encouragement and understanding.

This thesis is dedicated to the loving memory of David Roberts, whom is responsible for igniting my passion for science. Thank you, Granddad.

Chapter 1 │ Introduction

1.1│The Major Classes of Biological Macromolecules

Molecules, such as water, inorganic ions and a large array of small organic molecules such as sugars and fatty acids can be taken up by cells. By a series of chemical reactions cells have the ability to alter and manufacture many small organic molecules. These small organic molecules are pre-cursors to macromolecule synthesis and cells must manufacture large organic molecules via their polymerisation. There are four main classes: proteins, nucleic acids, carbohydrates and lipids [1].

1.1.1│Proteins

Proteins are the most abundant and functionally versatile of the four biological macromolecules in the cell. Proteins are comprised of amino acid sub-units, which are linked together by strong covalent peptide bonds. These linkages many be arranged in a linear polypeptide chain (Eq. 1) [1].

$$(1)$$
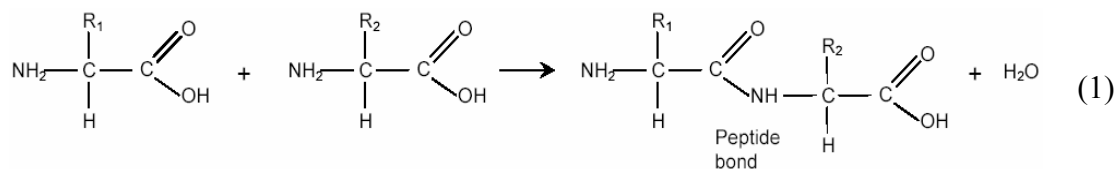
The peptide bonds link amino acid residues within proteins. This sequence of covalently linked amino acids is known as the primary (1°) structure of a protein (Fig. 1.1).

Protein secondary (2°) structure primarily concerns the polypeptide backbone rather than the side chains. Polar side chains are generally arranged on the outside and hydrophobic ones on the inside in a water-soluble protein. The backbone has hydrogen groups (C=O, N-H) capable of hydrogen bonding which, unless satisfied with actual bond formation, would lead to a destabilised protein structure. As the polypeptide backbone crosses back and forth, folding the molecule into a compact shape the backbone is exposed to the hydrophobic protein interior. The non-polar side groups of the amino acids in the protein interior are not capable of hydrogen

bonding, therefore hydrogen bonding occurs with groups on the same, or and adjacent polypeptide backbone. The two main classes of structure which satisfy this arrangement are the α helix (Fig.1.1 2˚(i)), where the backbone is arranged in a spiral-like coil, and the β sheet (Fig.1.1 2˚(ii)), where polypeptide backbones are extended, side by side. The helix and sheet structures are connected with connecting loops known as random coils (Fig.1.1 2˚(iii)) that do not fall into the previous defined categories. Their structure is associated with side chain interactions and may not necessarily satisfy the hydrogen bonding potentials of the polypeptide backbone, therefore such sections are mostly found at the protein exterior in contact with water [1].

Protein tertiary structure describes the arrangement of various secondary structure polypeptides that are folded up into the compact structure of a globular protein (Fig.1.1 3˚), defining the protein molecule. Hydrogen bonding and ionic interactions between side chains influence the folded structure while hydrophobic force is an essential contributor that drives the protein folding. Individual protein molecules in a structure are called subunits. Quaternary structure ((Fig.1.1 4˚) describes the arrangement of subunits into a singular functional complex [1].
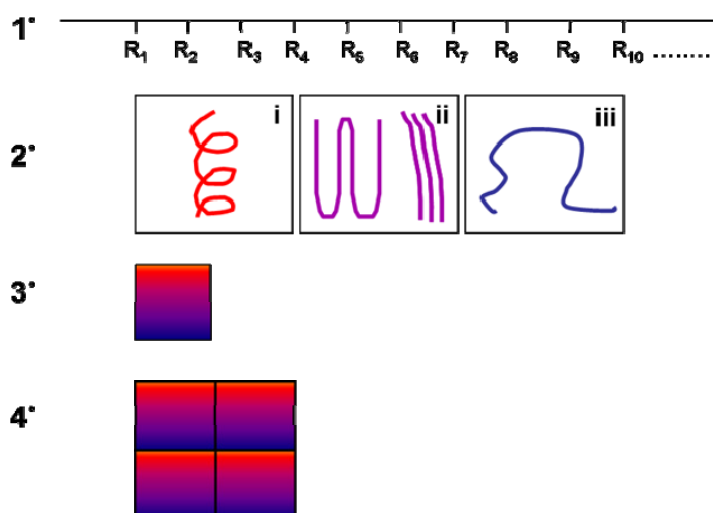
Figure 1.1 │**Diagrammatic illustration of the meaning of the four descriptions of protein structure; primary (1˚), secondary (2˚), tertiary (3˚) and quaternary (4˚). For secondary structure i-iii denotes alpha helix, beta sheet and random coil respectively [1].**

Each type of protein usually has a unique function associated with its particular composition and feature. Roles include biological catalysis, contributing to organism structure, participating in cell signalling, recognition and immunity.

1.1.2 │ Nucleic acids

Nucleic acids consist of two closely related forms; deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Their function involves the storage and conversion of genetic information into proteins. They are comprised of a sugar-phosphate backbone and a series of base pairs which are the informational coding part of the nucleic acid structure. The term nucleic acid was derived from the fact that DNA was first isolated from cell nuclei, and is acidic due to the phosphate groups [1]. DNA is double stranded, comprised of two polynucleotide molecules paired together by hydrogen-bonded complimentary base pairing (Fig. 1.2) [2]. The base pairs in DNA are adenine to thymine and guanine to cytosine (In RNA, thymine is substituted for uracil).
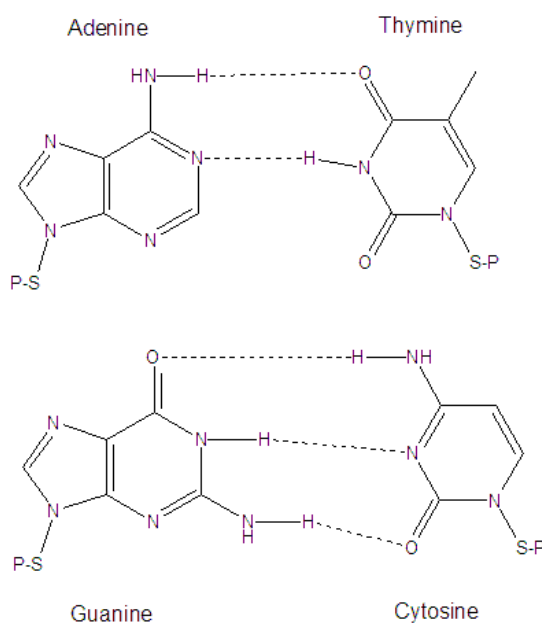
Figure 1.2 │ The Watson and Crick hydrogen-bonded base pairs of DNA, where S-P denotes the sugar-phosphate backbone. Two hydrogen bonds are formed between adenine and thymine, which is situated on the opposite strand. Three hydrogen bonds are formed between guanine and cytosine, which is also on the opposite strand.

The sugar-phosphate backbone is comprised of a series of 2'-deoxyribose molecules (or ribose in RNA) linked together by phosphodiesters (Fig. 1.3). When attached to ribose or deoxyribose, the bases are termed nuleosides and are associated as guanosine, thymidine, adenosine and cytidine [1].
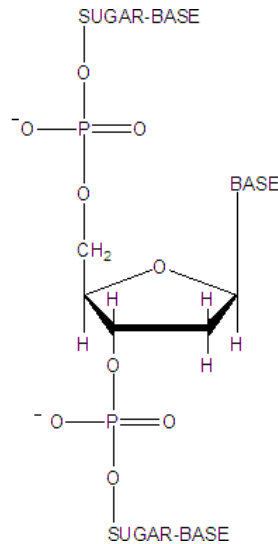
Figure 1.3 │ The sugar-phosphate backbone of DNA: A polymer with an alternating sequence of sugar-phosphate. The deoxyribose sugar is joined at both the 3'-hydroxyl and 5'-hydroxyl groups to phosphate groups in ester links, known as phosphodiester bonds.

1.1.3 │ Lipids

Lipids are the main components of membranes in all cells and can also serve as energy storage molecules. Lipids are neutral fats derived from fatty acids, which have the structure *RCOOH*, where *R* is a long hydrocarbon chain [1].
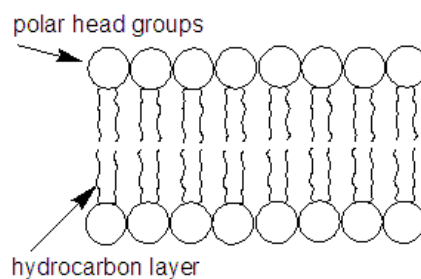
Figure 1.4 │membrane lipid arrangement in the lipid bilayer

Membrane lipids are amphipathic compounds comprising of a polar, hydrophilic head group and a non-polar, hydrophobic tail. These lipids arrange themselves in a lipid bilayer and are the basic structure of all biological membranes. The lipid bilayer is more complex than described in Figure 1.4 as integral proteins, peripheral proteins and some carbohydrate attachments also form part of the membrane structure. Cell membranes play major roles, aside from retaining the cell contents. They transport substances in and out of the cell, they transmit signal from the outside of the cell to the inside, maintain cell shape and are key for cell to cell interactions [1].

1.1.4 │ Carbohydrates

Carbohydrate molecules are the main source of energy for most living organisms. When polymerised, they can function as structural supports, protective coatings and long term energy storage molecules. Glucose, the central ingredient for metabolism activity, is stored in the cell in the form of glycogen and converted during glycogenolysis to glucose molecules when required (Eq. 2). Glycogen is catabolised via interaction with inorganic phosphate from the enzyme glycogen phosphorylase resulting in the cleavage of a glucose monomer to produce glucose-1-phosphate (G1P) [1]. G1P is then converted to glucose-6-phosphate (G6P) by the enzyme phosphoglucomutase. G6P can be further hydrolysed by the enzyme glucose-6-phoshatase to give free glucose to enable the start of the glycolysis [1].

(2)
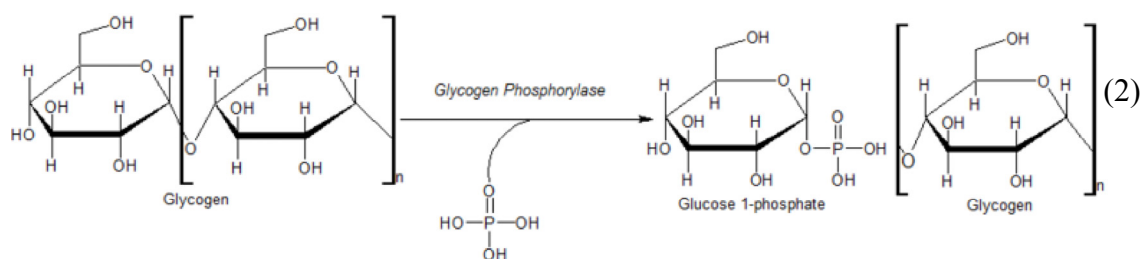
Glycoproteins are proteins having covalently bound carbohydrate. The carbohydrate units are involved in various biological activities, such as influencing protein

folding and conformation. The carbohydrate unit in a transmembrane glycoprotein aids the molecule orientation and anchorage of the molecule in the lipid bilayer [3].

1.2 │ The Molecular Biology of Cancer:

1.2.1 │ Cell cycle

Active, replicating cells pass through a series of stages that are collectively known as the cell cycle (Fig 1.5). There are two gap phases, G/G1 and G2, a synthesis phase, S, and a dividing phase, M.

Interphase describes the processes involved in preparation for mitosis including cell growth and reparation in the G, S and G2 phases. During the processes there are certain check-points that determine whether or not cell progression should proceed.
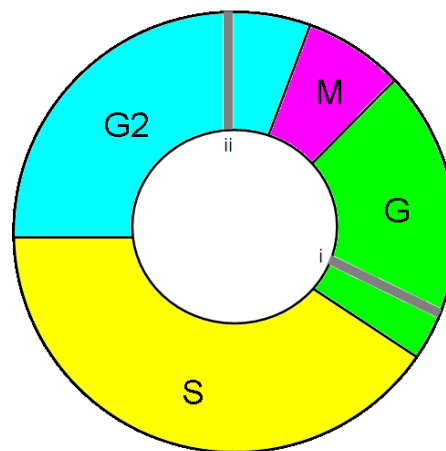
Figure 1.5 │Simplified diagram of cell cycle check points depicting G, S , G2 and M phases. Cell cycle arrest points are (i) if no growth factor signal received and (ii) if DNA damage or incomplete duplication .

The first stage of the cell cycle gap 'G' or 'G1', is the period of time when the cell prepares for DNA synthesis in the 'S' phase. It is during G1 that the decision is made whether or not the cell should proceed to divide. If the decision is positive then the cycle progresses to S phase where DNA synthesis occurs and the genome is duplicated. After the 'S' phase the cell enters another gap, 'G2' where the decision

is again made whether the cell should progress to mitosis. If the cellular processes have not been executed correctly up to this point the cycle is halted. This second check-point is designed to ensure that if DNA is damaged, or incomplete replication has occurred, the process is halted. The consequences for not halting the process would result in progeny cells that may have acquired mutations or have a lethally incomplete compliments of genes. Halting the process ensures time for replication to finish, or DNA repair to occur, or for the cell to undergo programmed cell death [1].

An individual cell itself does not make these decisions can be signalled from extracellular signalling molecules, cytokines and growth factors that are secreted by other cells so that cell division in the tissue is co-ordinated. In brief, growth factors transmit signals to the nucleus via signal transduction pathways resulting in the control of specific genes and their activity. A mitogenic growth factor stimulates the synthesis of proteins necessary for proceeding to S phase [1].

Cell cycle progression is halted if the mitogenic signal is not received. The cells enter a quiescent G0 phase in which its metabolism continues as normal, however, cell division does not occur. In general, cells remain in G0 phase for considerable time before switching to G1 phase to enter the cell cycle.

M phase in the cell cycle accounts for cell division, which is called mitosis in eukaryote cells. Without division, the cell membrane and cell volume will steadily enlarge. As the surface of a spherical organism increases by the square of its radius (r), its volume increases at a rate of (r^3). An increase in cell size (r) produces a relatively smaller increase in surface area (r^2) than in volume (r^3). The inner components have less surface area from which to obtain food, metabolites and oxygen. There is also less surface area to remove waste products and therefore cell death would quickly occur in the absence of cell division [1].

1.2.2 │ Protein kinases

Protein kinases are essential to maintaining control of the cell cycle. Controlling the activation of proteins involves the chemical process of phosphorylation and the reverse, de-phosphorylation. In G1 phase, cycline proteins are synthesised if growth factors or cytokines activate their synthesis. Cyclines enable the activation of protein kinases.

In every cell cycle the cyclin proteins are synthesised in G1 and destroyed in S phase. When the cell enters S phase the cyclins relevant to the G1 phase kinase activity are destroyed. New cyclins are also synthesised to enable progression from S phase into mitosis. These mitotic-related cyclins accumulate during S and G2 phase and combine with specific kinases. The kinase complex resides in the inactive form as the kinase is phosphorylated in three positions. Just before mitosis occurs, two phosphate groups are removed which causes activation of the complex [1].

1.2.3 │ Growth regulation

Cell proliferation is a result of cell growth and cell division, resulting in an increase in the number of cells in a population. In situations where a balance in cell proliferation needs to be met, cells will undergo programmed cell death such as apoptosis, where the cell is given a signal to die. This is regulated by tumour suppressor genes which are a heterogeneous set of genes that inhibit the cancer-related phenotype development in cells [4]. These genes are responsible for health levels of cell growth and differentiation.

Tumour cells are said to have altered versions of these oncogenes, resulting in inactivation or loss the tumour suppressor genes. Inactivation of tumour suppressor genes plays a crucial role in cancer pathogenesis and is just as important as the activation of oncogenes in many cancers.

1.3│Chemotherapeutic Resistance

1.3.1│Multi-drug resistance (MDR) pumps

The study of multidrug resistance (MDR) in tumour cell lines has led to the discovery of the plasma membrane P-glycoprotein (P-gp) molecule [5]. One explanation for high resistance towards drug treatment can be found due to the high level of expression of the plasma glycoprotein 'P-gp'.

It is likely that there are many factors involved that determine the intrinsic MDR phenotype in renal cell carcinoma, however, P-gp expression is certainly an important factor [6].

The protein is a member of the ATP-binding cassette transporter (ABC) super-family. These proteins are ATP-dependent cell membrane pumps that are responsible for the transport of substrates of a broad specificity. They operate as a drug efflux pump for xenobiotic compounds responsible for decreased drug accumulation in multidrug-resistant cells and often mediate the development of resistance to anticancer drugs [7].

1.3.2│Cancer stem cell theory

Adult stem cells, originally described in the haemopoietic system, have been described in many normal solid tissues including bladder, brain, breast, gut, kidney, liver and prostate [9-19]. They are rare by comparison with the number of amplifying and terminally differentiated cells of the organ of origin. Stem cells are required for tissue development and are stimulated to differentiate in response to tissue injury in addition to general replacement and repair. Unlike their differentiated progeny, stem cells are long-lived and present throughout the lifetime of the organ. This longevity makes them susceptible to the accumulation of genetic mutations.

This has given rise to the cancer stem cell hypothesis [19-21], which proposes that within a given tumour there exists a small population of cells with stem-like ability

(Fig. 1.6). They have a capacity to exhibit properties including self renewal and the ability to undergo asymmetric division leading to the generation of progeny that can differentiate and propagate to produce a tumour with a heterogeneous phenotype.

It is hypothesized that cells with stem-cell like properties may be influential in carcinogenesis, possessing the ability to self renew, produce differentiated daughter cells and resist environmental or therapeutic injury. Although there are mixed opinions across different cancer specialties, this has led to a surge in interest in identifying and characterizing the tumour initiating or cancer stem cell (CSC), with the aim of discovering novel diagnostic and prognostic markers [19-21]. The ultimate aim is to generate new therapeutic approaches and possible biomarkers.
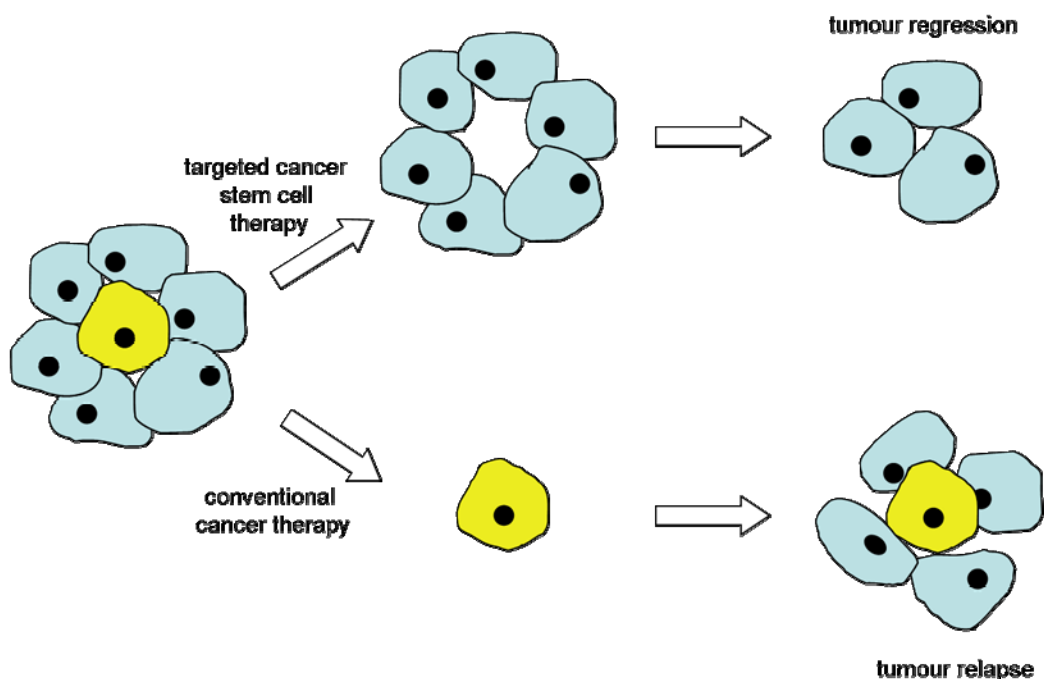
Figure 1.6 │ Targeted therapeutic cancer stem cell treatment proposed to inhibit tumour relapse

Tumours with 'stem-like' cells are more aggressive. They are said to have an over-expression of MDR pumps and therefore have enhanced efflux capabilities of chemotherapy agents. The aim of cancer stem cell targeted therapy is to try and overcome MDR pump activity. This is challenging and difficult to achieve as adult stem cells or putative cancer stem cells are rare and lack of specific stem cell markers. This makes isolation and characterisation difficult [17].

1.4 │ Renal Cell Carcinoma

Renal cell carcinoma (RCC) accounts for 2 – 3% of cancers and the highest rate of occurrence is found in developed countries. RCC is accountable for the most frequently occurring solid lesion within the kidney and there are different types; each with individual histopathological and genetic characteristics. Three significant subtypes of RCC are conventional (clear cell), papillary and chromophobe with prevalence levels of 80-90%, at 10-15% and 4-5% respectively [22-23].

RCC is defined as carcinoma of the renal parenchyma, composed of tubular cells in varying arrangements that originate from the Proximal renal tubule epithelium. This is the lining of the tiny convoluted tubes that filter blood near the Bowman's capsule. It is among the most resistant of tumours to therapy and unlike various other urological malignancies, the disease is currently fatal for over half of patients. The standard treatment involves surgery however 10 – 28 % will demonstrate a local reoccurrence or metastasis after a nephrectomy [22-23].

Many renal tumours are asymptomatic until late in the progression of the disease. Many cases are found incidentally when using non-invasive imaging when evaluating a variety of symptoms and over half of RCCs are detected this way. Some of the most common conditions are hypertension, weight loss, anaemia and abnormal liver function [22-23].

The general recommended system for clinical and scientific classification is the TNM stage system. Histological factors include the Fuhrman nuclear grade which is the most widely accepted histological grading system in RCC, remaining an independent prognostic factor, although subject to intra- and inter- observer discrepancies [22-23].

Metastatic RCC (mRCC) in particular has very poor prognosis and limited treatment options. In chemotherapy, 5-fluroruracil (5FU) is said to be effective, but

only in the presence of immunotherapeutic agents such as Interferon-alpha (IFN-α) [22-23].

Chemotherapy as a monotherapy should not be considered effective in patients with mRCC according to the European Association of Urology (EAU) guidelines. Equally surgical treatment alone is not a likely curative measure in mRCC unless all tumour deposits can be successfully removed. It has been found however that tumour nephrectomy used in combination with immunotherapeutic cytokine IFN-alpha can improve the clinical prognosis [23].

Therapy with similar agents such as interferon-alpha and interleukin 2 was the medical treatment standard for metastatic renal cell carcinoma (RCC) before 2006, although it carried a poor median survival rate of just one year [24].

Today medical treatment for metastatic RCC is a hot topic in the urological oncology field due to an increased understanding of the molecular mechanisms that dictate tumour angiogenesis. Signalling pathways with abnormal function are believed to contribute to the pathogenesis of many malignancies, particularly in terms of renal cancer [25].

The pathogenesis of renal cancers is linked to loss of the von Hippel-Lindau tumour suppressor gene, leading to accumulation of hypoxia-inducible factor 1 (HIF-1) and the subsequent over-expression of HIF-1 target gene products, such as vascular endothelial growth factor. (Activation of mTOR can also lead to an increased expression of HIF-1).

These HIF-1 induced factors are thought to be the key drivers of tumour angiogenesis; leading to the growth and progression of renal cancers [25] and studies have indicated HIF-1 mediates resistance to chemotherapy and radiation [26]. This has led to an increased focus on inhibiting key effectors such mTOR, VEGF and its receptor by multiple targeted therapy. Temsirolimus is a mammalian target of rapamycin (mTOR) protein kinase inhibitor, while inhibitors Sunitinib and Sorafenib target several protein kinases, including the VEGF receptor. According to

research, Temsirolimus should be considered as first-line treatment for low-risk patients while Sunitinib is recommended for first-line therapy for patients with good and intermediate-risk with Sorafenib as a second-line treatment [27]. In fact the EAU recommends that tyrosine kinase inhibitors should be considered as first or second line treatment for mRCC patients as a grade A recommendation [22-23].

However many of the drugs used in targeted therapy are in their early stages of implementation and therefore the combination of surgery and systemic therapies represents the best management option for most patients with metastatic renal cell carcinoma. Many issues such as the use cyto-reductive nephrectomy, metastasectomy, and consolidative surgery will change as systemic therapies become more effective [28].

1.5 │ Aims of Study

The main research focus is to develop Fourier transform infrared spectroscopy as a reliable analytical tool for analysing drug-based responses in chemo-biological material by investigating sample preparation, spectral methods of analysis and data interpretation.

The intention is to lay the groundwork for future use whereby FTIR may play a role in drug-screening processes in the developmental stages of novel therapeutic compounds. In such a complex field, where FTIR spectroscopy is not yet a major player, it is important to maintain a multidisciplinary flavour to the research in order to validate and support findings and equally highlight the limitations.

In the research field of bio-spectroscopy, analysis of cancerous material with FTIR-spectroscopy and multivariate analysis is widely accepted based on the quantity of published research [29-36]. It has yet to be seen, however, if the technique will be implemented as a standard practice in a clinical setting.

Issues of spectral distortion, such as the previously described as 'anomalous dispersion' artefact, has held back the progression of the field as spectral peak

assignments could not be assumed to be correct, as described in chapter 3. Recently, however, greater understanding of the nature of scattering in biological material has been determined and a new algorithm implemented to correct for the phenomena now known as resonant Mie scattering (RMieS) [37-39].

To determine if confidence in FTIR spectral interpretation can now be increased, RMieS-EMSC-corrected biological spectra will be validated using multi-discipline methods to characterise a set of non-established renal cell lines. This part of the research lays the foundation for this project, and is presented in chapter 4.

A second test to prove the capability of FTIR will be to extend characterisation to the next level of difficulty. Characterising the spectral profiles of cell lines that are established from different patients with different properties is a relatively straightforward process. Attempting to characterise a sub-population of cells from the same cell line, however, is somewhat more challenging. The question needs to be addressed if discrete chemical differences can be highlighted within a cell population of the same origin. In-keeping with the cancer stem cell theory, it is hypothesized that 'cancer stem cells' may be influential in carcinogenesis [40]. The 'side population' is a sub-population of cells said to be enriched for stem-like cells, however they are difficult to isolate and characterise by conventional '-omic means'. Characterisation of these cell populations with synchrotron radiation FTIR micro-spectroscopy is attempted in chapter 5.

Finally, the intention is to determine the feasibly of FTIR spectroscopy for cellular response detection to therapeutic agents in the context of renal cell carcinoma. The proposed model is to use 5-fluorouracil. In clinical practice, 5-fluorouracil does not excel in the therapeutic treatment of RCC [21-22], however unlike the more modern therapeutics, it is a well established agent with a wealth of information known about its mechanisms of action [41-42]. It is also an example of an agent that causes cell cycle arrest and is therefore interesting to study in FTIR spectroscopy.

Another arm of the investigation involves novel gold-based therapeutic compounds whereby very little is known about the drug mode of action (Fig. 1.7).
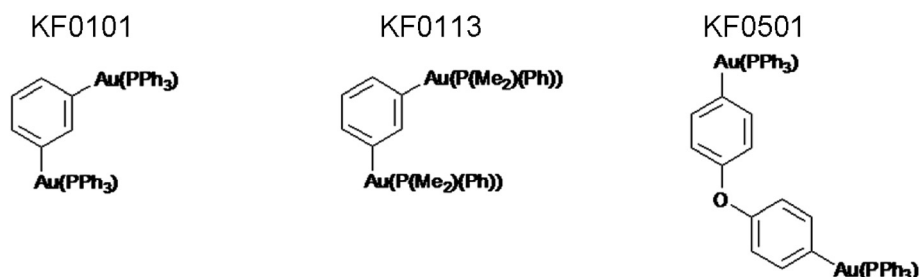
Figure 1.7 │ The di-gold(I) organometallic complex KF0101 was found to exhibit cytotoxic activity in vitro (NCI Panel Testing) against a range of cancer cell types. KF0113 and KF0501 are series analogues.

It is thought that these novel compounds have increased stability next to older agents such as Cisplatin [43]. Renal carcinoma, being one of the most unresponsive cancers to chemotherapy alone, would provide an extremely challenging test for metal based agents to undergo.

In an attempt to determine the feasibility of the technique, Chapter 6 includes initial testing and exploratory analysis of the different methods of FTIR spectroscopy that can be evoked. Chapter 7 seeks to separate some of the complicating spectroscopic factors so that spectral interpretation of purely drug-cell interactions may be accounted for.

1.6 │ References

1. Elliott WH, Elliott, DC: **Biochemistry and Molecular Biology**. *Oxford University Press* 2001: 495-502.

2. Watson JD, Crick, FHC: **Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid**. *Nature* 1953, **171**(4356):737-738

3. Shylaja M, Seshadri HS: **Glycoproteins: An overview**. *Biochemical Education* 1989, **17**(4):170–178.

4. Anderson MLM, Spandidos DA: **Onco-suppressor genes and their involvement in cancer (review).** *Anticancer Research* 1988, **8**(5 A):873-880.

5. Juranka PF, Zastawny RL, Ling V: P-glycoprotein: **Multi-drug resistance and a superfamily of membrane-associated transport proteins**. *FASEB Journal* 1989, **3**(14):2583-2592

6. Naito S, Sakamoto N, Kotoh S, Goto K, Matsumoto T, Kumazawa J. :**Expression of P-glycoprotein and multidrug resistance in renal cell carcinoma.** *Eur Urol.* 1993; **24**(1):156-60.

7. Del Moral RG, Olmo A, Aguilar M, O'Vaue F: **P glycoprotein: A new mechanism to control drug-induced nephrotoxicity**. *Experimental Nephrology* 1998, **6**(2):89-97.

8. Kondo Y, Hamada J, Kobayashi C, Nakamura R, Suzuki Y, Kimata R, Nishimura T, Kitagawa T, Kunimoto M, Imura N et al.: **Over expression of hypoxia-inducible factor-1α in renal and bladder cancer cells increases tumorigenic potency.** *Journal of Urology* 2005, **173**(5):1762-1766.

9. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD, Dirks PB: **Identification of human brain tumour initiating cells.** *Nature* 2004, **432**(7015):396-401.

10. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF: **Prospective identification of tumorigenic breast cancer cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(7):3983-3988.

11. Federici G, Espina V, Liotta L, Edmiston KH: **Breast cancer stem cells: a new target for therapy**. *Oncology* 2011, **25**(1):25-28, 30.

12. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, Wicha M, Clarke MF, Simeone DM: **Identification of pancreatic cancer stem cells.** *Cancer Research* 2007, **67**(3):1030-1037.

13. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R: **Identification and expansion of human colon-cancer-initiating cells.** *Nature* 2007, **445**(7123):111-115.

14. Papailiou J, Bramis KJ, Gazouli M, Theodoropoulos G: **Stem cells in colon cancer. A new era in cancer theory begins.** *International Journal of Colorectal Disease* 2011, **26**(1):1-11.

15. Florek M, Haase M, Marzesco AM, Freund D, Ehninger G, Huttner WB, Corbeil D: **Prominin-1/CD133, a neural and hematopoietic stem cell marker, is expressed in adult human differentiated cells and certain types of kidney cancer.** *Cell and Tissue Research* 2005, **319**(1):15-26.

16. Yin S, Li J, Hu C, Chen X, Yao M, Yan M, Jiang G, Ge C, Xie H, Wan D et al.**: CD133 positive hepatocellular carcinoma cells possess high capacity for tumorigenicity**. *International Journal of Cancer* 2007, **120**(7):1444-1450.

17. Bhatt RI, Brown MD, Hart CA, Gilmore P, Ramani VAC, George NJ, Clarke NW: **Novel Method for the Isolation and Characterisation of the Putative Prostatic Stem Cell**. Cytometry 2003, *54*(2):89-99.

18. Brown MD, Gilmore PE, Hart CA, Samuel JD, Ramani VAC, George NJ, Clarke NW: **Characterization of benign and malignant prostate epithelial Hoechst 33342 side populations**. *Prostate* 2007, **67**(13):1384-1396.

19. Grey BR, Oates JE, Brown MD, Clarke NW: CD133: **A marker of transit amplification rather than stem cell phenotype in the prostate?** *BJU International* 2009, **103**(7):856-858.

20. Rahman M, Deleyrolle L, Vedam-Mai V, Azari H, Abd-El-Barr M, Reynolds BA: **The cancer stem cell hypothesis: Failures and pitfalls.** *Neurosurgery* 2011, **68**(2):531-545.

21. Clayton S, Mousa SA: **Therapeutics formulated to target cancer stem cells: Is it in our future***? Cancer Cell International* 2011, 11:7.

22. Ljungberg B, Hanbury, D.C., Kuczyk, M.A., Merseburger, A.S., Mulders, P.F.A., Patard, J.J., Sinescu, I.C.: **Renal Cell Carcinoma Guideline**. *European Urology* 2007, 51(6):1502-1510.

23. Ljungberg B, Hanbury DC, Kuczyk MA, Merseburger AS, Mulders PFA, Patard JJ, Sinescu IC: **Renal Cell Carcinoma Guideline.** *European Urology* 2007, 51(6):1502-1510.

24. Pouessel D, Culine S, Verhoest G, Patard JJ: **Renal cell carcinoma and antiangiogenic therapies.** *Antiangiogenèse et cancer du rein* 2008, **37**(4 PART 2):628-633.

25. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, Grünwald V, Thompson JA, Figlin RA, Hollaender N et al.: **Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial.** *The Lancet* 2008, **372**(9637):449-456.

26. Semenza GL: **Targeting HIF-1 for cancer therapy.** *Nature Reviews Cancer* 2003, **3**(10):721-732.

27. Radulovic S, Bjelogrlic SK: **Sunitinib, sorafenib and mTOR inhibitors in renal cancer.** *Journal of BUON* 2007, **12**(1):151-162.

28. Thomas AA, Rini BI, Campbell SC: **Integration of surgery and systemic therapy in the management of metastatic renal cancer**. *Current Urology Reports* 2009, **10**(1):35-41.

29. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

30. Braiman MS, Rothschild KJ: **Fourier transform infrared techniques for probing membrane protein structure**. *Annual review of biophysics and biophysical chemistry* 1988, **17**:541-570.

31. Carmona P, Rodriguez-Casado A, Alvarez I, de Miguel E, Toledano A: **FTIR microspectroscopic analysis of the effects of certain drugs on oxidative stress and brain protein structure.** *Biopolymers* 2008, **89**(6):548-554.

32. Gasper R, Mijatovic T, Benard A, Derenne A, Kiss R, Goormaghtigh E: **FTIR spectral signature of the effect of cardiotonic steroids with antitumoral properties on a prostate cancer cell line**. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 2010, **1802**(11):1087-1094.

33. Lamberti A, Sanges C, Arcari P: **FT-IR spectromicroscopy of mammalian cell cultures during necrosis and apoptosis induced by drugs**. *Spectroscop* 2010, **24**(5):535-546.

34. Ahmed GAR, Khorshid FAR, Kumosani TA: **FT-IR spectroscopy as a tool for identification of apoptosis-induced structural changes in A549 cells treated with PM 701**. *International Journal of Nano and Biomaterials* 2009, **2**(1-5):396-408.

35. Mourant JR, Yamada YR, Carpenter S, Dominique LR, Freyer JP: **FTIR spectroscopy demonstrates biochemical differences in mammalian cell cultures at different growth stages.** *Biophysical Journal* 2003, **85**(3):1938-1947.

36. Derenne A, Gasper R, Goormaghtigh E: **The FTIR spectrum of prostate cancer cells allows the classification of anticancer drugs according to their mode of action**. *Analyst* 2011, **136**(6):1134-1141.

37. Bassan P, Byrne HJ, Lee J, Bonnier F, Clarke C, Dumas P, Gazi E, Brown MD, Clarke NW, Gardner P: **Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells**. *Analyst* 2009, **134**(6):1171-1175.

38. 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**(8):1586-1593.

39. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, Gazi E, Brown M, Clarke N, Gardner P: **Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples**. *Analyst* 2010, **135**(2):268-277

40. Addla SK, Brown MD, Hart CA, Ramani VAC, Clarke NW: **Characterization of the Hoechst 33342 side population from normal and malignant human renal epithelial cells.** *American Journal of Physiology - Renal Physiology* 2008, **295**(3).

41. Thomas DM, Zalcberg JR: **5-Fluorouracil: A pharmacological paradigm in the use of cytotoxics.** *Clinical and Experimental Pharmacology and Physiology* 1998, **25**(11):887-895.

42. Longley DB, Harkin DP, Johnston PG: **5-Fluorouracil: Mechanisms of action and clinical strategies.** *Nature Reviews Cancer* 2003, **3**(5):330-338.

43. Flower KR, Khalifa I, Bassan P, Demoulin D, Jackson E, Lockyer NP, McGown AT, Miles P, Vaccari L, Gardner P: **Synchrotron FTIR analysis of drug treated ovarian A2780 cells: An ability to differentiate cell response to different drugs?** *Analyst* 2011, **136**(3):498-507.

Chapter 2 │ Experimental Principals and Methodology

2.1 │ Fundamentals of FTIR Spectroscopy

Fourier transform infrared (FTIR) spectroscopy monitors the vibrational bending and stretching modes of molecules that are resonant with the frequency of infrared. The wavelengths at which these groups absorb the infrared radiation are measured, resulting in a unique spectroscopic fingerprint with relative intensities dependant upon sample composition.

2.1.1 │ Electromagnetic radiation and the photon

Radiation can be described as an electromagnetic sine wave, comprising of both electric and magnetic field components. Electromagnetic interactions result from the exchange of photons as the electromagnetic field is quantised, in quantum mechanical terms. Electromagnetic radiation is classified according to the frequency of its wave in the electromagnetic spectrum [1].

Spectroscopy is defined as the study the interaction of radiation with matter. Radiation is characterised by its energy, *E*. Radiation is emitted and absorbed in the tiny packets called photons. Photons can be described in two ways. Firstly, they can be said to hold particle properties, moving in straight lines at the speed of light. Photons have no detectable 'rest' mass, which enables them to travel at light speed with respect to all observers. Photons do, however, have both energy and momentum and the photon energy will be conserved unless emitted or absorbed by a charged particle. Einstein's equation ($E=MC^2$) cannot describe their energy however as they have no mass. Plank devised a constant to describe the relationship between photon energy and frequency [1].

The energy of the photon particle can be described mathematically (eq. 3) where *h* is Planck's constant (6.626×10^{-34} J s) and ν is the frequency (Hz).

$$E = h\nu \tag{3}$$

According to 'wave-particle duality' in quantum physics, it is natural for the photon to display either particle properties or wave properties, according to its circumstances. Waves have two important characteristics, frequency ν and wavelength λ related by the speed of light c in a vacuum (2.998×10^8 m/s) [1] (eq. 4).

$$\nu = \frac{c}{\lambda} \tag{4}$$

Therefore radiation, characterised by its energy E, is linked to the frequency ν and wavelength λ of the radiation by the Planck relationship, extending equation (2). The energy of the electromagnetic wave is proportional to the frequency, or inversely proportional to the wavelength (eq. 5).

$$E = h\nu = \frac{hc}{\lambda} \tag{5}$$

Vibrational spectroscopy in particular uses the reciprocal wavelength ($1 / \lambda$) which is the number of wavelengths per unit distance, denoted as wavenumber [1]. The wavenumber is usually written in the form (eq. 6):

$$\tilde{\nu}[cm^{-1}] = \frac{10000}{\lambda[cm]} \tag{6}$$

2.1.2 │ Wave propagation and absorption

One way to make the connection between particle and wave descriptions of a photon is to visualise a propagating wave packet.

The speed of a wave depends upon the medium in which it propagates. The propagation of an electromagnetic wave in a non-absorbing medium can be described mathematically as (eq. 7):

$$A(\varphi) = A_0(\varphi)e^{i(\omega t - \delta)} \tag{7}$$

where A is the amplitude, ω is the angular frequency, t is the time, δ is the phase angle and φ is the polarisation angle [1]. The angular frequency is a measure of how fast an object is rotating about its axis and is expressed by the wavelength and refractive index n [1] (eq. 8)

$$\omega = \frac{c}{\lambda n} \tag{8}$$

The refractive index must be modified to its complex form in the case for an absorbing medium (eq. 9).

$$n^* = n + ik \tag{9}$$

Here, the real part of the refractive index n is the ratio between c and the speed at which light travels in a material, known as the phase velocity υ_p (eq. 10).

$$n = \frac{c}{\upsilon_p} \tag{10}$$

When infrared radiation passes through a material, or medium, some intensity is absorbed through interaction with the molecules, and some intensity passes through without interaction. For an absorbing medium, the absorption coefficient α is one of many ways to describe the absorption of electromagnetic waves. α can be expressed in terms of the imaginary part of the refractive index k and the wavelength of light (eq. 11). The imaginary part k indicates the amount of absorption loss when the wave propagates through the material [1].

$$\alpha = \frac{4\pi k}{\lambda} \tag{11}$$

Based on eq. 7, the transmittance of radiation at intensity I emerging through a material is related to the intensity of incident radiation I_0 at the front face of the

material, at a particular wavenumber (eq.12). The Beer Lambert Law results directly from this equation where λ (cm) is the path length of the wave within the absorbing medium (optical thickness) and ε is the molar absorption coefficient and c the concentration (mol dm^{-3}) of the absorbing material (eq. 13) [1].

$$I = I_0 e^{-\varepsilon c l} \tag{12}$$

$$T = \frac{I}{I_0} = 10^{-\varepsilon c l} \tag{13}$$

The proportion of absorbed intensity of radiation over the total intensity that enters the material has a direct relation to the concentration of the material. Absorbance becomes linear with concentration (eqn. 14) [2].

$$A = \log\left(\frac{I_0}{I}\right) = \varepsilon c l \tag{14}$$

Absorbance is the negative logarithm of the transmittance (T) (eq. 15).

$$A = \log\left(\frac{1}{T}\right) = -\log(T) = -\log\left(\frac{I}{I_0}\right) \tag{15}$$

In the Beer-Lambert law, the concentration of an unknown material can be found by using the absorbance at a single wavelength. An absorbance spectrum, however, is a distribution of absorbance intensities for radiation passing through a sample over a series of wavelengths. Absorption varies with wavelength as the absorption coefficient has a different value at each wavelength [2].

When an interferogram is Fourier transformed it results in an output of light intensity at the detector versus the optical frequency. This raw, single beam spectrum contains information about the whole instrument; the sample, the sources, the ambient air, the optical components as well as any possible contamination in the

optical path. As only the information due to the sample itself is of interest, first a reference spectrum called the background must be acquired and used to remove unwanted information from the raw spectrum. The reference spectrum only contains information about the environment in the instrument and all reference spectra for FTIR spectrometers have the same general shape. The use of a reference spectrum is covered further in the next chapter (Ch. 2.2.2.).

To obtain the absorbance spectrum of the sample, the value at each wavelength in the sample spectrum is divided against the corresponding value at the same wavelength in the reference spectrum and the negative logarithm is taken (eq. 16). [2].

$$A = -\log\left(\frac{sample}{reference}\right) \tag{16}$$

2.1.3 │ Molecular vibrations

The absorption of light in a medium causes transitions from an energetic ground state to a particular excited state. Depending on the chemical nature of the interacting compound and the energy of radiation, the excited states may differ.

Molecular vibrations range from the simple motions of two atoms in a diatomic molecule to the motions of every atom in a large polyatomic molecule. Molecules with *N* atoms have *3N* degrees of freedom. *3N-6* degrees of freedom denote the number of ways that the atoms in a non-linear molecule can vibrate (the '*-6*' explains translational and rotational modes about the *x, y* and *z* axes). Each mode involves displacement of the atoms from their equilibrium positions and for each mode, *i,* all the atoms vibrate at a certain characteristic frequency, v_i [2].

Rotations and vibrations are excited in the infrared range and IR spectra result from transitions between quantized vibrational energy states. The energy difference for

transitions between the ground state ($v_i = 0$) and the first excited state ($v_i=1$) of most vibrational modes corresponds to the energy of radiation in the mid-infrared region.

Molecules transition to the excited state only if their dipole moment μ, changes during the vibration. Some molecules will possess certain elements of symmetry such that some vibrational modes will be degenerate or forbidden. Homonuclear diatomic molecules, for example, are not infrared active.

Due to this degeneracy, the number of vibrational modes usually observed is often less than *3N-6*. For a linear molecule, for example, rotation about the axis of the bond does not involve any atom displacement, therefore one of the rotational degrees of freedom is lost. Thus the number of modes of a linear molecules is 3*N*-5 so that a heteronuclear diatomic molecule (N=2) has a single vibrational mode.

At certain frequencies, large atom displacements may only occur for a select number of atoms for many vibrational modes, while neighbouring atoms remain stationary. The frequencies of such modes are characteristic to the functional group in which the motion is centred thus the observation of spectral features in a certain spectral region often indicates a specific chemical functional group of the molecule in question. [2].

Complex organic molecules may involve significant atom displacements involving several atoms, such that the frequency may vary from one molecule to another for a particular functional group. These modes are known as fingerprint bands and are useful to distinguish similar molecules. Every molecule will have slightly different vibrational modes from all other molecules and so the vibrational spectrum of a given molecule is unique and can be used to identify that molecule.

2.2 │Infrared Instrumentation

2.2.1 │Michelson interferometer

Waves can interact with one another, if they come from a similar source or frequency, causing interference effects. Two waves sharing the same frequency with amplitude will interact favourably. Their peaks and troughs will line up and the resultant wave will have amplitude of the summation of the two, known as constructive interference. Two waves interacting out of phase result in destructive interference. One wave's crests will coincide with another wave's troughs and will cancel each other out, resulting in amplitude of zero. The interaction of waves is the key concept of an interferometer [2].

An FTIR spectrometer consists of a globar light source, a two-beam interferometer (comprised of a fixed mirror, an adjustable mirror and a beam splitter) and a detector, such as mercury cadmium telluride (MCT). Modern interferometers are designed based on the original Michelson interferometer (Fig. 2.1). The Michelson interferometer divides the infrared beam into two paths by use of a beamsplitter which is reflected by two mirrors; one of fixed position and one that moves. The beams are subsequently recombined after path differences have been introduced, creating the opportunity for interference to occur between the beams.
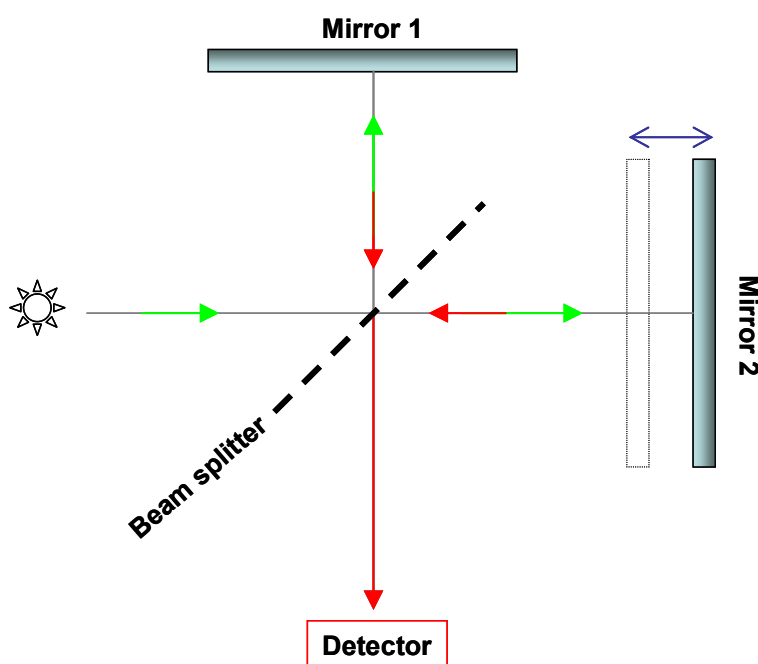
Figure 2.1 │Schematic of a Michelson interferometer: The initial IR beam is split by the beamsplitter. The two split beams move different distances and have different phase delays when recombined, hence causing an interference pattern (an interferogram)

The split beams travel to different distances and hence have different phase delays when recombined. The resulting interferogram records the summation of cosine and sine contributions that display the signal intensity as a function of path length. The interferogram is Fourier-transformed which produces the sample spectrogram [2].

2.2.2 │ The importance of system purging

FTIR systems are always in need of a constant supply of nitrogen or `air that has been filtered through a desiccator to remove moisture. This is because an absorbance sample spectrum is susceptible to water and carbon dioxide contributions. If the carbon dioxide and water vapour levels are too high, absorptions of infrared energy by the gases will result in unwanted peaks that may obscure absorbing sample peaks.

In the vapour phase small molecules display fine structure because both vibrational and rotational transitions can occur simultaneously. This is rarely observed in larger, vapour phase molecules, as the individual rotational transitions are too close together to be resolved. These transitions are also not observed in liquids as the rate of molecular collisions occurs faster than the rotational frequency [2]. The shape of the vibration-rotation spectra of water and carbon dioxide observed in a spectrum acquired with a poorly-purged system are characteristic (Fig 2.2).

In a vibration-rotation spectrum of a diatomic molecule the rotational energy levels are characterised by a single rotational quantum number J. Assuming a rigid rotor (constant bond lengths) $E_J = BJ(J+1)$ where B is the rotational constant [2].

The selection rule for a diatomic molecule in terms of a transition between rotational energy states is $\Delta J = \pm 1$ vibration-rotation spectrum of a rigid diatomic molecule, displaying the singular vibrational mode v_0, results in a series of equally spaced lines above and below the vibrational mode at the R branch; $\Delta J = +1$ and P branch; $\Delta J = -1$ [2].

In reality molecules are not rigid rotors and the length of bond between *X-Y* will increase due to increasing angular velocity from centrifugal forces. Due to this centrifugal distortion the spacing of the lines in the *P* branch increase, and the *R* branch decreases, as the distance from the vibrational mode increases. The spacing of the rotational modes in larger and less symmetric molecules decreases and the spectra become more complex [2].

For carbon dioxide, there is no change in the dipole moment of linear carbon dioxide during symmetric stretching and therefore the vibrational mode is forbidden. The anti-symmetric mode, however, does result in a change in dipole and is therefore allowed. Linear molecules have two equal moments of inertia whilst non linear molecules usually have three different moments of inertia. Further complications arise, however, as linearity is lost in the bending and rotational modes, resulting in the *Q* branch of carbon dioxide near 670 cm^{-1} for example (Fig. 2.2).
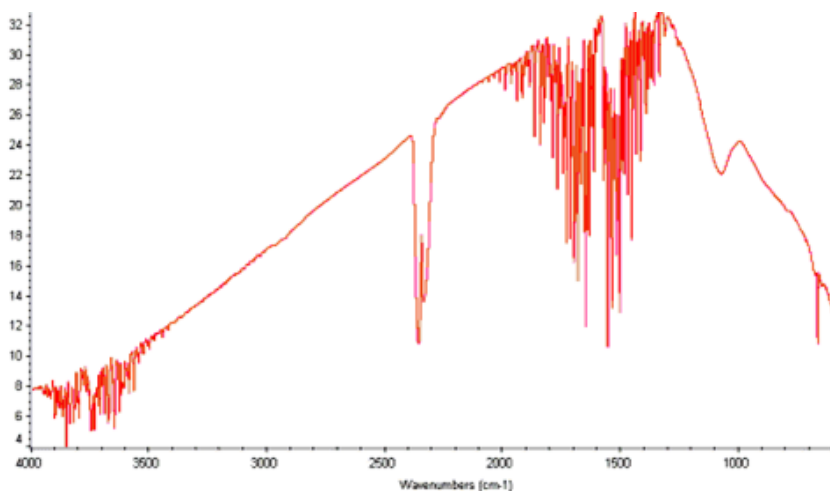
Figure 2.2 │ A typical background spectrum with bands at ~2350 cm^{-1} and 667 cm^{-1} are due to carbon dioxide and bands at ~3500 cm^{-1} and 1630 cm^{-1} are due to water vapour [3].

2.2.3│Basic principles of Fourier transforms

The Fourier transform (FT) changes a signal (or any data) from the time domain to the frequency domain (and back again through the inverse FT) where *f(t)* is the signal in the time domain and $F(\omega)$ is the signal in the frequency domain (eq. 17):

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\partial t \qquad (17)$$

Any signal in the time domain is a sum of numbers at discrete points in time. Instead of a summation, it is common to let the terms approach zero and do an integral to reveal a convolution of the delta function that returns *f(t)* (eq. 18).

$$\sum f(t_n)\partial(t-t_n) \longrightarrow \int f(t_n)\partial(t-t_n) \qquad (18)$$

As opposed to the summing up all these points, wave functions (eq. 19) are used to cover all points in time using Euler's formula.

$$e^{ix} = \cos(x) + i\sin(x) \qquad (19)$$

Wave packets have magnitude *M* with an exponential term e^{iwt}. *Acos(ωt)* is the real part, proportional to cosine with amplitude *A* and imaginary part proportional to the sin with amplitude *Bsin(ωt)*. Any time domain signal can be represented by a sum of all possible combinations of sinusoidal waves with ω_h, frequencies and A_n, B_n amplitudes that stretch over all time at the right magnitude (eq. 20) [2].

$$Me^{i\omega t} = A\cos(\omega t) + iB\sin(\omega t) \qquad (20)$$

In simplistic terms, eq.16 states the amount of signal with frequency (ω) in *f(t)* is equal to (eq. 21):

$$\frac{f(t)}{e^{iwt}} = f(t)e^{-iwt}$$

(21)

This is akin to asking the quantity of *x* in *y*, which would be (*y/x*) or that there is example, there is *f(t)/20Hz* amount of a *20Hz* frequency in a signal. However, the above equation states as *f(t)* is a single number and does not have frequency components. More correctly, the amount of signal $F(\omega)$ is calculated by integrating over all values of *t* (eq. 22):

$$F(\omega) = \frac{f(t)}{e^{iwt}} = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\partial t \qquad (22)$$

For each frequency, the amplitudes of the real and imaginary parts as a function of omega $F(\omega)$. The output from the DFT produces a complex number where the magnitude is the real part of the signal and the phase which is the initial angle of the wave [2].

In Fourier transform spectroscopy, all wavelengths of light enter in parallel, simultaneously producing interference patterns for each one. The multiplex (Fellgett) principle states this as an advantage of reduced measurement time in comparison with a continuous wave spectrometer that observes only a single wavelength at a time [2]. The FT, a decomposition of a sequence of values into components of different frequencies, is often too slow to compute directly from the definition. Modern spectrometers use the fast Fourier transform (FFT), which is an algorithm to compute the same result more quickly [2].

There is another advantage to using over FT spectrometers over continuous wave spectrometers. The Jacquinot advantage states that the signal-to-noise-ratio is higher as the light throughput is higher. There are no requirements for entrance and exit slits as wavelengths are being measured in parallel so the interferometer's output light intensity is almost equal to the input intensity, which makes signal detection easier [2].

2.3│ Methodology Principles

2.3.1 │FTIR microspectroscopy

A beam of IR light first enters the interferometer which modulates the wavelength of light. Infrared microscopes are similar to conventional microscopes, with the exception that the optics are mirrors instead of lenses, allowing the IR beam to pass through the system without being absorbed. The modulated infrared beam is then focused onto the sample by an objective, in conjunction with a visible beam of light for illumination, so that the sample can be viewed. Typically, a camera is attached to the microscope so that an optical image can be obtained. The jaws of the aperture can be adjusted so that only data for the sample region of interest can be acquired

In general, IR microscopes allow for two modes of data collection. In transmission mode, the IR light passes through the sample and below the stage and is collected by a second IR objective that re-collimates the beam and sends it to the detector [4].

If the sample signal is collected in reflectance mode the light reflected back off the sample is collected using the same objective used to focus the beam onto the sample. Highly reflective substrates such as MirrIR can be used, comprising of an infrared-reflective surface coated by a conducting metal. When the thickness of the sample on a metallic substrate is similar to; or thicker than the wavelength of incident radiation, the reflection spectrum is similar to the transmission spectrum of the sample. The beam passes through the sample twice as the beam transmits through the cell and is reflected off the substrate. The mode of collection term is modified to transflection spectroscopy and results in absorbance bands that are approximately twice as great as the corresponding transmission spectrum [2].

2.3.2 │Synchrotron radiation (SR) FTIR

Spatial resolution in infrared microspectroscopy using a standard thermal source can be limited to a diameter of ~20 μm due to high signal to noise ratio (SNR) requirements. This is due to low intrinsic brightness (photon flux or power emitted per source area and solid angle) of the thermal infrared source [5].

Synchrotron radiation (SR) is emitted by a particle when it is accelerated (or decelerated) along the same direction of the particle in motion. Synchrotrons (Fig 2.3) are electron storage rings that use magnetic fields to bend electrons into a closed orbit, producing radiation at each bending magnet. Infrared radiation is generated by these electrons, travelling at relativistic velocities either inside a curved path through a constant magnetic field (bending magnet radiation) or by longitudinal acceleration / deceleration when leaving/entering a magnetic section (edge radiation). The SR beam is then collimated and directed into the interferometer and then to the microscope. The two setups have equivalent flux; the option depends on the engineering preferences at a particular synchrotron [5].
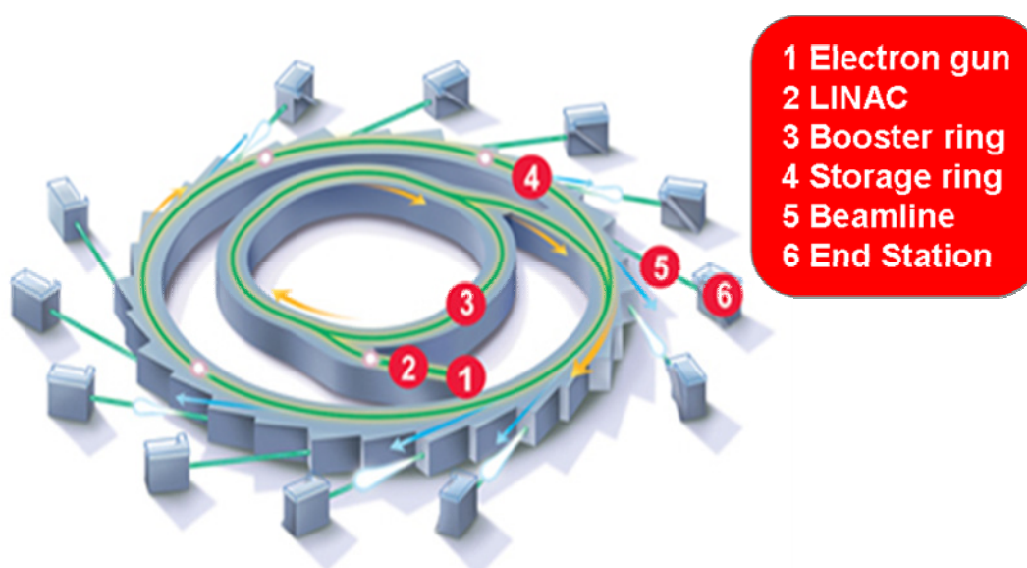
Figure 2.3 │**Simplified schematic of a typical Synchrotron Station comprising of 6 main sections [6].**

There are several advantages to using this IR source for micro-spectroscopy. Very few modifications are needed to adapt commercial instruments to operate with a SR source. Instruments fitted with a SR source can achieve high spatial resolution and excellent spectral quality with a good SNR. For single cell cancer tracking, the use of a synchrotron light source has been implemented to reach the highest spatial resolution, up to 5 x 5 μm, enabling the spectroscopist to monitor subtle changes at the biomolecular level [7-9].

2.3.3 │ Imaging FTIR microspectroscopy

FTIR imaging shares the fundamentals from the conventional technique but provides additional information in the form of spatially resolved spectra, combining digital imaging with the attributes of spectroscopic measurements.

Conventionally, a single element detector is used in FTIR microspectroscopy. Single-point mapping traditionally uses confocal apertures to select the region of interest in a sample. Spectra are measured one point at a time, creating an average signal from the area contained within the range of the selected aperture [2]. The aperture can be reduced to increase spatial resolution, allowing finer sample detail to be observed. Modern, motorised microscope stages are also computer-controlled, enabling single point spectral maps to be acquired with high spatial precision.

The ability to resolve detail, however, is limited by diffraction. Diffraction occurs when a wave encounters an obstacle and can affect all sound, water and electromagnetic types of wave. Diffraction effects cause light of a wavelength larger than the aperture to be lost and therefore any aperture with a finite size will cause diffraction and hence its resolution will be limited [1]. Systems with a focal plane array detector do not use apertures, but an array of detectors. Therefore, diffraction-limited spatial resolution is possible at each wavelength.

In terms of single point mapping, the sample upon the stage has to be moved point by point which is a time consuming process. Array detectors improve acquisition time. Linear array detectors are available, typically consisting of a 16 to 28 pixel array. The sample on the stage, however, still has to be moved line by line. Spectral imaging utilising a focal plane array (FPA) detector, typically consisting of 64×64 or 128×128 elemental pixels, is the most advanced and fastest technique. A two dimensional FPA detector captures the full field of view at each wavelength simultaneously with no movement of the stage. This enables the ability to collect thousands of spectra at one time, creating hyperspectral images [2].

The field of view is related to the spatial resolution, where spatial resolution describes the minimum distance that is required to resolve and detect two adjacent objects. Magnification is a ratio of the physical area of the array detector (number of pixels) divided by the area of the sample field of view; therefore higher magnifications used with the same detector would image a smaller sample area [2].

2.3.4 │ FTIR Photoacoustic spectroscopy (PAS)

The photothermal class of techniques involve an impinging light beam which is absorbed by the sample, altering its thermal state. This thermal state change can be manifested as a change in temperature or sample density, for example.

Photoacoustic spectroscopy is a photothermal technique where modulated infrared radiation is absorbed by a sample in a gas microphone cell and the optical properties of the sample dictate the wavelengths of radiation that are absorbed. This absorption induces a temperature oscillation in the sample at depth x and width dx, proportional to $P_0(1-R)\beta e\text{-}\beta x$ where P_0 is the incident power, R is the fraction of incident light reflected by the sample surface and β (cm^{-1}) is the absorption coefficient of the solid sample for the wavelength λ [10].

The infrared beam is modulated so that the cyclic warming and cooling of the absorbing sample occurs faster than the sample can expand and contract in appropriately. The thermal waves that are generated as a result propagate through the sample and into the gas layer within the photoacoustic cell. The waves decay according to the thermal diffusion length μ_s where α is the thermal diffusivity of the sample and f is the modulation infrared beam frequency (eq. 23). The subsequent pressure wave is detected by the microphone that generates a voltage that is amplified and finally processed by FTIR spectrometer.

$$\mu_s = \sqrt{\left(\frac{\alpha}{\pi f}\right)} \tag{23}$$

The key advantage of photoacoustic spectroscopy is that only the absorbed light is converted to sound thus scattered radiation is not detected [11]. In other vibrational spectroscopic methods, if the wavelength of the IR used to probe the cells is similar in size to the cell nucleus, significant Mie scattering from the nucleus can occur. This can result in distortion of line shapes and observable shifts in peak positions.

2.3.5 │ Cell culture

In mammalian cell culture, the cells are typically cultured in plastic flasks or wells in a suitable environment of culture medium with appropriate nutrients. All renal carcinoma cell lines studied were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 1% L-Glutamine and 10% foetal calf serum (FCS). Both are media supplements that support cell growth of cells that have high energy demands.

Cells are cultured in a humidified atmosphere with 5% carbon dioxide (CO_2) at 37°C to simulate real mammalian physiological CO_2 conditions. The blood in the human body contains a partial pressure of 40 mmHg CO_2, which is close to 5%. In the blood, CO_2 exists in the bicarbonate form (hydrogen carbonate, HCO_3^-), acting as a pH buffer to allow for fluctuations in nutrient, gas and metabolites. Maintaining CO_2 at 5% in cell culture ensures the same amount of CO_2 is dissolved in solution as the bicarbonate. [12]

Depending on the cell type, cells may be suspended in solution during culture, or will attach to the bottom surface of their container, either directly onto the tissue culture plastic, or a substrate such as a glass slide. Adherent cells attach in-vitro similarly to the in-vivo process where adhesion is mediated by cell adhesion transmembrane proteins [13]. Once cells have reached an appropriate density, cells are detached from the substrate by tryptic digestion of the cell adhesion proteins.

2.3.6 │ The use of chemical fixatives

Care must be taken to consider the experiment setup that permits the researcher to obtain the optimal spectra for the investigation in hand. Variations of sample

preparation will have a significant effect on the quality of the spectra. For example, cells can be dried, chemically fixed or probed when live. It is known that the IR spectra of cellular chemical components will change with hydration level. Dehydration will cause an irreversible change in molecular structure of various cell components including cell membranes, ultimately leading cell death caused by an increase in membrane permeability.

Mourant et al. [14] reported the significance of air drying M1 Fibroblast cells and their cancerous equivalent, MR1 Fibroblast cells. In dehydrated cells Protein Amide II shifted with a decreased wavenumber and reduced intensity. Secondly, phosphate absorption bands of RNA and DNA (1050-1150 cm^{-1}) of dried cells were not as clearly defined as those of cells in aqueous media which displayed much narrower bands. However the use of an aqueous media translates to strong water absorption resulting in an unattainable protein amide I (1575- 1710 cm^{-1}) measurement.

For both investigative research and medicinal application, sample fixation is a more practical method, although it is critical to match the method of fixation with the intended analytical technique. Some types of analysis can be incompatible with certain fixation techniques with varying degrees of artefact contribution observed in the spectra.

Formalin fixation at a 4% concentration is a widely utilised fixing agent in the biomedical sector and at low concentrations preserves lipid, phosphate and protein components without significant influence to the cellular infrared spectrum [15-17]. Cross-linking fixation via aldehydes is very effective for proteins and nucleic acids, forming intra- and intermolecular covalent bonds between adjacent amine-containing groups through a Schiff acid-base reaction.

Fixation by alcohol involves dehydration by chemical removal of water from the sample. The potential problems of dehydration are shrinkage and the removal of soluble components from the cells. A cell in a hypertonic environment has a higher concentration of solute in the environment than in the inside of the cell, making the

net flow of water move out of the cell. Ethanol fixation has also been shown to poorly preserve lipid structure [17].

2.3.7 │ Supporting multidiscipline techniques

2.3.7.1│ *Cytotoxicity sulforhodamine B (SRB) assay*

The SRB assay is one of the most used methods for in vitro cytotoxicity screening for determining the IC_{50}, which is the concentration of agent needed to inhibit a given cell process by half. In this case, the process is cell proliferation measured by protein quantification. Sulforhodamine B is a fluorescent dye with uses spanning from laser-induced fluorescence, to the quantification of cellular proteins of cultured cells.

SRB is an aminoxanthene dye with two sulfonic groups which binds to the basic amino acid residues under mild acidic and consequently dissociating under basic conditions. SRB binds to the protein components of cells that have been fixed in trichloroacetic acid (TCA) in a stoichiometric reaction and therefore the quantity of dye extracted from stained cells is directly proportional to the cell mass [18].

The SRB assay has a high level of sensitivity, adaptability to a 96-well plate format (Fig. 2.4) and also a good endpoint stability due to the fixing process, meaning the assay is not only applied in scientific research, but on larger-scale screening, such as at the National Cancer Institute (NCI) [18]. The assay is often compared to that of the tetrazolium dye, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT).

The MTT assay has the advantage of detecting only viable cells whereas the SRB cannot distinguish them. This is because the MTT assay requires cellular activity to reduce the colourless tetazolium dye to its purple coloured formazan dye form. This may mean that IC_{50} values determined may be slightly higher than that of MTT, but this does not detract from the ability of the SRB to detect cytotoxic effects.

The major disadvantage of the MTT assay, however, is that some compounds can directly interfere with MTT reduction without having any effects on cell viability and SRB assays are rarely affected by this interference [18].

Figure 2.4 │ SRB-stained cell protein aggregates in a 96 well plate reader. Cells were seeded at ascending densities to establish the optimum density for the cytotoxicity assays using the renal carcinoma cell lines

2.3.7.2│ Label-free liquid chromatography mass spectrometry (LC/MS)

Liquid chromatography–mass spectrometry (LC/MS) is a powerful tandem technique that combines the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry. In brief, a LC/MS system comprises of a high-performance liquid chromatography column (HPLC) connected to a mass spectrometer through an ionising interface. The HPLC system can practically separate any dissolvable compound, while the mass spectrometer can then ionize the separated solution, providing a molecular weight for each peak component.

Mass tags can be introduced into proteins or peptides metabolically, by chemical means, enzymatically, or by spiked synthetic peptide standards. Most label-based quantification approaches include limitations such as complex sample preparation, the demand for high sample concentration and incomplete labelling [19] Instead, quantitative, label-free approaches have been introduced. [20-21] In contrast, label-free quantification aims to correlate the mass spectrometry signal of peptides, or peptide sequencing events with relative or absolute protein quantities directly [22].

An LC-MS/MS system can fragment the parent ion into a distinctive pattern, separating daughter ions for identification and quantisation [23]. The MS/MS spectrum is comprised of peptide fragments resulting from collision induced dissociation. MS/MS spectral matches use the un-interpreted peaks in a peptide fragment spectrum to match to a theoretical fragment spectrum in a sequence database. In a protein ID experiment multiple peptides are usually found and all of their fragment spectra are used to correlate to a protein. The larger the number of peptides identified, the greater the confidence in the protein correlation.

In order for a sequence database search algorithm to perform peptide and protein ID, a number of constraints are places as the MS/MS data heterogeneity is rich. The process typically initiates with an enzyme specificity constraint, using index searching. Theoretical tryptic peptides are generated and mass lists pre-made from a protein database. Next the parent mass of the intact peptide is matched to a short list of database peptides. The theoretical fragment masses are compared to the experimentally derived fragment spectrum. The peptides are then ranked by how many of the fragment masses match the theoretical fragment masses in the sequence database. If more than one peptide is searched all of the peptides found are correlated to their prospective proteins and the protein with the greatest number of well correlated peptides is usually the most significant hit and most programs include a probability number for the proposed match.

2.3.7.3│ Western blotting

Gel electrophoresis enables analytical separation of small protein quantities, working on the principal that charged molecules migrate in an electric field. A polyacrylamide gel is used as the stationary phase because it is solid and easier to handle. Polymerisation of the gel occurs in situ between two glass plates, made to a recipe of suitable porosity so that the proteins migrate in the electric field appropriately [24]. The porous gel acts as a molecular sieve, where small molecule move faster through the gel and the largest ones are slow, or may not even move at all. The key feature of the gel is the detergent sodium dodecylsuphate (SDS), a hydrophobic tail and a negatively charged sulphate group. The protein sample itself it also dissolved in SDS prior to introduction to the gel and it denatured because of it. The SDS molecules attach to the protein by the hydrophobic tail while the negatively charged SDS swamp the native protein charge. A major advantage is that SDS also solubilises water-insoluble hydrophobic membrane proteins so that they may also be studied. When the SDS-protein complexes are electrophoresed on the SDS gel they all move towards the anode and separation is dependent of the size of the molecules. The result is a band of proteins, visualised by Coomassie or Ponceau blue staining [24].

To make the proteins accessible to antibody detection they are transferred from the gel to a nitrocellulose membrane by electro-blotting, which uses a current to pull proteins from the gel onto the membrane. It is necessary to first block any interaction between the membrane and the antibodies used to detect the target protein as the membrane itself is chosen for its ability to bind protein. Blocking of non-specific binding reduces background noise at the end result and eliminates false positives. This is achieved by the addition of a dilute protein milk solution which attaches to the membrane in all areas where the target proteins have not attached [25].

To detect the target protein a primary antibody and a secondary antibody is required. After non-specific blocking, the primary antibody is incubated with the membrane. Primary antibodies bind with high affinity to their unique target antigen and are generated when a cell culture or host is exposed to the target as part of the

immune response. The secondary antibody is added to recognise the primary antibody and has an enzyme attached to it. When a luminescent substrate is exposed to the enzyme on the secondary antibody, chemiluminescence occurs, proportional to the amount of target protein staining. This is recorded by a CCD camera that quantifies the result in terms of optical density [24].

2.3.7.4│ Fluorescence activated cell sorting (FACS)

FACS is a powerful method used to sort cells at high purity. Individual cells in solution flow through a liquid chamber that passes through one or more laser beams, causing light to scatter. Fluorescent dyes in the cell also emit light at certain frequencies. The way the light bounces off each cell gives information about the cell's physical characteristics. Light bounced off at small angles is called forward scatter. Light bounced off in other directions is called side scatter. Forward and side scatter are used for preliminary identification of cells and to exclude debris and dead cells. Forward Scatter measures approximate cell size, and side scatter measures the granularity of the cell [26].

Cell structure or function can be investigated by fluorescent staining. For example, FACS can be used to study DNA cell content. Propidium iodide (PI) and Hoechst dyes bind to DNA and become fluorescent. Some Hoechst dyes can enter live cells. DNA staining can be used to study the cell division cycle. Relative DNA content shows the proportion of cells in G1, and S and G2M phases. By using a DNA binding dye such as propidium iodide, a DNA Histogram can be obtained, profiling the DNA content thereby identifying each cell cycle phase [27-28].

2.4 │ Experimental Methods and Materials

2.4.1│Cell culture protocol

A cell cryovial, removed from storage in liquid nitrogen, was left to warm up for 15 seconds. There were approximately 3~5 million cells from cryovial contained cells from a confluent T75 flask (A flask which has capacity of 75 cm^2 surface area). The cryovials were warmed in a water bath at 37ºC until the last slither of ice liquefied

in the vial. After transport to a biological safety cabinet the cryovial contents was aspirated and dispensed in a drop wise manner into 15 mL falcon tube containing Dulbecco's Modified Eagle's Medium (DMEM). 1 ml of the media was added to the original cryovial and aspirated and put back into the falcon tube containing cells in media, to ensure the complete removal of any residual cells left in the cryovial. The contents of the tubes were then centrifuged for 1600 rpm for 5 minutes to form a cell pellet at the bottom of the tube. The culture media was removed and the cells were re-suspended cells in 6 mL of fresh, complete culture media (DMEM containing 10 % foetal calf serum (FCS) and 1% L-glutamine).

For sub-culturing with a cell density of 1 in 3 creating a new cell passage twice a week, the cells/media were distributed in 2 mL aliquots into new T75 flasks containing 9 mL of complete culture media and incubated at 37ºC with 5% CO_2.

2.4.2│FTIR sample preparation

2.4.2.1│ Cultured slides

For sample preparation using infrared slides, the substrates (MirrIR or calcium fluoride CaF_2) were sterilised in 70% ethanol (EtOH) for 1 hour and left to dry for 1 hour. The sterile slides were then placed in 6-well plate tissue culture plastic. Culture medium from an 80% cell confluent T75 flask was removed and the cells were washed twice in PBS (10 ml for each wash). Trypsin (2 mL) was then added and the flask was placed back into the incubator at 37˚C for 1 minute. Following this, the flask was removed from the incubator firmly tapped to loosen cells from the bottom of the flask. Using an optical microscope the flask was checked to ensure that the majority of cells have detached from the flask bottom. Culture media (4 mL) was then added so that the FCS contained within the media neutralised the effect of trypsin and no further proteins were digested from the cell surface. The contents of the flask was removed and centrifuged at 1600 rpm to obtain a cell pellet. The media was removed and replaced with fresh media (1 mL). The cell pellet was aspirated before a 10 µL aliquot containing cells and media was removed and mixed with 10 µL of Trypan blue. 10 µL of the mixture was pipetted onto an

improved Neubauer hemocytometer to determine the estimated concentration of viable cells (unstained by Trypan blue) per 1 mL in the sample. The cell density was then adjusted to the required density needed for each well containing the IR slide (typically to achieve a total starting concentration of 1×10^5 cells per well) and incubated at 37ºC with 5% CO_2 for the required period of time. When required, cells were fixed by removing the growth medium, washing twice in PBS and then adding 4% formalin (in PBS) to each well. The samples were left for approximately 40 minutes before removing the cell cultured-slide. The slide was then rinsed in deionised water by dipping in three times with 2 seconds each dip and then left to air-dry.

2.4.2.2│ *Cyto-spinning for single cell experiments*

For single cell work the cells must be proportionally dispersed throughout the sample. Cyto-spinning is means of doing this to achieve spread out single cells. Conveniently, the cells can be fixed in their Falcon centrifuge tube and used at any time. Using cell pellet of predefined density, cells are re-spun in PBS (X2) to remove all traces of media and then suspended in 4% formalin for at least 40 minutes. Fixed cells were spun at 800g for 5 minutes using 200 μl aliquots. After spinning it was important to leave the samples to air dry for 24 hours before washing with de-ionised water, as the centrifugal force may have squashed the cells onto the slide, but washing with water straight away might have loosened them. For particularly small cells that may 'bounce' off the slide during the spinning, a 5 minute spin at 400g then a further spin for 5 min at 800g was performed to ensure firm plating.

2.4.3│ FTIR methodology

2.4.3.1│ *'Bench-top' single point FTIR Spectroscopy*

Single point transflection spectra were taken using a BioRad FTS 7000 equipped with a liquid nitrogen-cooled MCT detector and a KBr beam-splitter, attached to a microscope using WinIRPro Software (Varian Inc., U.S.A.). The cells were viewed on a movable x,y stage using a 15× objective lens. The aperture was opened to a

size of 250 μm^2 to sample a large population of cells. Sample images were ratioed against a background spectrum of a blank MirrIR slide that had been acquired before every new sample spectrum.

2.4.3.2│ *Single cell Synchrotron radiation FTIR spectroscopy (SR-FTIR)*

SR-FTIR single point transmission spectra were acquired on the SMIS beamline at the French Synchrotron Facility Soleil (Saint-Aubin, France) using a Nicolet Continuum XL microscope equipped with an MCT detector. The size of the aperture was adjusted to match the diameter of the cell such that it was fully illuminated, typically 12-16 μm^2. Sample spectra were ratioed against a background spectrum of a blank MirrIR slide that had been acquired before every new sample spectrum.

2.4.3.3│ *Focal plane array (FPA) imaging FTIR spectroscopy*

Transflection-mode FTIR imaging spectroscopy was performed using a Varian 670-IR spectrometer coupled with a Varian 620-IR imaging microscope. The microscope was equipped with a 128 x 128 pixel focal planar array detector and a 10 × eye-piece coupled with a 15 x IR objective lens. Each pixel covered a 5.5 μm x 5.5 μm sampling area; the hyperspectral image covered a total area of 700 μm^2 consisting of 16,384 spectra. The hyperspectral image was collected in the 850-3700 cm^{-1} range, at a resolution of 4 cm^{-1} with 128 co-scans for the sample and background. Sample images were ratioed against a background spectrum of a blank MirrIR slide that had been acquired before every new sample spectrum.

2.4.3.4│ *Photoacoustic spectroscopy (FTIR-PAS)*

Measurements were taken using BioRad Win-IR Pro® software (Varian Inc., USA, formerly BioRad) with an MTEC (Model 300) photoacoustic cell (MTEC Photoacoustics, USA) coupled to a Biorad FTS 6000 spectrometer. Per spectrum, 256 scans were co-added using the triangular apodization function in rapid scan mode with a mirror velocity of 0.32 cm s^{-1}, acquired at a 16 cm^{-1} spectral resolution

and ratioed against a carbon black spectrum. The contribution of the substrate to the spectra was removed by subtracting the spectrum of a blank MirrIR substrate.

2.4.4│Cytotoxic SRB assay

Before cytotoxic SRB assays may be performed, an initial seeding density assay must assess the starting requirements of the assay, which can differ depending on the nature of the cell. Cells were seeded at increasing densities between 0 and 8000 cells/well (Fig. 2.4) for a range of renal cell carcinoma cell lines (2220R, 2245R, 2246R, 2247R, A-498 and Caki-2) .The optical density readings for the experiments were plotted against initial seeding density to determine the optimal starting point which would be just before the log phase of growth. On all cell lines this was determined to be approximately 3000 cells per well (Fig. 2.5).
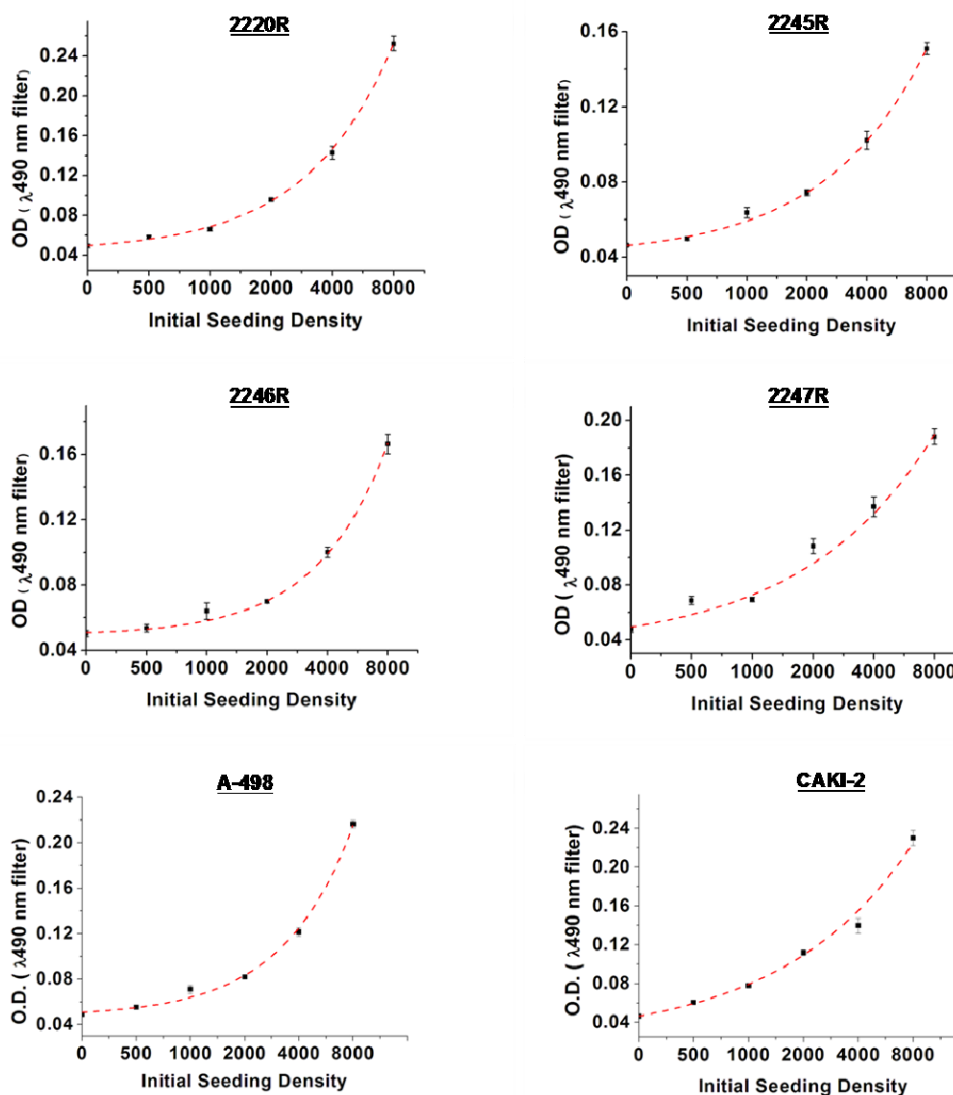
Figure 2.5│ Determination of cell density for the SRB assay using 6 renal cell lines, 2220R, 2245R, 2246R, 2247R, A-498 and CAKI-2. Levels of fluorescence recorded for the SRB stained protein aggregates, excited at a wavelength of 490 nm. An appropriate density at the start of exponential growth at 3000 cells per well was selected for all renal cell lines used.

To deduce inhibitory concentrations (IC) for the drugs against the renal cell lines Caki-2 and A-498, the widely used sulforhodamine B (SRB) cytotoxicity assay was implemented. Cells were seeded (3000 cells in 100 μl medium per well) in triplicates and left for 24 hours to ensure cells were well-established and adhered to the bottom of the well plate. A range of 12 drug concentrations between 0.15 and 20.0 μM were used by the addition of drug/media (100 μL at twice the desired

concentration) to the established cells in media (100 µL) giving the correct final concentration range. Each concentration point was performed in triplicate. For the control group, media only was added (100 µL). The well plates were incubated at 37°C for 3 days in 5% CO_2. After 3 days the media was discarded and wells washed with PBS (X2) (100 µL). The remaining cells were fixed with TCA (100 µL; 10% w/v double distilled (dd) H_2O), refrigerating at 4˚C for 1 hour. The wells were washed with PBS(X2) (100 µL) and dried at room temperature (RT) for 1-2 hours. The cells were stained with SRB (100 µL; 0.4% w/v dd H_2O) and left for 15 minutes at RT. The plates were then washed with acetic acid (1% w/v dd H_2O) until the acid ran clear, then dried at RT for 1-2 hours. The protein residues were then re-suspended with Tris-HCl (100 µL; 1.5M; pH 8.8) and agitated on shaker for 5 minutes before reading on a Versamax plate reader (Molecular Devices) at 490 nm absorbance. Each concentration point was repeated in triplicate as was each experiment to obtain mean absorbance values.

2.4.5│Label-free liquid chromatography mass spectrometry (LC/MS)

The LCMS sample preparation and instrument operation was conducted by the Biological Mass Spectrometry Facility, Paterson Institute for Cancer Research, University of Manchester.

For each cell line, cells were grown to confluence in three T75 flasks for cell lysis. The adherent cells were treated with four PBS washes (25 ml) and aspirated. Cells were then scraped into a solution of triethyl ammonium bicarbonate (TEAB) buffer (500 µl; 500 mM), sodium dodecyl sulphate (SDS) (0.05%) and PPase inhibitors I and II (both at 1%, Sigma Aldrich) and stored on ice for 5 minutes. The cell lysates were spun (4˚C; 1600 g) for 10 minutes and the supernatant transferred to a fresh tube. Cell lysates were separated into aliquots and stored at -80˚C.

Protein concentration was measured using 2D Quant Kit (GE Healthcare) as per manufactures instructions. In brief, the protein lysates are pelleted by centrifugation

and re-suspended in an alkaline solution of cupric ions, which bind to the polypeptide backbones of any present protein. A colorimetric agent is then added which reacts to unbound cupric ions. The colour density is inversely related to the protein concentration of the sample. Protein concentrations were calculated for the cell lines as 2245R [0.905 μg/μl]; 2246R [1.329 μg/μl] and 2247R [2.211 μg/μl]. Samples were scaled to a 20 μg reaction with a dilute of TEAB (500 μl; 500 mM), SDS (0.05%) and PPase (protein phosphatase) inhibitors I and II (both at 1%). The reducing agent tris(2-carboxyethyl)phosphine) (TCEP) (2 μl) was added used to break disulphide bonds within and between the proteins. The protein samples were then incubated for 60 minutes at 60˚C.

After allowing the samples to cool, proteins were alkylated with the cysteine blocker methyl methanethiosulphonate (MMTS) and incubated for 15 minutes at room temperature. Blocking any free cysteine residues prevents formation of new disulphide bonds. Tryptic digestion was performed at an enzyme: substrate ratio of 1:10 (2 μg trypsin to 20 μg protein). Sequencing grade trypsin (20 μg) was re-suspended in TEAB (1 ml, 500 mM) and 100 μl of this stock was added to each protein sample vial, resulting in a final reaction volume of 125 μl with an SDS concentration of 0.01% in the digest. The samples were incubated at 37˚C with gentle agitation for 24 hours before quenching the digest reaction by addition of formic acid (475 μl, 1%). The final solution therefore contained approximately 20 μg digest on 600 μl [33 ng/μl]. Each LCMS injection at 15 μl equated to approximately 500 ng per injection.

Peptides from each sample were analysed a total of 5 times. The inlet method of choice was a Nano-Acquity ultra performance liquid chromatography (UPLC) system (Waters). 500 ng of peptide was loaded onto a Waters C18 Symmetry trap column (180μm ID, 5 μm, 5 cm) in water, 0.1% (v/v) acetonitrile, 0.1% (v/v) formic acid at a flow rate of 7 μl/min. Peptides were then separated using a Waters NanoAcquity bridged ethylene hybrid BEH C18 column ((75 um ID, 1.7 um, 25 cm) with a gradient of 1 to 25% (v/v) of acetonitrile, 0.1% formic acid at a flow rate of 400 nl/min.

The nano-liquid chromatography (nLC) effluent was sprayed directly into the LTQ-Orbitrap XL mass spectrometer aided by the Proxeon nano source at a voltage offset of 1.8 KV. The mass spectrometer was operated in parallel data dependent mode where the MS survey scan was performed at a nominal resolution of 60, 000 (at m/z 400) resolution in the Orbitrap analyser between m/z range of 400-2000 (target value of 1 million ions, maximum fill time of 500 ms). Dynamic exclusion was enabled to prevent the selection of a formally targeted ion for a total of x seconds. nLCMS data was acquired for 50,000 multiply charged 'features, 10,000 of which MSMS data was also acquired.

All the nLCMSMS data was imported into and processed by Progenesis LCMS vx (Nonlinear Dynamics). Progenesis generated peak lists files were submitted to Mascot vy (Matrix science) for a database search using the following criteria. Database was Uniprot vx restricted to human with a precursor tolerance of 20 ppm, fragment tolerance 0.9Da, enzyme specificity trypsin, allowed missed cleavages. The false discovery rate was calculated with the use of decoy search. Progenesis LCMS calculated the relative abundances of all the multiply charged features detected. The most significant features were selected that had a fold change >2.47 and an ANOVA value (P) = $2.85e^{-5}$ with Power ≥ 0.9995.

2.4.6│ Western blotting

The Western blot was conducted by the Genito-Urinary Cancer Research Group at the School of Cancer and Enabling Sciences in the Patterson Institute for Cancer Research, University of Manchester.

In preparation a 13.5% electrophoresis resolving gel was prepared for sodium dodecyl sulphate polyacrylamide gel electrophoreises (SDS-PAGE). The following were components were added into the blot holder: Milipore water (2.6 mL), tris (1.5M; pH 8.8; 1 mL), acrylamide (4.1 mL), SDS (10%; 185 μL), ammonium persulfate (APS) (10%; 36 μL) and tetramethylethylenediamine (TEMED) (18 μL). Ethanol (2 mL) was added on top of the gel to remove air bubbles and to prevent

oxygen from getting into the gel which could oxidise it and inhibit polymerisation. The gel was left for 10 minutes to polymerise, after which the ethanol layer was removed and the gel was washed with Milipore water.

The stacking gel comprised of Milipore water (2.37 mL), tris(hydroxymethyl)aminomethane (0.5M; pH 6.8; 1 mL), acrylamide (0.59 mL), SDS (10%; 34 μL), ammonium persulfate APS (10%; 16 μL) and tetramethylethylenediamine (TEMED) (8 μL) was loaded into the resolving gel in the blot holder, and a comb was inserted to create the protein loading wells. The gel was left to polymerise for a further 10 minutes. The gel plates were connected to the Western blot apparatus and the electrophoresis buffer tris-buffered saline and Tween 20 (TBST) was pored into the chambers to allow for electric current.

Sample preparation: For protein reduction, ß-mercapto-ethanol (BME) (20 μL) was added to a ×4 loading buffer (80 μL) comprised of tris(hydroxymethyl) aminomethane (0.5M; pH 6.8; 25 ml); glycerol (20 ml); SDS (10%; 160 mL), Milipore water (14 mL) and bromophenol blue (40 mg). To each of the cell line, 10 μl, 8 μl and 5 μl of BME in loading buffer was added to the protein lysates at 27.2 μg, 13.7 μg and 6.7 μg respectively; a total of 9 samples to load. The samples were boiled for 5 minutes before being microcentrifuged (30 seconds at 8000 g), and loaded into the gel wells. The gel was run at 60 V for 5 minutes and then 130 V for 1.5 hours.

Wet gel transfer: The SDS-PAGE gel was stacked next to a nitrocellulose membrane, centred in between two layers of filter paper and two sponges, all submerged in cooled transfer buffer comprised of TRIS base (3.03 g), glycine (14.4 g), methanol (200 mL) and Milipore water (800 mL). To remove air bubbles a pipette was used as a roller. The sandwich was then submerged in the transfer tank so that the proteins migrated from the gel towards to the nitrocellulose membrane at the positive electrode (taking 1.5 hours at 100V). At the end of the transfer Ponceau blue staining was used to confirm the presence of proteins on the membrane. The membrane was kept in a container filled with TBST (a mixture of tris-buffered

saline and TWEEN 20) comprised of NaCl (8 g), TRIS base (3 g; pH 7.4), 0.1% TWEEN 20 (Polysorbate 20 detergent) (1 mL).

Antibody Staining: A 0.3% membrane blocking milk (0.15 g Marvel in 50 mL TBST) was used with a SNAP ID system under manufacturer protocol (Milipore). For the primary staining, 10 μl of Histone H2A Antibody (Cell Signalling technologies) was used in 3 mL milk and 2 μl in 5 mL milk for the anti-rabbit secondary antibody solution. A glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody was used as a control to confirm protein loading quantities (4 μl in 3 mL for primary; 2 μl in 5 mL for anti-rabbit secondary antibody). To obtain the blot images, the membranes were treated with chemiluminescence developer supplied with the SNAP ID kit and exposed in a BioSpectrum Imaging System (UVP) for 5 minutes.

2.4.7│ FACS sorting

A-498 cells were seeded at 5×10^5 cells in Falcon (T25) flasks. After 24 hours, agents KF0101, KF0113, KF0501 and 5FU were added at IC_{25} and IC_{50} and were left for 24 and 72 subsequent hours. For the control group no agent was added to the cells. After the appropriate time the media from each flask was removed, containing dead cells. The viable cells remaining adhered to the flask were washed with PBS. The cells were trypsinised, removed and centrifuged. The cell pellet was re-suspended in the smallest possible volume of PBS and agitated. PBS (1 mL) was added to the cell suspension, transferred to Eppendorf tubes and spun (800g; 5 min). The supernatant was aspirated and the Eppendorf agitated to break up the pellet. Ice cold EtOH(1 mL; 70% in H_2O) was added and the Eppendorf vortexed to prevent cell clumping. The cells were fixed for 30 min at RT. Cells were adjusted to approximately 5×10^5 cells/ mL and washed twice in PBS (in 1% BSA). Cells were spun (1000g; 5 min) and re-suspended in staining buffer (1 mL) comprised of propidium iodide (50 mg/mL), Triton–X (0.1 % v/v), PBS and ribonuclease A (50 mg/mL). Cells were incubated at RT for 20 minutes, protected from light. The cells were sorted using an Influx FACS analyser. Samples included cells treated with

5FU and KF1 at IC_{50} concentrations for 24 hours and 72 hours after agent introduction. Untreated cells were also sorted as a control for both time periods.

2.6│References

1. Gauglitz G, Vo-Ding, T, (ed.): **Handbook of Spectroscopy**. *WILEY-VCH* 2003, Ch.3-5.

2. Griffiths PR, de Haseth, J.A.: **Fourier Transform Infrared Spectrometry** 2nd edition: *Wiley*; 2007:Ch. 2,4,5,7,10

3. **Hebrew University of Jerusalem, faculty of agriculture, food and environment**, *http://departments.agri.huji.ac.il/zabam/FTIR.html*. Image accessed March 2011.

4. Miller LM: **Infrared Microspectroscopy and Imaging**. *http://www.nsls.bnl.gov/newsroom/publications/otherpubs/imaging/worksho pmillerhighres.pdf*, accessed March 2011

5. Dumas P, Tobin MJ: **A bright source for infrared microspectroscopy: Synchrotron radiation**. *Spectroscopy Europe* 2003, **15**(6):17-23.

6. *http://www.odec.ca/projects/2005/shar5a0/public_html/how_does_a_synchr otron_work.htm*. Image accessed January 2011

7. Gazi E, Gardner P, Lockyer NP, Hart CA, Brown MD, Clarke NW: **Direct evidence of lipid translocation between adipocytes and prostate cancer cells with imaging FTIR microspectroscopy**. *Journal of Lipid Research* 2007, **48**(8):1846-1856

8. Gazi E: **Applications of Fourier Transform Infrared Microspectroscopy and Time of Flight Secondary Ion Mass Spectrometry in Studies of Prostate Cancer Tissue and Cell Lines**. *The University of Manchester* Institute of Science and Technology; 2004.

9. Gazi E, Dwyer J, Lockyer NP, Miyan J, Gardner P, Hart CA, Brown MD, Clarke NW: **A study of cytokinetic and motile prostate cancer cells using synchrotron-based FTIR microspectroscopic imaging**. *Vibrational Spectroscopy* 2005, **38**(1-2):193-201.

10. Michaelian KH: **Photoacoustic Infrared Spectroscopy**: *John Wiley & Sons*; 2006: Ch. 1,2

11. Rosencwaig A, Gersho A: **Theory of the photoacoustic effect with solids**. *Journal of Applied Physics* 1976, **47**(1):64-69.

12. **Study Package for Medical College Entrance Examinations**, *Tata McGraw-Hill*, 2007:5,20

13. Schwartz MA: **Signaling by Integrins: Implications for Tumorigenesis**, *Cancer Research* 1993, **5(**1);1503-1506

14. Mourant JR, Gibson RR, Johnson TM, Carpenter S, Short KW, Yamada YR, Freyer JP: **Methods for measuring the infrared spectra of biological cells**. *Physics in Medicine and Biology* 2003, **48**:243-257.

15. Gazi E, Dwyer J, Lockyer NP, Gardner P, Shanks JH, Roulson J, Hart CA, Clarke NW, Brown MD: **Biomolecular profiling of metastatic prostate cancer cells in bone marrow tissue using FTIR microspectroscopy: A pilot study**. *Analytical and Bioanalytical Chemistry* 2007, **387**(5):1621-1631.

16. Hopwood D: **The effect of formaldehyde fixation and dehydration on ox adrenal medulla with respect to the chromaffin reaction and post-chroming.** *Histochemistry and Cell Biology* 1967, **10**(1):98-106.

17. Gazi E, Dwyer J, Lockyer NP, Miyan J, Gardner P, Hart C, Brown M, Clarke NW: **Fixation protocols for subcellular imaging by synchrotron-based Fourier transform infrared microspectroscopy**. *Biopolymers* 2005, **77**(1):18-30.

18. Vichai V, Kirtikara, K.: **Sulforhodamine B colorimeric Assay for Cytotoxicity Screening**.: *Nature Protocols* 2006, vol. **1**;

19. Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, Scrivens JH: **A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches**, *Journal of Proteome Research* 2009, **8**, 3752–3759

20. Wang, W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH: **Quantification of proteins and metabolites by mass spectrometry without isotopic labelling or spiked standards**, *Anal. Chem.* 2003, **75**, 4818−4826

21. Radulovic D, Jelveh S, Ryu S, Hamilton TG: **Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry**, *Molecular & Cellular Proteomics* 2004, 3, 984-997.

22. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B: **Quantitative mass spectrometry in proteomics: A critical review.** *Analytical and Bioanalytical Chemistry* 2007, **389**(4):1017-1031.

23. McMaster MC: **LC/MS: a practical users's guide**, *John Wiley & Sons Inc*. 2005,**1**.

24. Elliott WH, Elliott, DC, **Biochemistry and Molecular Biology**, *Oxford University Press* 2001: 495-502.

25. Luttmann W, Bratke K, Kupper, M, **Immunology**, *Elsevier Inc*. 2006, Ch.5

26. Radbruch A, **Flow cytometry and cell sorting**, *Springer-Verlag,* 2000, 1, Ch. 1

27. Krishan A: **Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining**. *Journal of Cell Biology* 1975, **66**(1):188-193.

28. Poulin NM, Matthews JB, Skov KA, Palcic B: **Effects of fixation method on image cytometric measurement of DNA content and distribution in cells stained for fluorescence with propidium iodide**. *Journal of Histochemistry and Cytochemistry* 1994, **42**(8):1149-1156.

Chapter 3 │ Spectral Pre-processing and Analysis

3.1 │ Spectral Pre-processing

3.1.1 │ The influence of light scattering

Incident IR light with intensity I_0 may be partially reflected I_R, scattered I_S and absorbed I_A upon interaction with a sample. The remainder of light will be transmitted I_T. According to the law of conservation of energy, the energy balance (eq. 23) must be:

$$I_0 = I_A + I_R + I_T + I_S \qquad (23)$$
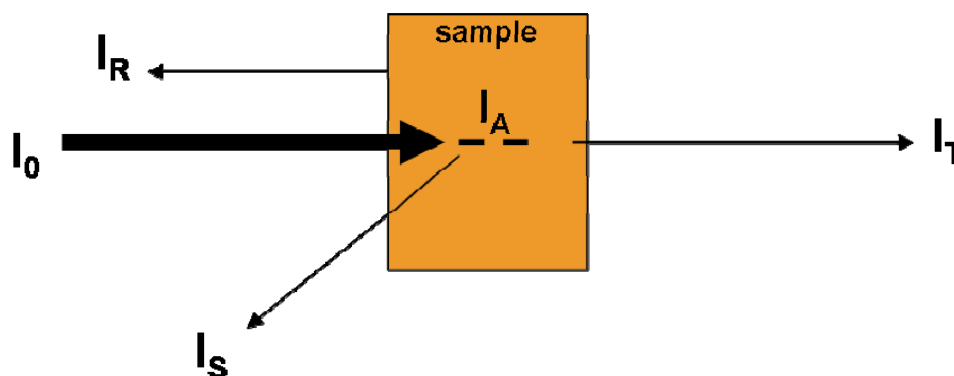
Figure 3.1 │ The interaction of incident light and the sample [44]

To evaluate I_A, ideally, I_0 and one of either I_S, I_T, or I_R is measured, with zero contribution from the other intensities.

Light scattering is the redirection of light that occurs when the incident light (an electromagnetic (EM) wave) encounters a non-homogeneous obstacle, i.e. the scattering sample. Light scattering, however, is not as straight-forward at Figure 3.1 suggests as it not simply a case of electromagnetic waves bouncing off the surface of the sample [1].

The EM wave will interact with the particle; periodically perturbing the electron orbits within the particles constituent molecules at the same frequency as the electric field (of the incident light wave). Electron cloud perturbation results in an

induced dipole moment, due to periodic charge separation within the molecule. The oscillating dipole moment causes the molecule to re-radiate the light in all directions. The majority of light scattered by the particle emits at the same incident frequency, which is termed elastic scattering (Figure 3.2).
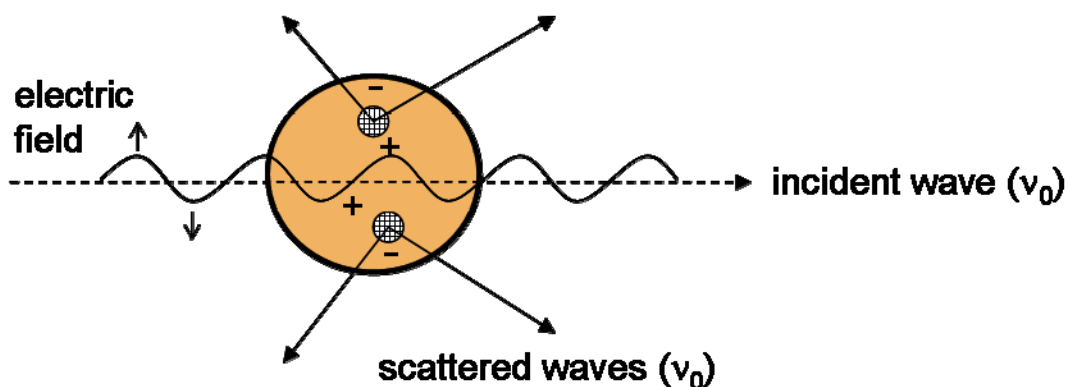
Figure 3.2 │**The interaction of the incident light as an electromagnetic wave, inducing temporary dipole moments in the electron clouds of the sample molecules. As a result scattered light is emitted at the same frequency as the light of incidence [45].**

Rayleigh and Mie scattering describe two forms of elastic light scattering. Rayleigh scattering concerns small spherical particles much smaller in diameter than the wavelength of the incidence and scattered wave and usually exhibit as gaseous, aerosol form in the nanometre range [1].

Mie scattering, however, applies to a general theory of the scattering of electromagnetic radiation by homogenous spherical particles. The formulae and calculations are more complex than those of Rayleigh scattering for both absorbing and non-absorbing spherical particles with no boundary on particle size. (Rayleigh scattering is simply a limiting case of Mie theory for particles much smaller than the incident wavelength).

The Van de Hulst approximation for Mie scattering efficiency is descriptive of a non-absorbing spherical particle [2]. Biological samples, however, are absorbing materials and can exhibit severe scattering. In recent advances, the approximation

was modified to describe absorbing materials, coining the term RMieS (Resonant Mie Scattering).

As Mie scattering depends non-linearly on α, the product of the radius and the imaginary part of the refractive index of the scatterer [3] (eq. 11) the approximation was modified by using a refractive index spectrum of a reference, rather than using a constant value.

The mid infrared range studied for biological materials is typically ~4000–700 cm^{-1}, equating to wavelengths of 2.5–14 μm. Incidentally this range is of similar magnitude to a biological cell and its cellular components [4].

Aside from the refractive index, there are a number of significant properties that affect the severity of the scattering sample. Firstly, scattering is more severe when the scattering sample is closer in diameter to the wavelength of light. Another factor is sample uniformity. A flat surface will not give rise to scattering and IR substrates are polished to high standards for this reason. An uneven surface will cause some of the incident light to be scattered in such a direction that it will not reach the detector, resulting in a 'false' absorption profile [5].

Three types of biological material were considered comprising of a microtomed tissue specimen, an adhesive cultured cell and cell that has been detached by trypsin and cytospun onto an infrared substrate. The spectrum of an isolated cell that has been dropped or deposited onto a slide such that is has no protein attachments to the substrate will present the most severe RMieS scattering. Not only is the cell smallest in diameter, closest to the wavelength of incident light, it is relatively spherical (Fig. 3.3c). The cultured cell will have extended across the surface of the substrate, appearing larger and relatively more uniform than the previous cell (Fig. 3.3b). A tissue specimen will be larger and more uniform, displaying the least amount of RMieS scattering. A uniform sample layer, such as a film of Matrigel, however, will not display scattering (Fig. 3.3a).
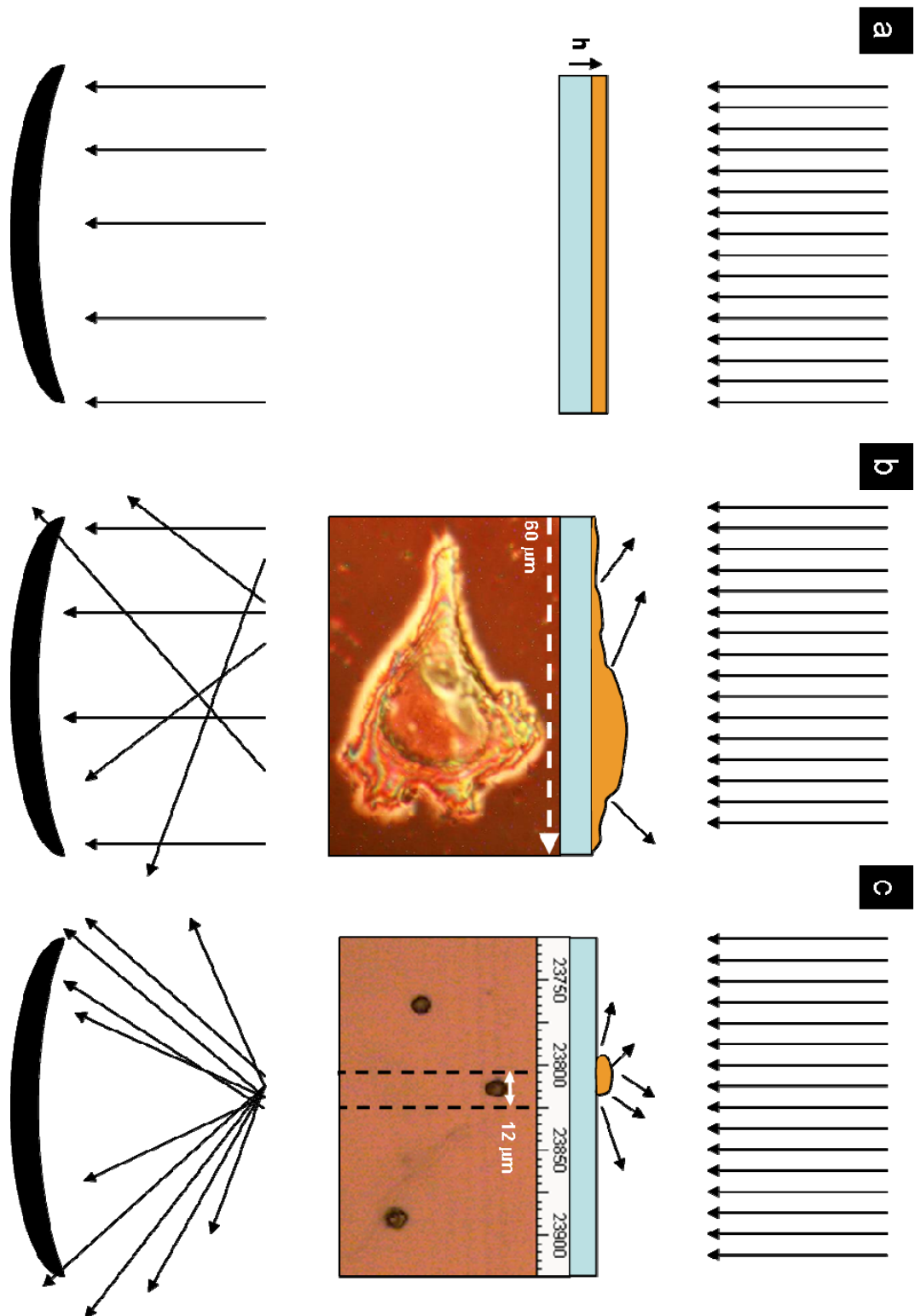
Figure 3.3 │ Schematic diagrams illustrating the occurrence of RMieS scattered light emitted from biological sample and reaching the detector. A matrigel film of uniform thickness will display no scattering (a). An adhesive cell will be fairly larger than the incident wavelength of light but will not be as uniform. (c) A single cell that has been spun onto a substrate will be most spherical and closest to the incident wavelength of light and will therefore exhibit the most severe scattering perturbations within its infrared spectrum.

The consequences of Mie scattering in an affected spectrum include broad oscillating baselines, shifted peak positions and distorted peak shapes. The oscillating baseline is a remnant of non-resonant Mie scattering, which the Extended Multiplicative Signal Correction (EMSC) algorithm developed by Kohler et al. successfully removes [3].
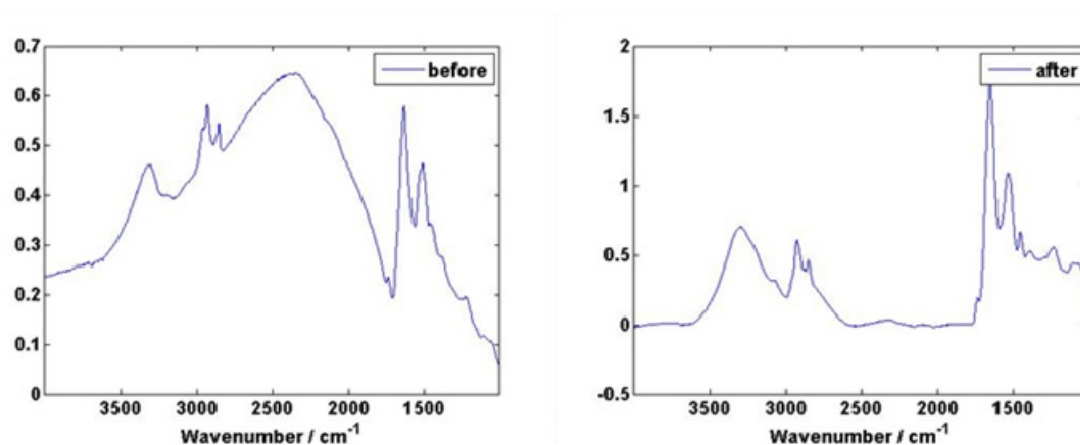
Figure 3.4 │ An example of a highly RMieS scattering single cell, acquired at the SMIS beamline at Soleil, Paris. The spectrum has been corrected with the iterative version of the RMieS-EMSC algorithm with 100 iterations. The cell was ~12 μm in diameter and had been cytospun onto CaF$_2$.

Another notable Mie distortion can be observed as a classic negative dip near the amide I region of the fingerprint at ~1750 cm^{-1}, formerly known as the 'dispersion artefact'. This artefact is now understood to be a consequence of absorbing Mie scattering materials, such as biological samples. The RMieS-EMSC correction algorithm combines the corrective benefits of EMSC and additionally removes the 'anomalous dispersion artefact' to a considerable degree (Fig. 3.4); if not completely in some circumstances. This, however, can only be determined with simulated data, where the 'true answer' for the sample spectrum is known.

The RMieS-EMSC algorithm works by linear regression with the use of an appropriate reference spectrum, such as a uniform layer of Matrigel, which works as the next best solution to a reference of a pure absorbance spectrum. The initial RMieS corrected spectrum will exhibit spectral information from the reference and so improvements are made by iteratively looping the RMieS corrected output and

feeding that into the algorithm such that the new corrected spectrum becomes the new reference spectrum and the original raw sample spectrum is then evaluated.

3.1.2 │ Spectral derivatives

Derivative spectra, obtained by differentiating spectral intensity with respect to wavelength, have a number of uses.

Flat baselines can be achieved as baseline drift is removed with the use of spectral derivatives. Narrower bandwidths i.e. higher resolution of differential spectra potentially allow for subtle differences between spectra to be more easily resolved. To resolve overlapping bands, mathematical derivates are used to narrow their full width at half height value (FWHH). First-order derivative spectra measure the change in the peak slope, with the absorbance peak maxima coinciding with zero on the y-axis of the derivative spectra.

Odd derivatives typically have a dispersive character whereas the even derivates have a symmetrical band shape, parallel to the original band [1]. Therefore the second derivative can be more useful for spectral interpretation as the peak frequency identical to the original peak frequency.

The drawback of implementing derivatives is a cost in the SNR. Noise is higher than in non-derivative spectra and becomes progressively worse as the derivative order increases, as observed in Figure 3.5.
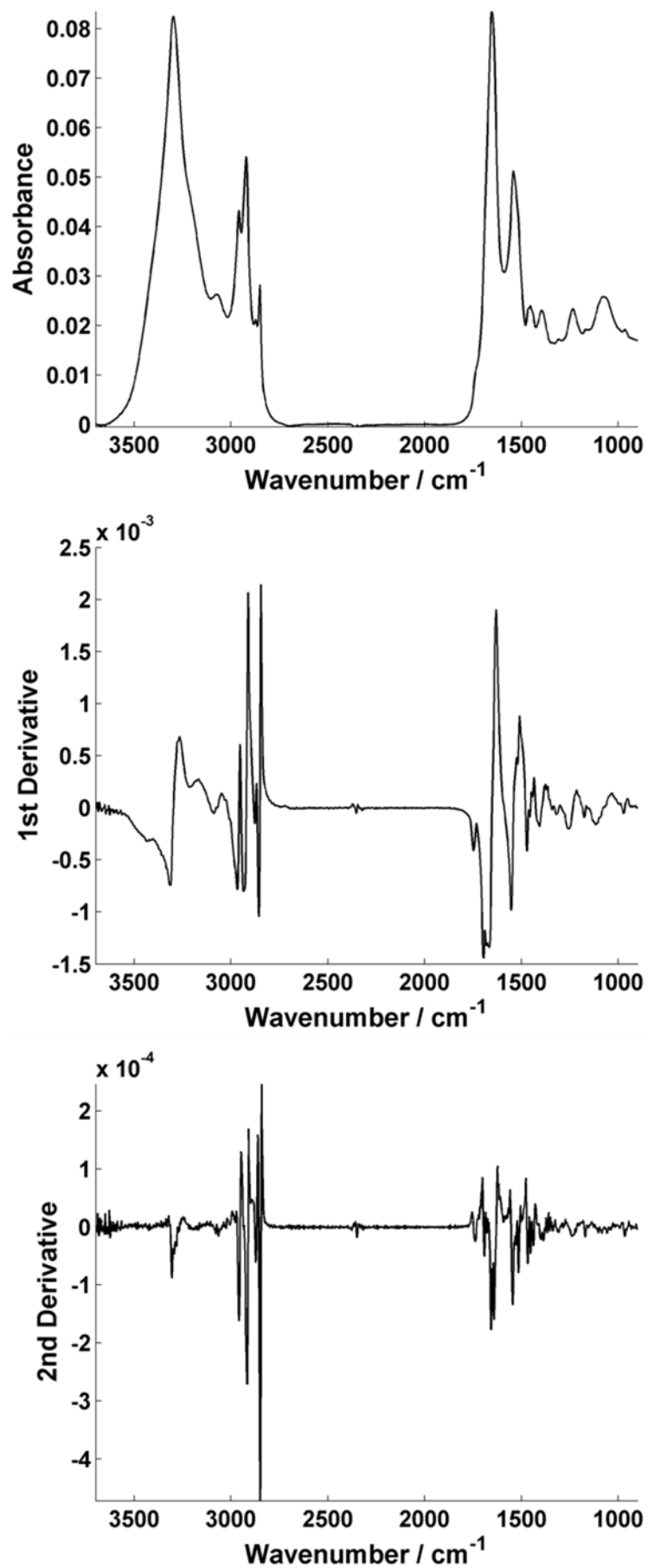
Figure 3.5 **A typical infrared spectrum from a biological cell (upper) and its first (middle) and second derivative transformations (lower).**

3.1.4 │ Fourier self deconvolution

Fourier self deconvolution (FSD) is another mathematical procedure that allows the width of spectral bands to be narrowed. It is a deconvolution approach as the inherent line shape function of each band in the spectrum is altered. FSD aims to reduce the bandwidth without affecting positional or area. The FWHH can often be reduced by as much as a factor of 2 without introducing excessive noise [1]. Over-deconvolution can occur, however, whereby the line width has narrowed to such an extent that the line shape adopts sidelobes. A contribution of sidelobes from two adjacent peaks can lead to a mis-identification of a new peak. FSD can be a powerful tool under correct use. Conversely, it can also result in misleading interpretations.

3.1.5 │ Vector normalisation

It is fairly common place in chemometric analysis of FTIR spectra to normalize the derivatives during the pre-processing step before analysis to remove thickness differences between sample spectra.

To carry out vector normalisation, the average value of the absorbance is calculated for each spectral region. This value is then subtracted from the spectrum such that a new average value equals zero. Finally, the spectra are scaled such, that the sum squared deviation over the indicated wavelengths equals one [6].

3.2 │ Spectral Signals of Cancer Bio-molecules

Fourier transform infrared spectroscopy monitors the vibrational bending and stretching modes of molecules that are resonant with the frequency of infrared. The wavelengths at which these groups absorb the infrared radiation are measured, resulting in a unique spectroscopic fingerprint with relative intensities dependant upon sample composition. FTIR is an attractive option for biomedical applications. Sample preparation is minimal, simple, reproducible and non-destructive to the material and usually, depending on the material type, only small quantities are required, such as micron-thick tissue specimens for example.

Absorbance peaks in vibrational spectra correspond to different molecule bonds, ratios or conformations of bio-molecules. Differences in the chemical environment of these bonds can cause changes in shape, position or intensity of these peaks. Spectral analysis of biomaterial is heavily based on the interpretation of spectral changes associated with spectral features that dominate the infrared signal at specific wavenumbers and Figure 3.6 displays the characteristic peaks from protein, lipid, nucleic acid and carbohydrate signals. Underlying contributions from bio-groups can also contribute to the signal at particular wavenumbers. To analyse underlying bands, spectral processing can be performed in an attempt resolve these peaks (Ch. 3.1.3/4).

The infrared amide I band can be useful for predicting protein secondary structure. The amide I stretching carbonyl mode has been determined to have specific frequencies associates with α-helices, β-sheets, and other characteristic structures such as turns and bends and random coils. The approximate wave numbers corresponding to the three common structures found in proteins are: α-helix, 1650 cm^{-1}; β-sheet, 1632 cm^{-1} and 1685 cm^{-1}; and random coil at 1658 cm^{-1} [7].

In a review by Movasaghi et al. [8], it is apparent that spectral interpretation of comparable areas of bio-material reveals a good agreement in wavenumber assignments, despite the use of different sample preparation techniques, suggesting a certain level of confidence of peak position.
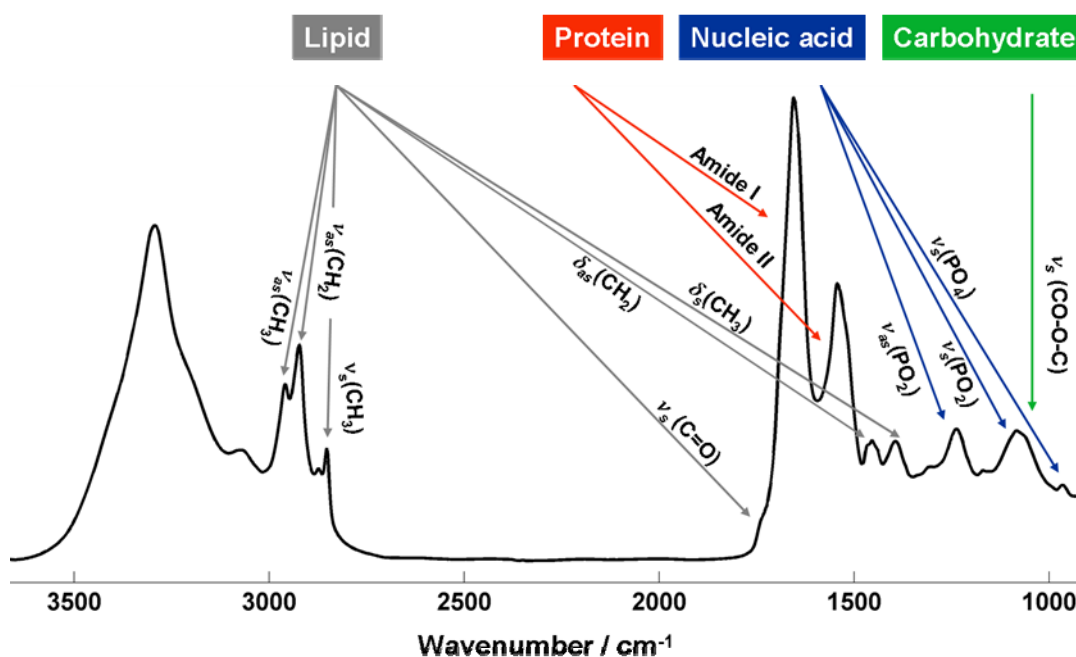
Figure 3.6 │ Assignments of the four biomolecular groups in a typical biological spectrum: lipids, proteins, nucleic acids and carbohydrates

Significant research, with a substantial quantity of cancer-related biomaterial, has been conducted over recent years with a view to clinical application. Studies have included cervical, melanoma, prostate, oesophagus, lymphoma, gastric, endometrial adenocarcinomma and fibrosarcoma [9-16]. Many studies have investigated the difference in cancer signatures by use of peak ratios, spectral shifts and absorbance differences.

As Figure 3.6 suggests, there are a considerable number of spectral peaks that are associated with lipid absorbencies in a typical biological sample and hence possible biomarkers of carcinoma. For instance, the ratio of methyl (CH_3) and methylene (CH_2) peaks have been assessed in a number of studies. Fung et al. reported on differences between normal and malignant epithelial tissue in endometrial adenocarcinomma between the membrane lipid methylene stretching mode at 2853 cm^{-1} and an asymmetric methyl stretch from cellular proteins nucleic acids and lipids at 2959 cm^{-1} [16]. The ratio between these two lipid markers decreased in methyl/methylene in the malignant epithelial tissue, relative to the normal tissue. This is said to be consistent with the hypomethylation observed during carcinogenesis in other tissues, which may alter gene expression and facilitate

mutations [18]. This conclusion was also reached by Maziak etl al [13] who in addition found an increase in methylene peak absorption suggested an accumulation of triglycerides in malignant tissue.

Conversely, Andrus et al. [19] claimed an increasing CH_3/CH_2 ratio with increasing grade of malignant non-Hodgkin's lymphoma. Similarly, Yao et al. [20] recently found that the lipid ratios of 2925/1460 cm^{-1} and 1740/1460 cm^{-1} decreased in the cancer spectra, indicating lower quantity of lipids in the colon neoplasm. Liu et al. also found a marked decrease in lipid content in papillary thyroid carcinoma [21].

The discrepancy between the findings of Fung and Maziak et al. compared to Andrus and Yao et al. suggests that these spectral biomarkers may not be translatable to all cancers. For instance, Mordechai et al. compared spectra of both cervical and melanoma tissue and found that while a reduction in the glycogen carbohydrate was a good diagnostic marker for malignant cervical tissue, the same trend was not observed for the malignant melanoma [9]. Despite the lack of correlation with these cancers there is a striking case to support a relative reduction in glycogen in malignant tissue, compared with non-malignant tissue.

The glycogen/phosphate ratio with absorbance intensities at 1030 and 1080 cm^{-1} has been investigated as a marker in malignant tissue. Reduced glycogen and increased phosphate levels were found in cancer of the cervix [11], prostate [12] and oesophagus [13]. The mutual trend may be indicative of malignancy due to an increased metabolic turnover, and is explored further in chapter four.

Perhaps the most obvious spectral signals that exhibit characteristic differences in malignant samples will be the peaks associated with nucleic acids due to the differences in genetic sequencing. A difference in the phosphodiester symmetric and anti-symmetric stretching bands associated with the sugar-phosphate backbone of DNA and also phosphorylated proteins and phospholipids has been observed in many studies. Pronounced increases in these bands have been observed in malignant cervical and gastric samples [11, 15]. Yao et al. found a peak shift to a

higher wavenumber in the phosphodiester band at 1080 cm^{-1} in malignant tissue relative to benign colon neoplasm [20].

For ribose in particular, it has been observed that malignant samples display an increased level. Mordechai et al. found the RNA/DNA ratio at 1121/1020 cm^{-1} increased in malignant cervical and melanoma tissue [9]. This result is supported by the findings of Andrus et al., where an increase in ribose content at 1121 cm^{-1} was seen to correlate with an increase in the 996/966 cm^{-1} ratio, another index of RNA/DNA, with increasing malignancy grade in lymphoma [19].

While Maziak et al. observed an increase in the nuclei-to-cytoplasm ratio in tissue samples of adenocarcinoma of the oesophagus, the relative amount of DNA increased and the content of RNA decreased, suggesting the opposite to the studies by Mordechai and Andrus et al. [13].

The research suggests that while there are spectral markers that consistently discriminate differences in the biochemistry between non-malignant and malignant biological samples. The trends of these markers, however, appear to differ and seem to be cancer specific. In future research, now that there is increased understanding about the issue of the 'dispersion artefact', there will hopefully be an increase in the number of studies that concentrate on spectral assignments at the single cell level, which may give us increased understanding about how we can relate any infrared biomarkers to the biological processes that govern their change.

3.3 │ Chemometric Techniques for Spectral Analysis

3.3.1 │ Principal component analysis

PCA is a method of finding patterns in a set of data, highlighting similarities and differences when patterns are difficult to find in data of high dimension. When a dataset is comprised of two variables, the dataset is two-dimensional and therefore fairly easy to analyse. An infrared spectrum, however, will include hundreds or thousands of data points, depending on the chosen resolution and spectral range.

The data is 'n dimensional' and extremely complicated to interpret. Therefore spectroscopists regularly use PCA as a way to explore the reasoning behind which particular spectra may class together and what spectral features may determine their classification. There are several examples of the use of this technique in characterising cancer cell lines [22-25].

The initial dataset coordinates are transformed into new versions in order to display the variance of the experimental data much more clearly [26]. The principal components are calculated in such a way so that the largest amounts of variance are covered. They consist of linear combinations which are weighted sums calculated from the initial variables, which are the absorbencies measured at each wavenumber in a spectrum. In terms of spectroscopy, the new linear combinations do not relate to particular spectral bands, but a collection of bands, each with different weighting or variance contribution to the PC.

The input matrix is the experimental observation data consisting of absorbance data points of particular spectrum at a particular wavenumber. The raw data is usually mean centred initially the mean is subtracted from each spectrum, producing a normalised data set with mean of zero [26].

Variance is a measure of the spread of data within a data set, very similar to standard deviation, only applies to one dimension. Covariance is calculated in PCA as data of more than one dimension can be assessed and can allow comparisons of how much two dimensions will vary with respect to one another. Positive covariance occurs when two dimensions increase together, i.e. they vary together in positive correlation. If covariance is negative, then one dimension increases while the other dimension decreases. Zero covariance indicates that the dimensions are independent of each other. If there are more than two dimensions in a data set then more than one covariance can be calculated and a covariance matrix is calculated [26].

The covariance matrix is then split into the score matrix, which is comprised of the scaling coefficients (also known as eigenvalues), and the loading matrix (also

known as eigenvectors or principal components). The loadings or eigenvectors are non-zero vectors and can only be found for square matrices that are multiplied with a vector that results in a new vector that has been scaled from the original vector's position (Fig. 3.7).
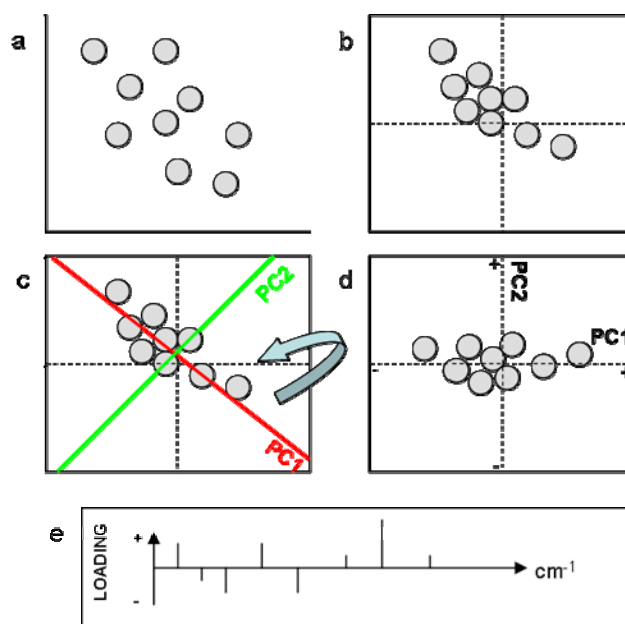
Figure 3.7 │**PCA in practice: The raw data (a) is first normalized (b). The axes are then rotated to explain maximum variance in the dataset (c). This is shown by the PC1 line in red which explains the most variance. The next best explanation is the line through PC2 in green. Each data point can be classified as a summation of the contributions from each of those lines (d). The PC1 loading plot will determine the sub-component (i.e. spectral peaks) contributions to variance in the PC, each with a different weighting (e).**

Eigenvectors allow the most important patterns in a data set to be pinpointed and displayed in a PCA plot. All eigenvectors are orthogonal to one another. This is important in PCA as data can be expressed in terms of these perpendicular eigenvectors, for example comparing the data in PC1 versus PC2 [26].

Since the calculation of the principal components uses the criteria that the largest amount of variance should be covered in the experimental data, the first principal component features the maximum variance, with sub sequential PCs carrying decreasingly less variance. If the dataset was not mean centred initially, the first PC would be equal to the mean of the dataset and the second PC would explain the

most variance in the dataset. The principal components towards the upper limit usually describe miniscule data variance and merely contain spectral noise as well as unnecessary information from the experimental spectra.

3.3.2 │Linear discriminant analysis

LDA is different to PCA in that it directly discriminates data between classes whereas PCA only examines the principal component analysis as a whole set of data and does not give much indication of underlying class structure [27].

A set of independent variables are used to separate data points based on the categorical dependent variables which are the groups that are user defined. New variables create a line onto which the centroid of different clustered groups are projected as far apart as possible, based on linear combinations of the dataset (hence the term linear discriminant analysis is used). If the model is for classification purposes the model can be assessed for classification efficiency by a report of correctly assigned groups using training and test datasets.

Usually the dataset encompasses a series of principal components which have been calculated in PCA prior to LDA. PC-LDA is a method of improving intra-group clustering whilst maximising inter-group clustering (Fig. 3.8).
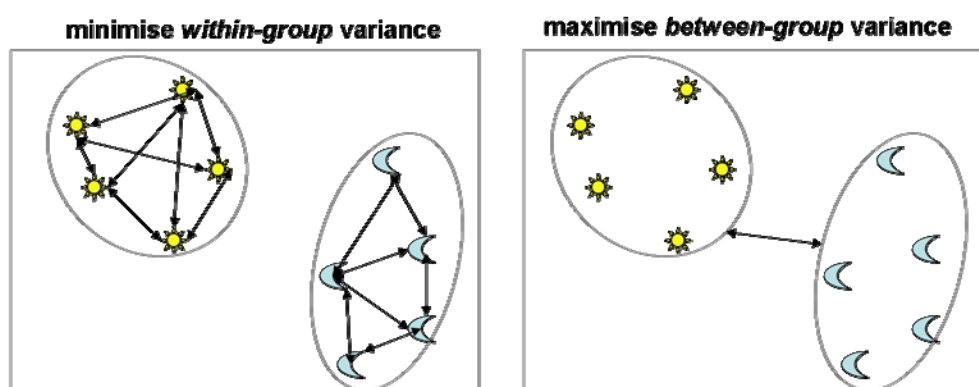
Figure 3.8│ LDA allows for reduction in inter-group variance whilst maximising the distance between different cluster groups

There is, however, a level of subjectivity associated with the technique, namely the number of principal components to use. The topic has been debated extensively and there are a number of methods offered in the literature to determine this factor, such as the Kaiser rule, parallel analysis, Cattell's scree plot analysis, or simply taking the number of PCs that account for a fixed level of variance explained [28-31]. LDA will be used in the experimental analysis of various renal cancer cell lines and will be further explored and discussed in chapters five and six.

3.3.3 │ Cluster analysis

One method of analysis, k-means clustering, aims to separate data points in k clusters in which each data point belongs to the cluster with nearest mean. The user is asked to define how many clusters are desired [27].
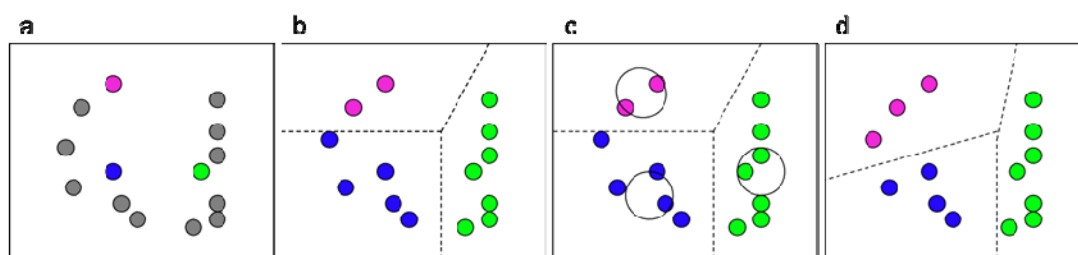
Figure 3.9 │ A simplified demonstration of standard K-means clustering

The initial k-means are randomly assigned (Fig 3.9a) and each individual data point is grouped to whatever k-mean the data point is in closest proximity to (Fig 3.9b). Each k means centre then calculates its centroid based on its acquired data points (Fig 3.9c). The centroid of each cluster becomes the new mean (Fig 3.9d) and each individual data point surrounding the new mean is grouped. This continues until convergence has been reached.

3.4 │ References

1. Griffiths PR, de Haseth, J.A.: **Fourier Transform Infrared Spectrometry** 2nd edition: *Wiley* 2007: Ch. 2,4,5,7,10.

2. Bassan P, Gardner P: **Biomedical Applications of Synchrotron Infrared Microspectroscopy**. *RSC Publishing* 2011: Ch. 8.

3. Kohler A, Sulé-Suso J, Sockalingum GD, Tobin M, Bahrami F, Yang Y, Pijanka J, Dumas P, Cotte M, van Pittius DG, Parkes G, Martens H: **Estimating and Correcting Mie Scattering in Synchrotron-Based Microscopic Fourier Transform Infrared Spectra by Extended Multiplicative Signal Correction**. *Applied Spectroscopy 2008*, **62**, 3, pp. 259-266.

4. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, Gazi E, Brown M, Clarke N, Gardner P: **Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples**. *Analyst* 2010, **135**(2):268-277.

5. Bassan P, Byrne HJ, Lee J, Bonnier F, Clarke C, Dumas P, Gazi E, Brown MD, Clarke NW, Gardner P: **Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells**. *Analyst* 2009, **134**(6):1171-1175.

6. Lasch P, Haensch W, Naumann D, Diem M: **Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis**. *Biochimica et Biophysica Acta- Molecular Basis of Disease* 2004, **1688**(2,2);176-186.

7. Stuart B: **Biological Applications of Infrared Spectroscopy**. *John Wiley & Sons*; 1997: p119.

8. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

9. Mordechai S, Sahu, R.K., Hammody, Z., Mark, S., Kantarovich, K., Guterman, H., Podshyvalov, A., Goldstein, J., Argov, S.: **Possible common**

biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. *Journal of Microscopy* 2004, **215**(1):86-91.

10. Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, Diem M: **Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix.** *Biospectroscopy* 1998, **4**(1):47-53.

11. Wood BR, Quinn MA, Burden FR, McNaughton D: **An investigation into FTIR spectroscopy as a biodiagnostic tool for cervical cancer.** *Biospectroscopy* 1996, **2**(3):143-153.

12. Gazi E, Dwyer, J., Gardner, P., Ghanbari-Siahkali, A., Wade, A.P, Miyan, J., Lockyer, N.P., Vickermann, J.C., Clarke, N.W., Shanks, J.H., Scott., L.J., Hart, C.A., Brown, M.: **Applications of Fourier transform infrared microspectoscopy in studies of benign prostate and prostate cancer. A pilot study.** *Journal of Pathology* 2003, **201**(1):99-108.

13. Maziak DE, Do MT, Shamji FM, Sundaresan SR, Perkins DG, Wong PTT: **Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: An exploratory study.** *Cancer Detection and Prevention* 2007, **31**(3):244-253.

14. Sahu RK, Mordechai S, Manor E: **Nucleic acids absorbance in Mid IR and its effect on diagnostic variates during cell division: A case study with lymphoblastic cells**. *Biopolymers* 2008, **89**(11):993-1001.

15. Fujioka N, Morimoto Y, Arai T, Kikuchi M: **Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy.** *Cancer Detection and Prevention* 2004, **28**(1):32-36.

16. Fung MFK, Senterman MK, Mikhael NZ, Lacelle S, Wong PTT: **Pressure-tuning Fourier transform infrared spectroscopic study of carcinogenesis in human endometrium.** *Biospectroscopy* 1996, **2**(3):155-165.

17. Yang D, Castro DJ, El-Sayed IH, El-Sayed MA, Saxton RE, Zhang NY: A **Fourier-transform infrared spectroscopic comparison of cultured human fibroblast and fibrosarcoma cells: A new method for detection of malignancies.** *Journal of Clinical Laser Medicine and Surgery* 1995, **13**(2):55-59.

18. Counts JL, Goodman JI: **Alterations in DNA methylation may play a variety of roles in carcinogenesis**. *Cell* 1995, **83**(1):13-15.

19. Andrus PG: **Cancer monitoring by FTIR spectroscopy**. *Technology in Cancer Research and Treatment* 2006, **5**(2):157-167.

20. Yao HW, Liu YQ, Fu W, Shi XY, Zhang YF, Xu YZ: **Initial research on Fourier transform infrared spectroscopy for the diagnosis of colon neoplasms.** *Spectroscopy and Spectral Analysis* 2011, **31**(2):297-301.

21. Liu Y, Xu Y, Liu Y, Zhang Y, Wang D, Xiu D, Xu Z, Zhou X, Wu J, Ling X: **Detection of cervical metastatic lymph nodes in papillary thyroid carcinoma by Fourier transform infrared spectroscopy**. *British Journal of Surgery* 2011, **98**(3):380-384.

22. Harvey TJ, Gazi E, Henderson A, Snook RD, Clarke NW, Brown M, Gardner P: **Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy**. *Analyst* 2009, **134**(6):1083-1091.

23. Harvey TJ, Henderson A, Gazi E, Clarke NW, Brown M, Faria EC, Snook RD, Gardner P: **Discrimination of prostate cancer cells by reflection mode FTIR photoacoustic spectroscopy**. *Analyst* 2007, **132**(4):292-295.

24. Murali Krishna C, Kegelaer G, Adt I, Rubin S, Kartha VB, Manfait M, Sockalingum GD: **Characterisation of uterine sarcoma cell lines exhibiting MDR phenotype by vibrational spectroscopy**. *Biochimica et Biophysica Acta - General Subjects* 2005, **1726**(2):160-167.

25. Walsh MJ, German MJ, Singh M, Pollock HM, Hammiche A, Kyrgiou M, Stringfellow HF, Paraskevaidis E, Martin-Hirsch PL, Martin FL: **IR microspectroscopy: potential applications in cervical cancer screening.** *Cancer Letters* 2007, **246**(1-2):1-11.

26. Gauglitz G, Vo-Ding, T, (ed.): **Handbook of Spectroscopy**. *WILEY-VCH* 2003, Ch.3-5.

27. Martinez AM, Martinez AM, Kak AC: **PCA versus LDA** *Transactions on Pattern Analysis and Machine Intelligence* 2001, **23**(2):228-233.

28. Kaiser HF, Norman WT: **Coefficients alpha for components**. *Psychological Reports* 1991, **69**(1):111-114.

29. Kanyongo GY: **The influence of reliability on four rules for determining the number of components to retain**. *Journal of Modern Applied Statistical Methods* 2006, **5**(2):332-343.

30. Watkins MW: Monte **Carlo PCA for Parallel Analysis**: *State College, PA: Ed & Psych Associates*; 2000.

31. Ledesma RD, Valero-Mora, P.: **Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis.** In: *Practical Assessment, Research & Evaluation*. vol. **12**; 2007.

Chapter 4 │ Multidisciplinary Validation of FTIR Spectroscopy for Cell Characterisation

4.1 │ Introduction and Aims

Evidence suggests that FTIR spectroscopy can provide a means to analyse biological material, with an emphasis on clinically relevant research [1]. It has been shown to enable identification of cancerous tissue biopsies and in some cases provide an estimation of cancer grade with comparable sensitivity and specificity to conventional approaches [2-6]. There is also emerging evidence that FTIR could be implemented to assess cellular drug interactions that may aid screening for therapeutic efficacy in the future [7-11].

There has been a major limitation which has held back the progression of Mid-IR FTIR spectroscopy. As previously mentioned (Ch. 3.1), FTIR spectral analysis of biological material has been hampered due to spectral distortions associated with Resonant Mie Scattering (RMieS). Examples of such distortions include sinusoidal baselines and peak shifts, particularly in the carbonyl $\nu(C=O)$ peak at ~1730 cm^{-1}, associated with lipid esters [12].

Until recently RMieS was not fully understood. Measures to avoid the issue included simply ignoring the amide I band, or baseline correcting using the rubber band method [13]. This involves a conceptual rubber band that is stretched beneath the spectrum. A new baseline is constructed where the rubber band touches the spectrum at selected points in the spectrum. This type of correction can introduce bandshape and intensity errors. Consequently, spectral results may be mis-leading and therefore conclusions formed on the basis of such data are likely to be erroneous.

High demands are placed on any clinically-targeted technology in terms of their performance, confidence and accuracy. If the spectroscopy community cannot be confident in the results of FTIR spectral analysis for bio-materials then it is probable it will not be implemented in the clinic. The RMies-EMSC correction algorithm has provided a means to increase our confidence in terms of ability to separate real chemical information from spectral perturbations caused by physical

affects. The algorithm has been demonstrated to work well using simulated data, where it was possible to assess the robustness of the correction [14].

The cell lines, selected to specifically test the robustness of the algorithm, the renal cell lines were 2245R and 2247R, established from primary kidney resections, and 2246R derived from a metastatic pleural nodule, derived from different surgical specimens from different patients diagnosed with clear-cell renal carcinoma at the N.C.I., U.S.A. Cell line 2247R has a Fuhrman nuclear grade 3. No grades were reported for cell lines 2245R and 2246R.

For the FTIR-based characterisation of the renal cell lines, three different approaches were taken in terms of sample preparation and spectral acquisition, as outlined in Figure 4.1.
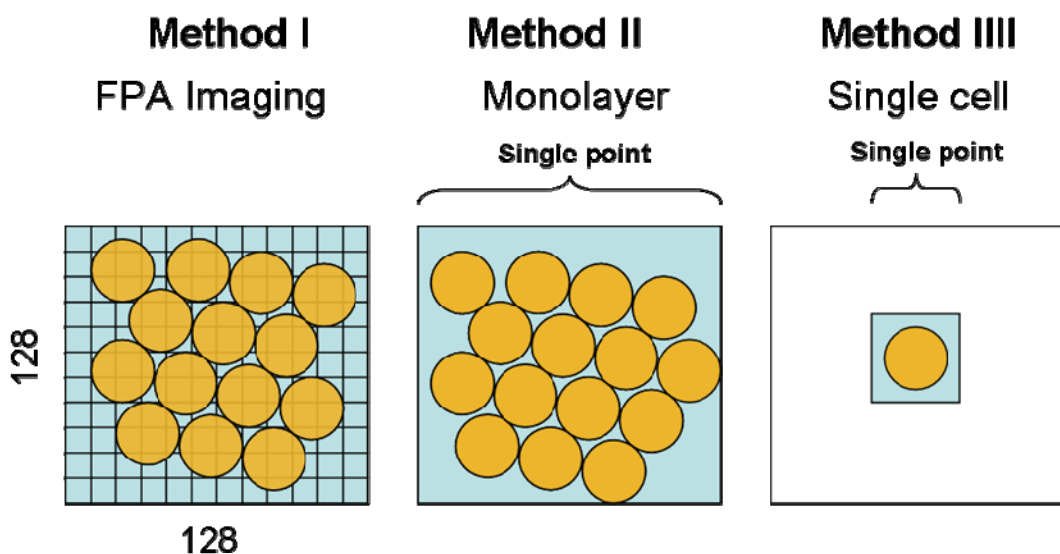
Figure 4.1│ Schematic diagram of the different spectral methods employed in the cell characterisation study: Method I utilises focal plane array FTIR imaging technology with 128 ×128 elements, simultaneously acquiring the hyperspectral image of the sample at a resolution of 5.5 microns per pixel. Method II uses conventional bench top single point Microspectroscopy where the aperture (in blue) is opened to record the average signal of numerous cells (orange) in a population. Method III is also single point micro-spectroscopy but utilises a synchrotron radiation source for high brilliance, increasing the signal to noise of a single cell spectrum relative to a bench top infrared source.

Recently factors affecting cell line discrimination using principal component analysis have been investigated, such as use of different cell culture media [15]. The results indicated that the use of different culture media did not provide the basis for chemometric separation in different prostate cell lines. Method I (Ch. 4.2.2) was designed to take the idea of this investigation further. Firstly, the cell lines were cultured under the same conditions. Secondly, the cells were co-cultured onto the same substrate at different points of reference and allowed to reach complete confluence. In doing so, more confidence could be placed in the sample preparation, i.e. culture media or use of different substrates was not a factor of separation between cell lines.

In addition, the use of the glycogen/phosphate ratio at 1030 cm^{-1}/1080 cm^{-1} peaks has been investigated to assess cancer biopsies and is thought to be a metabolic marker [2, 16]. Method I provided a means to relate chemical information in terms of phosphate and glycogen absorptions to the results of the biological proliferation assay to test this biomarker of metabolism, as a high rate of metabolic turnover could have a link with a high rate of proliferation.

The second stage of analysis involved single point spectra of the cell monolayers using two approaches (Method II, Ch. 4.2.3). Conventional FTIR transflection measurements were made as well as measurements by FTIR-PAS, a technique relatively free from scattering perturbations (Ch. 2.2.4). Method II provided optimum conditions for the FTIR-PAS measurements and so a direct comparison could be made with the conventional FTIR approach, using transflection MirrIR substrates. It was determined that alternative substrates such as calcium fluoride (CaF$_2$), silica chips and quartz gave poorer results in terms of signal-to-noise (SNR); the SNR for a CaF$_2$ substrate was 25 compared to a typical spectrum obtained from a MirrIR substrate at 88 [17]. It has been proposed that the highly reflective silver coating on the MirrIR substrates give rise to a 'double absorption' effect whereby light reflected off the surface of the substrate, as well as incident light, will be absorbed by the sample, further increasing the intensity of the signal [17].

Single point spectra of a cell monolayer that provide an average signal from a cell population radiation whereby hundreds of cells are simultaneously irradiated, result in a spectrum with excellent signal to noise and an overall chemical signature of a cell population. Method III enabled single cell infrared measurements to be made. Good signal to noise spectra for single cells may be obtained with the use of synchrotron radiation [18]. The spectrum of a single cell may present more detailed information in comparison to that of the average signal from cell monolayers, however they typically suffer the most severe spectral distortion due to RMieS. Single cell spectra therefore present the biggest challenge in terms of RMieS correction.

When implementing the RMieS-EMSC algorithm to assess experimental (non-simulated) sample data, the 'true answer' in terms of result validity could not be known. It was accepted that a multidisciplinary approach were required to validate corrected experimental spectra. Protein-based techniques namely LCMS (Ch. 2.3.7.2) and Western blotting (Ch. 2.3.7.3) were employed to validate results from the FTIR spectroscopy analysis.

4.2│Experimental Methods

4.2.1│ Proliferation

All three cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) with foetal calf serum (FCS) (10%) and L-glutamine (1%), grown at 37˚C in a humidified atmosphere (5% CO_2) until confluent. Cells were seeded at a density of $1x10^5$ cells per well in a six well plate, one for each day. At each time point cells were removed. To do this, the cells in each specific well were pre-washed twice in phosphate buffered saline (PBS) before trypsin (1 mL) was added and incubated at 37˚C for approximately 3 minutes to detach the cells. After cell detachment DMEM media with 10% FCS was added and the cells were transferred to Falcon tubes for centrifugation (800g for 5 minutes), along with the PBS washes. A Trypan blue exclusion assay was performed for the proliferation characterisation. Total viable

cell number per day was counted using a hemocytometer and a ratio of 1:1 Trypan blue: cells in media. The Trypan blue dye permeated the membranes of dead cells which were excluded from the count. For robustness each time point would be counted with triplicate wells, taking the average result as the final count.

4.2.2│FTIR method I: Imaging of dual-cell monolayer

The targeted positions for 2245R and 2247R cells were marked on the non-reflective side of the MirrIR substrate. Micro-droplets (1000 cells in 20 μl /media) were positioned on top of the IR reflective surface such that the two cell type micro-droplets did not mix (Fig. 4.2). H_20 (1 mL) was added to the tissue culture plastic well holding the slide and incubated for 15 minutes to allow the cells to adhere without drying out. Media (DMEM with 10% FCS, 1% L-Glutamine) was then added to submerge the slide. The cells were left to proliferate until 100 % confluence was reached.
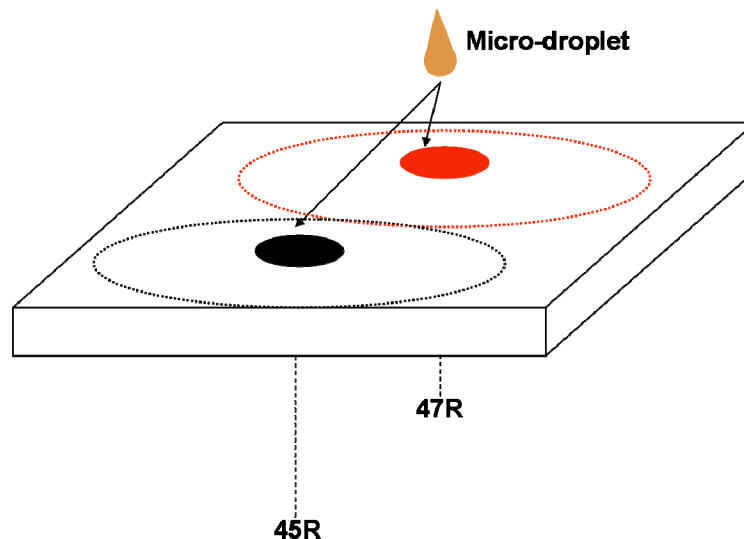
Figure 4.2 │**Schematic diagram for method I sample preparation: Micro-droplets of concentrated 2245R and 2247R cells were placed at designated points on the IR reflective side of the substrate. Designated points were marked on the non-reflective side of the MirrIR slide. The cells were left to establish as a monolayer until an interface between the two cell types was reached.**

Transflection-mode FTIR imaging spectroscopy was performed using the method outlined in Ch. 2.4.5.3. The obtained infrared image was then corrected with the RMieS-EMSC algorithm with 10 iterations and vector normalised. For analysis of

the glycogen/phosphate ratio, the peak heights at 1030 cm^{-1} and 1080 cm^{-1} were assessed for every spectrum.

4.2.3│FTIR method II: Cell monolayers

Cells were cultured and prepared for FTIR as outlined in Ch. 2.4.1 and 2.4.2.1. The FTIR samples were cultured onto standard size slides (2.5 x 7.5 cm^2). Single point transflection spectra were taken using the method described in Ch. 2.4.3.1. The dataset was acquired with a spectral resolution of 4 cm^{-1} with 256 co-added scans per spectrum. The data comprised of 20 spectra taken from 10 samples per cell line, in randomised order and on different sampling days.

The sample preparation of the FTIR-PAS samples followed the same protocol as for standard FTIR spectroscopy; however they were cultured onto smaller MirrIR substrates (0.5 x 0.5 cm^2) to fit the sample chamber of the photoacoustic accessory. Measurements were taken using the instrumentation described in Chapter. 2.4.3.4. The system was allowed to purge for a period of 4 minutes per spectrum. The data set comprised of 20 spectra acquired from 10 samples per cell line.

4.2.4│FTIR method III: Single cell SR-FTIR microspectroscopy

Cells were cultured then cytospun onto calcium fluoride (CaF$_2$) slides as described in Ch. 2.4.2.2. SR-FTIR transmission spectra were acquired on the SMIS beamline as described in Ch. 2.4.3.2. 240 spectra were recorded at 4 cm^{-1} resolution with 256 co-scans and the size of the aperture was typically 14 μm^2.

4.2.5│Spectral pre-processing

All FTIR datasets were transformed in Matlab (The Mathworks Inc.) to second derivative with 7 point Savitsky-Golay smoothing and a polynomial order of 3, over a selected analytical range of 1825-1015 cm^{-1}. The datasets were finally vector normalised and mean centred prior to principal component analysis, performed as a single routine in Matlab.

4.2.6│Amide I analysis

Fourier self deconvolution (FSD) and curve fitting was performed using Omnic® software (Thermo Electron Corp.). The Fourier self-deconvolution (FSD) process can be manipulated to optimise the result and therefore it is not a quantitative solution. It does, however, enable an estimation to be made for protein secondary structure. The parameters that can be manipulated include the bandwidth of the overlapped bands and the enhancement factor. The enhancement factor is a measure of the degree to which features are revealed and determines the 'strength' of the resolving power applied to the data.

FSD was carried out at full band width at half height of the peak (FWHH) at 16 cm^{-1} and a resolution enhancement of 1.4 for the spectral range covering the amide I band between 1720 and 1600 cm^{-1}. The enhancement factor used was subtle enough to avoid over processing the spectrum which would result in distorted and negative band shapes of complex functions which no longer describe the data. Peak fitting employed a Voigt fitting profile with low sensitivity in order to avoid over-fitting of bands (Fig. 4.3).
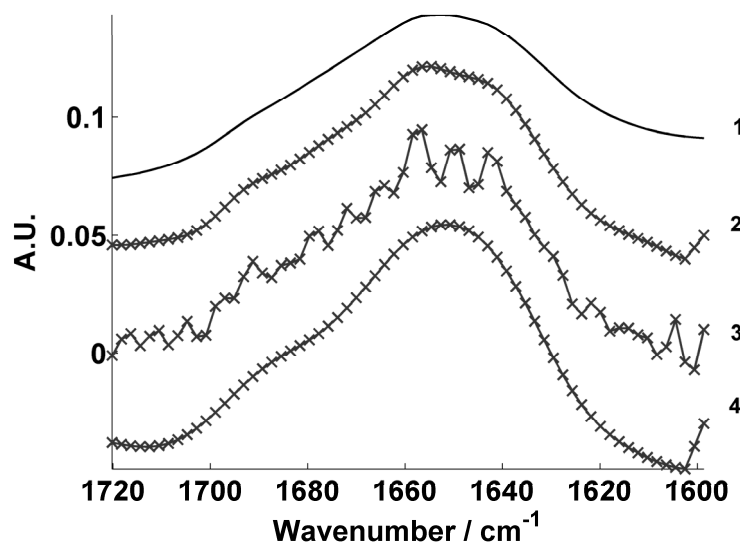
Figure 4.3 │Comparing different FSD spectra from (1) the original mean spectrum of 2245R (2) displays the chosen parameters for FSD; a bandwidth of 16 cm^{-1} and enhancement factor of 1.4. The FSD spectrum is similar in shape to the original, but features have been further resolved. (3, 4) illustrate examples where FSD is over-manipulated; the resulting spectrum is no longer similar to the original spectrum. FSD Spectrum (3) has a bandwidth of 16 cm^{-1} but a greater enhancement factor at 2.8 while FSD spectrum (4) has a bandwidth of 32 cm^{-1} and an enhancement of 1.4.

4.2.7│Label-free LCMS

The LCMS sample preparation and instrument operation was conducted by the Biological Mass Spectrometry Facility, Paterson Institute for Cancer Research, University of Manchester.

Sample preparation is detailed in Ch. 2.4.5. Peptides from each sample were analysed a total of five times in five equivalent LCMS injections (Fig. 4.4). A technical replicate of 2245R was prepared to ensure consistency in the sample preparation technique.
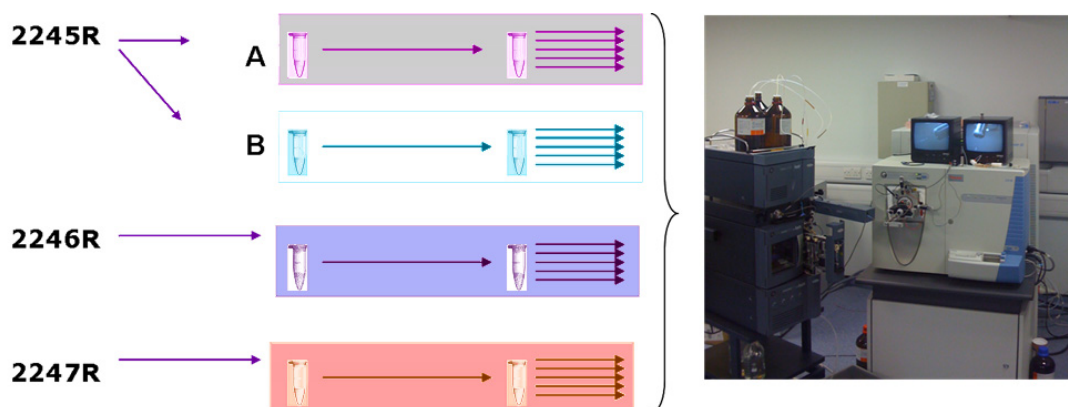
Figure 4.4 │Schematic for proteomic sample preparation; Protein lysates obtained from whole cells are injected into the LC-MS for analysis. A technical replicate was prepared for cell line 2245R to test the robustness of sample preparation and instrumentation.

4.2.8│Western blotting

The Western Blot was conducted by the Genito-Urinary Cancer Research Group at the School of Cancer and Enabling Sciences in the Paterson Institute for Cancer Research, University of Manchester.

The level of the protein Histone H2A expression was assessed by Western blot for all cell lines following standard protocol, described in Ch. 2.4.6. A total of nine samples were loaded: protein lysates at 27.2 µg, 13.7 µg and 6.7 µg were loaded

respectively for each cell line; to ensure loading consistency. For a control, GAPDH was also loaded.

4.3│Results and Discussion

4.3.1│Assessment of growth characteristics

Cell lines were initially characterised by their morphology and growth pattern. This was achieved by seeding each cell line at the same density (1×10^5 cells per well) and monitoring changes in the cell wells over a period of 5 days.

Cell lines 2245R and 2246R were similar in morphology with a long, oval-shaped cellular structure and extensive cytoplasm characteristic of clear cell renal carcinoma. The 2247R cells, however, were more rounded with stellate growth characteristics and radial expansions similar to mesenchymal morphology (Fig. 4.5).
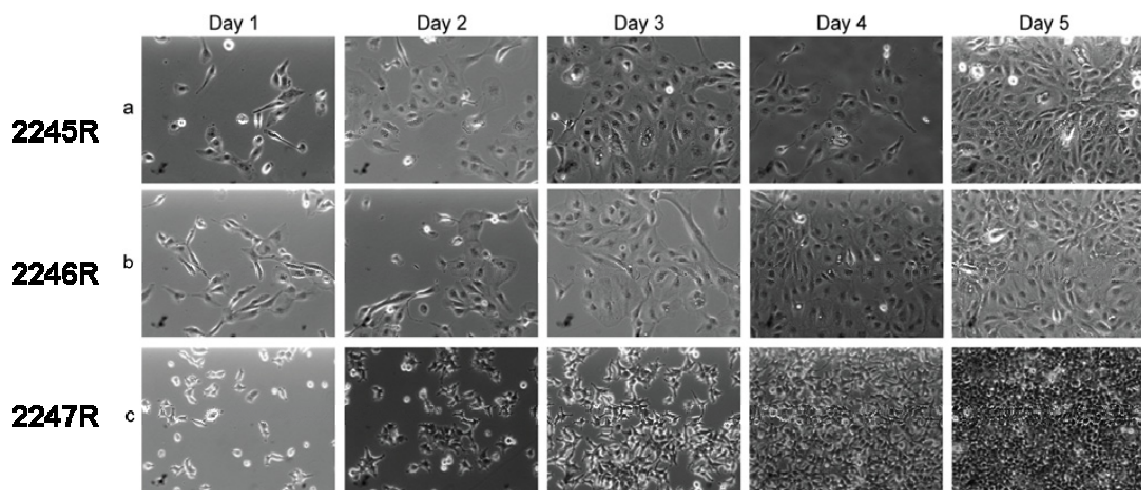
Figure 4.5│ In vitro photo-micrographs of cell lines over time. Cell lines (a) 2245R, (b) 2246R and (c) 2247R were seeded at an initial density of 1×10^5 cells per well in standard growth medium. Morphology and growth patterns were assessed by taking daily photo-micrographs with a 10× magnification under phase contrast.

A viable cell count was also performed during this study, as described in the experimental section (Ch. 4.2.1). The 2247R cell line exhibited a higher proliferation rate, in relation to cell lines 2245/6R, resulting in an exponential increase after two days (Fig. 4.6).
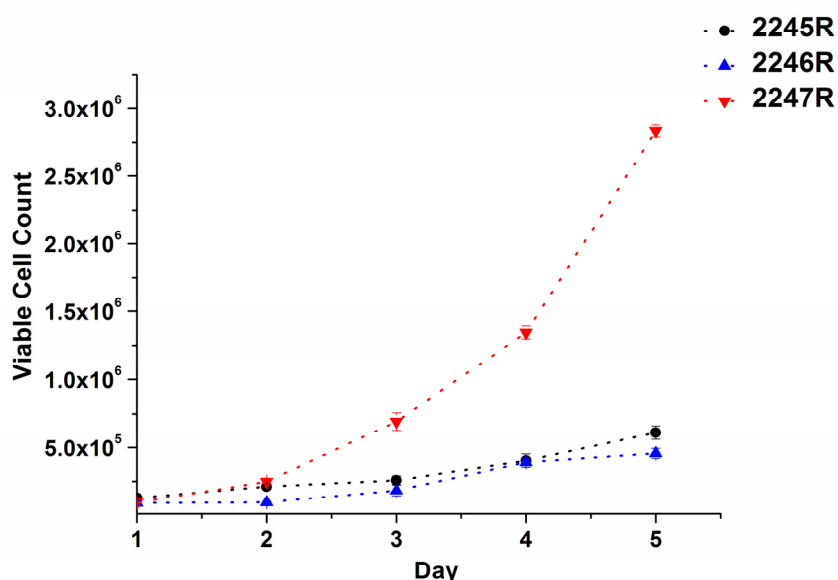
Figure 4.6│Growth characteristics of the renal cell lines: cells were seeded at 1×10^5 cells initially and incubated at 37°C with 5% CO_2 for a period of 5 days. Cells were counted daily using Trypan blue; viable cell quantities are plotted for each cell line.

4.3.2│FTIR imaging of the dual cell line monolayer

A dual cell line monolayer was established by cell culture onto a MirrIR slide (experimental section 4.2.2). In terms of growth behaviour, 2245R cells appeared to spread, extending outwards across the substrate surface in a 2D fashion. In contrast the 2247R cells started to stack on top of each other once 100% confluence had been reached across the surface. This can be observed in the optical image where the 2247R cells appear much darker than their counterparts due to the higher optical density of the 2247R section of the monolayer (Fig. 4.7).
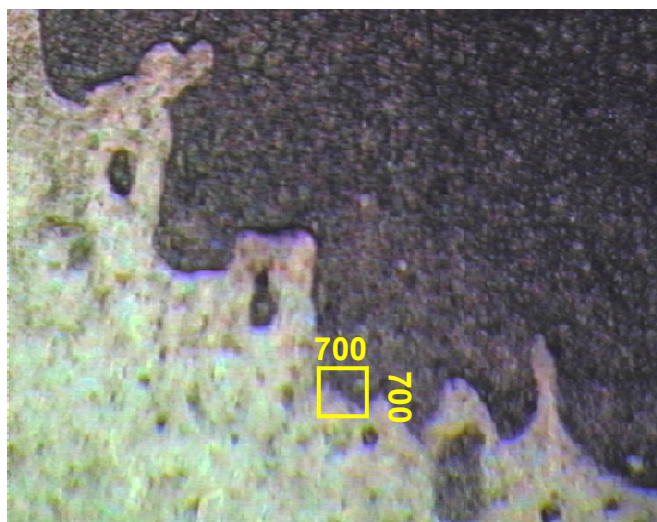
Figure 4.7 │ Optical image taken using a 4× objective using the FTIR-Microscope. The 2247R cell line appears much darker in colour than the 2245R cell line due to its higher cell density. The position of the section used for the FTIR spectral acquisition with a ×36 aperture is highlighted in the yellow box: a sampling area of 700 x 700 microns.

To compare the 1030 / 1080 cm^{-1} glycogen / phosphate ratio, which is said to be indicative of metabolic turnover, a hyperspectral 128×128 image, consisting of 16384 spectra was acquired from the 2245R-2247R monolayer interface. Figure 4.8 displays the pre-processing steps applied to the image. The total intensity raw image (Fig. 4.8a) maps out the two cell lines, with the more abundant 2247R cell line on the upper right-hand side and the 2245R cell line on the lower-left side, as observed in the larger optical image (Fig. 4.7).

Figure 4.8b displays the resulting image if RMieS-EMSC is not used prior to vector normalisation, where the total intensity chemistry still appears to distinguish between the cell lines. Once the raw image had been RMieS-EMSC corrected and vector normalised (Fig. 4.8c,d) the total intensity appeared to be more homogeneous across the image.

It was noted that in the 2245R region there are still patches of areas that are thinner and this may be due to the fact that the cell line has more contact inhibition, as seen in Figure 4.5.
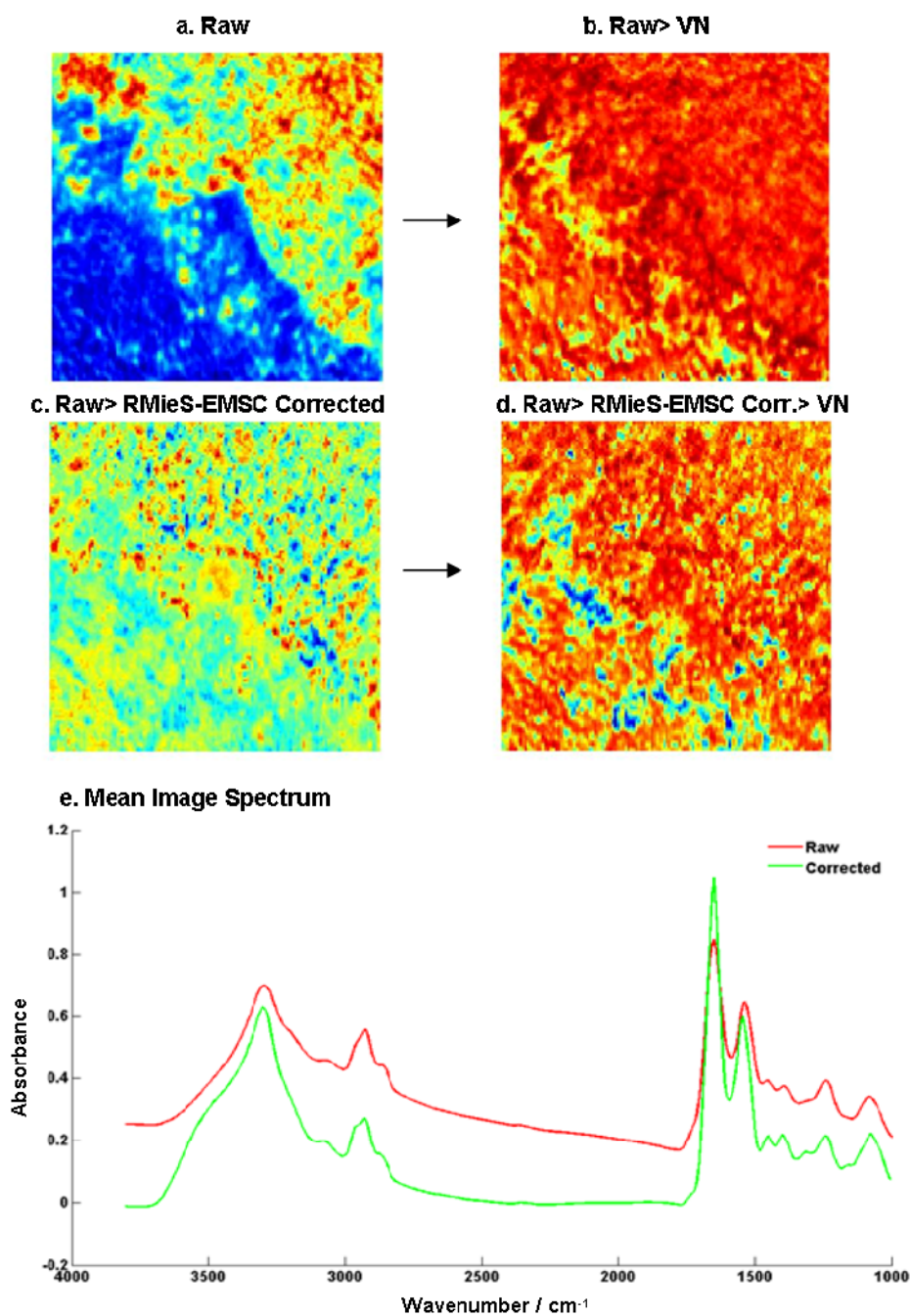
Figure 4.8 │ **(a-d) Total intensity heat maps of the hyperspectral image of the 2245R/2247R monolayer during pre-processing and analysis. All intensity maps were auto-scaled in Matlab. Low absorbance is shown in blue and high absorbance in red. Images (a,b) display the raw intensity image and its vector-normalised format (c,d) Display the pre-processing steps used in the analysis. The raw image was RMieS-EMSC corrected with 10 iterations before vector normalisation. (e). The mean spectrum is displayed for the raw (red) and RMieS-EMSC corrected spectrum (green).**

107

The RMieS effect in highly confluent monolayers is relatively subtle, in contrast to single cell spectra. Use of the RMieS-EMSC algorithm, however, is still required. Figure 4.8e displays the mean spectrum of raw and RMieS-EMSC corrected data, in where the most notable difference is the correction of the baseline.

After the RMieS-EMSC correction and subsequent vector-normalisation, the peak heights of all 16384 spectra were determined at the spectral wavenumbers 1030 and 1080 cm^{-1} and ratioed. Figure 4.9 displays the heat map for the 1030/1080 cm^{-1} ratio across the interface of the two cell lines. A decreasing ratio is said to indicate a higher rate of metabolic turnover, where glycogen stores are depleted (Ch. 1.1.4) and may be indicative of increasing malignancy [16, 19]. The FTIR imaging result does appear to correlate with the higher rate of proliferation in 2247R cells, as noted by the cell viability assay (Fig. 4.6).
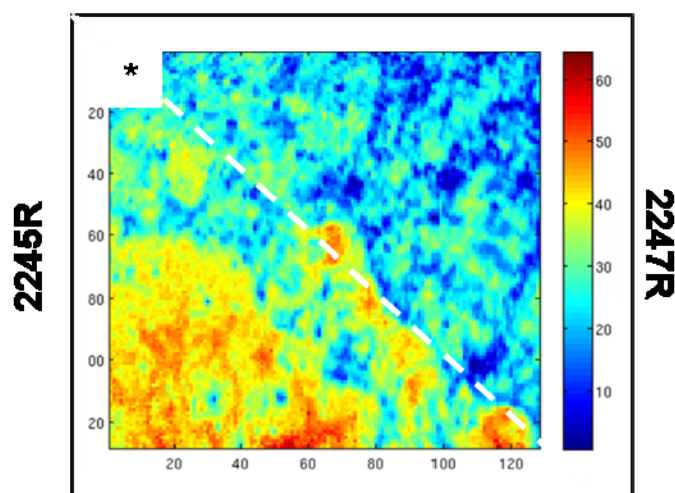
Figure 4.9│the glycogen/phosphate ratio intensity image was constructed using the peak heights at wavenumbers 1030/1080 cm^{-1} for each spectrum in the image. (*) shows the rough interface between the two cell lines where 2247R was cultured on the upper right hand side, and 2245R towards the lower left hand side.

4.3.3 │ Assessment of RMieS-EMSC correction

As mentioned previously, Resonant Mie scattering is inherent in all spectra from biological samples that are not comprised of a thin, homogenous film. Severity of the spectral perturbation can depend on the morphology of the 'scattering object' [12]. The impact of morphology in relation to RMieS has been highlighted by the two preparation methods. Mid-infrared radiation has a wavelength range of 2.5-14 µm and therefore samples exhibiting components closer to this range will suffer from severe distortion [14] and scattering from the edges of cells may also make a significant contribution [14].

The effect of RMieS is apparent in the spectral datasets for the methods II and III, which included single point cell monolayer spectra and spectra of single cells. As the cell lines used for the two methods were identical, the only difference was in morphology due to sample preparation (i.e. cells were cultured onto a surface in method II and cells were detached and spun onto the surface in method III). It was therefore been shown that samples of different morphology will have a different RMieS profile.

The adherent cells, cultured onto the substrate were able to extend across the 2D substrate surface, appearing significantly larger and more uniform in thickness and therefore displayed less obvious baseline distortions as observed in the corresponding raw spectra in Figure 4.10a. In contrast the single cells, which were lifted off the surface of the tissue culture plastic and cytospun onto substrates were typically more spherical and smaller in size (approximately 14 µm in diameter) and their spectra displayed severe distortions (Fig. 4.10c). Using the EMSC-RMieS algorithm the raw absorbance spectra were subjected to 150 iterations of the RMieS-EMSC algorithm and the results can be seen in Figures 4.10b and 4.10d illustrating the extent of correction to the distorted baseline. As it is difficult to suggest how many iterations of the algorithm to use, the general rule of thumb was used that the number of iterations increased until the using RMieS scattering distortions were no longer apparent.
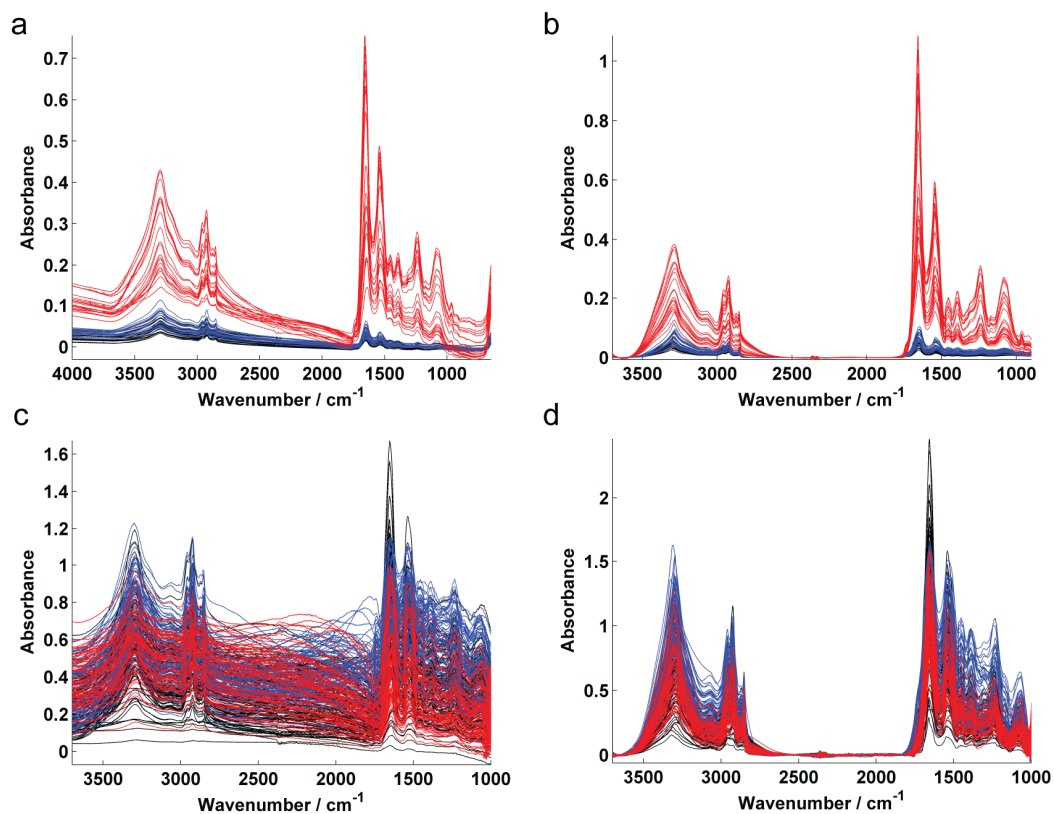
Figure 4.10 │ The raw and RMieS-EMSC corrected spectra for the cell monolayers (a, b) and for the single cell spectra respectively (c, d). Sample spectra were colour coded as 2245R (black), 2246R (blue) and 2247R (red).

4.3.4 │ Principal component analysis

The PCA results for the FTIR spectral datasets before and after EMSC-RMieS correction are shown in terms of their clustering, relative to the FTIR-Photoacoustic result in Figure 4.11.

It is apparent that cell line 2247R does display cluster separation, with respect to the two other cell lines in the uncorrected score plots (Fig. 4.11a, c). It is probable, however, that the different RMieS scattering profile of each cell line had a degree of influence in the cluster separation. The uncorrected score plots from the raw spectra illustrate that simply ignoring the issue of RMieS can potentially lead to misleading result conclusions.

Cell classification was desired in terms of the global chemistry differences within each cell type, rather than physical effects, such as different RMieS scattering profiles due to cell morphology. Figures 4.11b and 4.11d display the PCA score plots for data that has undergone RMieS-EMSC correction for the monolayer and single cell spectra respectively. The cell line clusters are more distinguishable and show a better relation to the PCA score plot from the FTIR-PAS data that did not require RMieS correction.

Cell line 2247R aside, the SR-FTIR single cell data (Fig 4.11d) displayed better discrimination of cell lines 2245R and 2246R in contrast to the monolayer datasets. This may reflect increased technique sensitivity in terms of cell classification and may be due to the specific targeting of cells during data collection, whereby the aperture is adjusted to capture the dimensions of the single cell itself. In contrast, the monolayer technique captures a broader range of cells, and possibly sections of surrounding protein secretions and blank substrate sections, which will also contribute to the averaged single point spectrum, possibly 'diluting' the cellular signal.
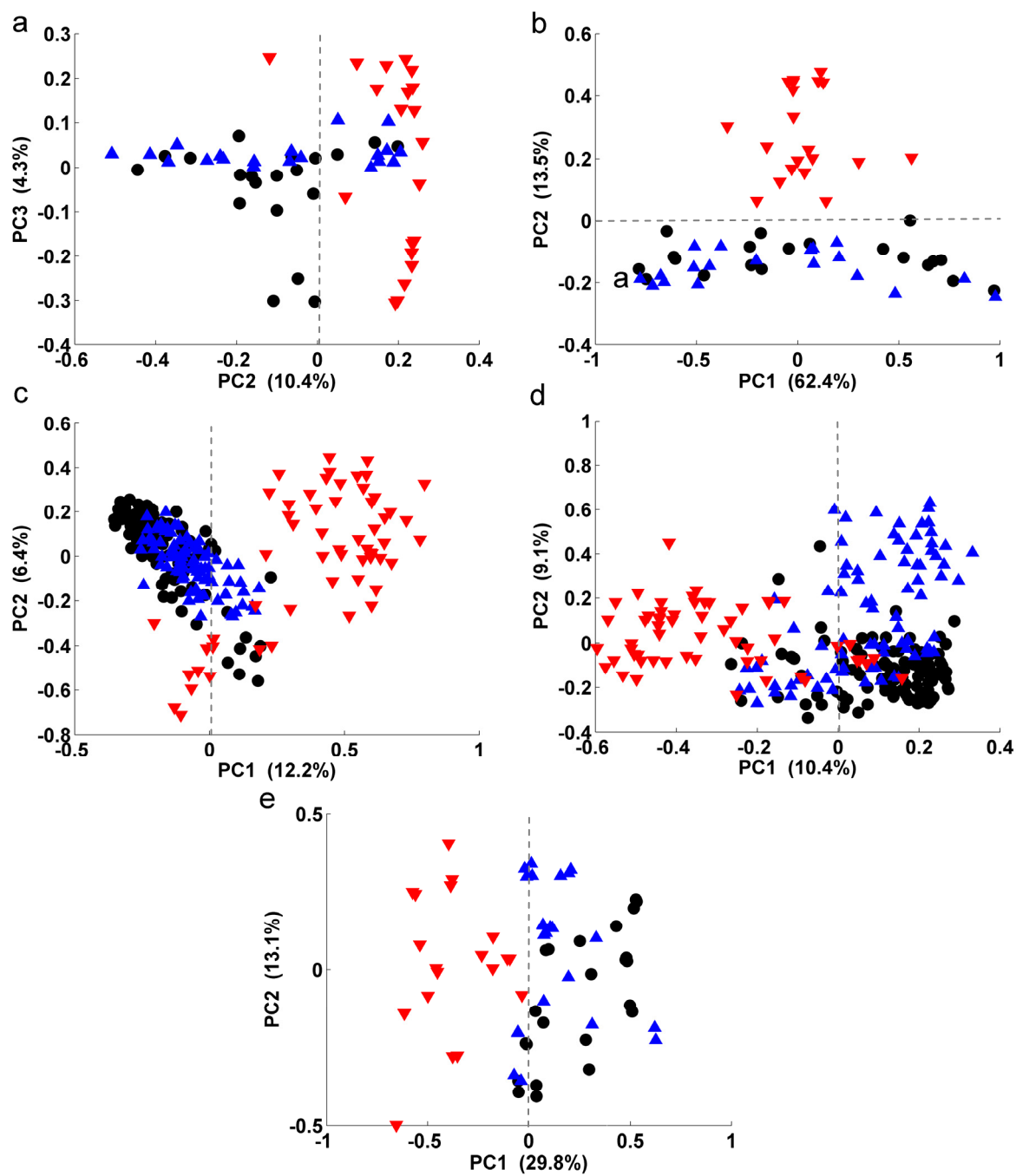
Figure 4.11 │ The PCA score plots of raw data and RMieS corrected data for method II FTIR monolayer spectra (a, b) respectively; method III SR-FTIR single cell spectra (c, d); method II FTIR-PAS monolayer spectra (e)

The single cell spectra are most difficult to interpret as they include intra-cell type variation; each measured cell may be slightly different due to biochemistry associated with cell phase or viability and have complex mean derivative profiles (Fig 4.12c) and this is reflected in the corresponding loading plot (Fig. 4.12d).

Analysis of the PCA loading plots generally suggest strong variance in the amide I region of the different cell spectra. This amide I variation is clearest both in the loading plot for the FTIR monolayer data (Fig. 4.12b), the dataset benefiting from the greatest signal to noise in the second derivative form (Fig 4.12a).

The fact that the 2245/6R spectral cluster lies in negative PC2 space whilst the 2247R cell spectral cluster in positive space of PC2 (Fig. 4.11b), suggests that these two groups are anti-correlated and that 2247 will be different in amide I composition, relative to cell lines 2245/6R which appear relatively similar. Due to the differences in these PCA loading plots, subsequent analysis focused on the protein compositions of the cell lines.
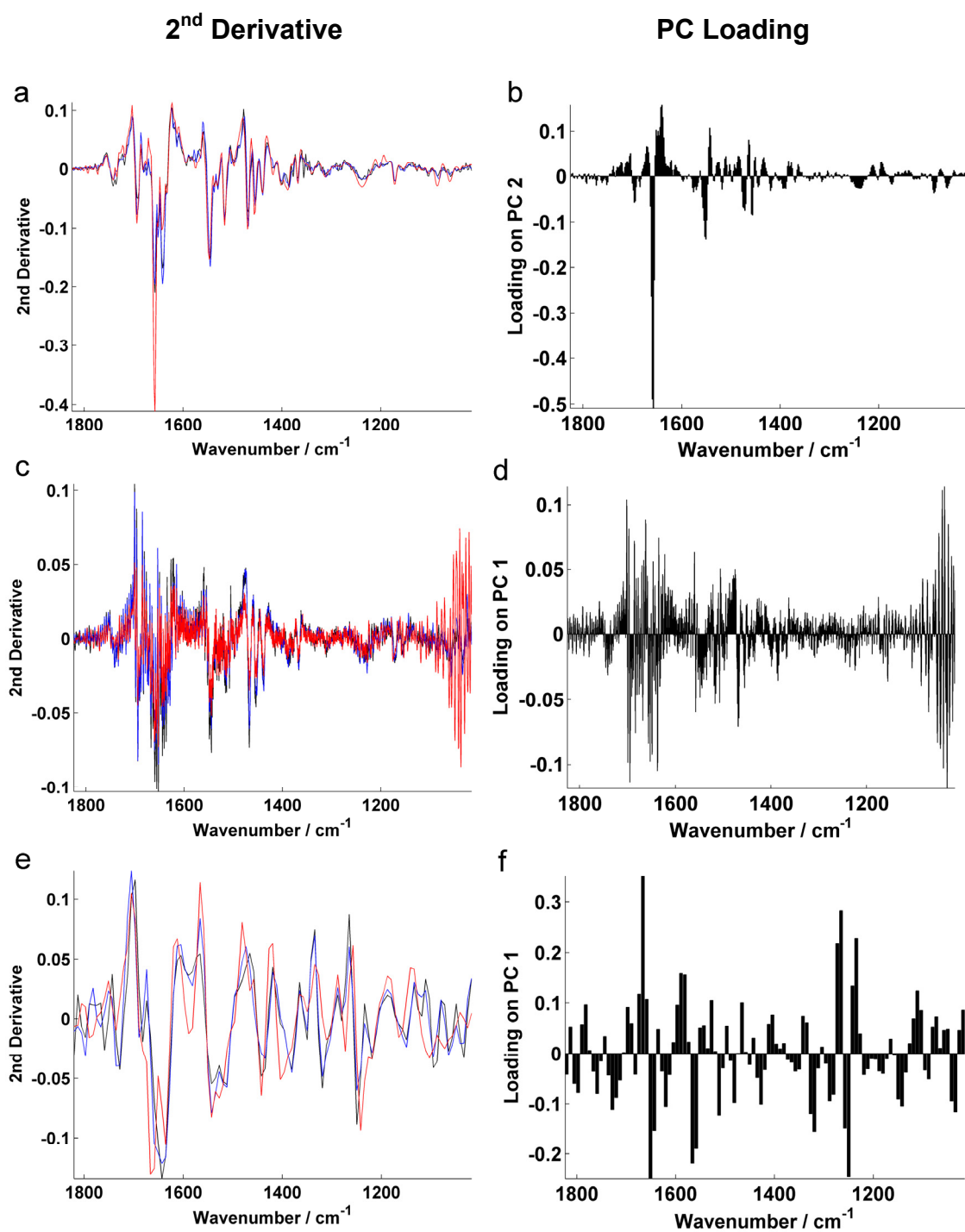
Figure 4.12 │ **The mean 2ⁿᵈ derivative spectra of the cell lines and the resulting PCA loading plots for method II FTIR monolayer spectra (a, b) respectively; method III SR-FTIR single cell spectra (c, d); method II FTIR-PAS monolayer spectra (e, f)**

4.3.5 │ Estimation of protein secondary structure

Protein secondary structure was analysed using the mean spectrum for each cell line from the Method II, the FTIR monolayer dataset. Deconvolution and curve fitting techniques were employed to maximize the amount of information retrieved from the complex amide I region. Voigt curve-fitting of the amide I band was used to give the best estimation of protein secondary structure, as the fractional areas of the fitted component bands are directly proportional to the relative quantity of structure they represent. The band centres were approximated to the frequencies observed in the deconvoluted spectra. The fitted curves were checked by second derivative analysis, whereby underlying peaks and shoulders should be resolved at the locations of negative peak minima in the derivative. Band widths and heights were allowed to attain best fit values by iteration (the resulting composite spectrum) (Fig. 4.13). The bands were fitted iteratively and a composite spectrum was created so that closely matched the FSD spectrum. Tentative band assignments were made in reference to bands characterised in isolated proteins [20].
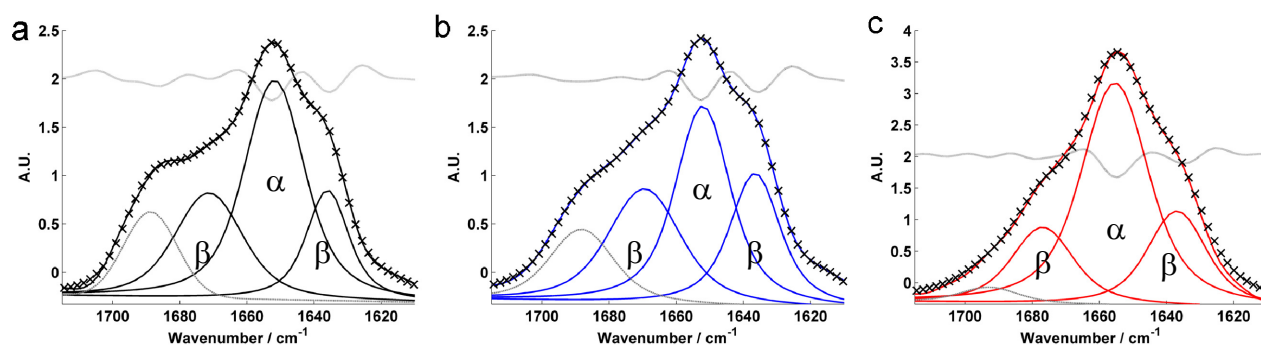
Figure 4.13 │ **The amide I secondary structure curves displaying alpha and beta components matched to the superimposed 2nd derivative spectrum at 2.0 units. In determination of the overall α/β ratio, the fitted curve associated with turns and bends ~1690 cm^{-1} (in grey) was not considered. For each cell type, the fitted composite spectrum (crosses) closely matched that of the original deconvoluted spectrum (solid line).**

The results (Table 4.1, Appendix) estimated a higher occurrence of α-helical contributions in the 2247R cell line compared to cells lines 2245/6R, which were by contrast richer in β-structure.

Analysis of protein secondary structure by means of Fourier self-deconvolution and curve fitting is only an estimate in samples of such complex biochemistry. The assignment of protein secondary structure is not straightforward process and can only be applied as a general rule, based on model calculations and previous findings from known proteins. Model calculations on the amide I bands of globular proteins have demonstrated that the spectral contributions of structural elements other than alpha and beta components, such as random coils turns and bends are spread over a wide wavenumber region.

A number of experimental issues are presented by use of band-narrowing and curve-fitting procedures. Secondary structure prediction accuracy can vary based on the band narrowing parameters used [21]. In both FSD and 2nd derivative techniques, noise is amplified and noise can be mis-interpreted as a real band [22]. A low enhancement factor in the FSD and low sensitivity in the subsequent curve fitting parameters were used to avoid this issue.

In order to demonstrate the difference with high sensitivity curve-fitting, the results were re-processed, resulting in an increased number of fitted-bands (Fig. 4.14). This is a more in-depth estimation of the nature of the protein secondary structure and these bands do have references for not only alpha helices and beta sheets, but also for random coil, and turns and bends. When analysed in terms of percentage expression (Fig. 4.15), the results were in agreement with the initial curve fitting using the low sensitivity parameters. The major caveat, however, is that the references for these bands are derived from pure protein samples. In an infrared spectrum of a complex biological cell, there will be underlying bands from macromolecules other than protein signals that are present in the same region. This makes any estimation of protein secondary structure difficult to substantiate.
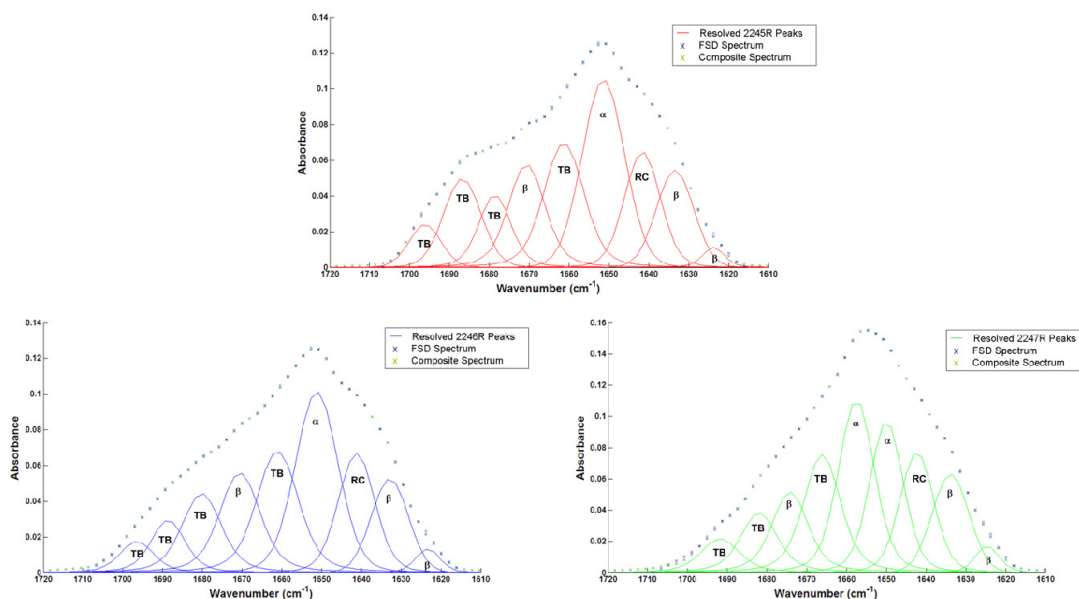
Figure 4.14│The repeated amide I secondary structure curves fitted with a higher sensitivity. For each cell type, the fitted composite spectrum closely matched that of the original deconvoluted spectrum, however in this instance there are more fitted bands. These bands do correlate with the spectral wavenumber positions of known alpha helix, beta sheet, random coil, turns and bends, however the reference material for these positions is comprised of protein alone [20].
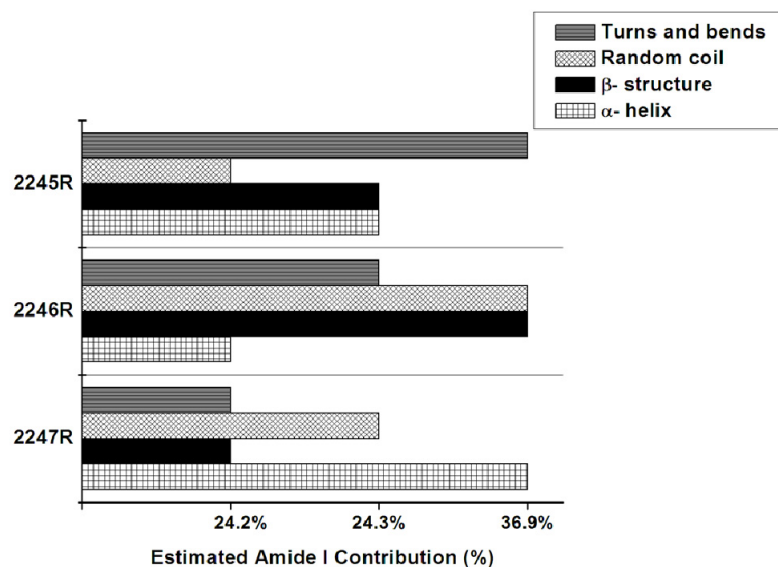
Figure 4.15│ High sensitivity-fitted bands were analysed as a percentage of α-helix, β-structure, turns and bends, estimated by addition of the areas of all the component bands assigned to each of the structures and then expressing the sum as a fraction of the total amide I area. Increased expression of α-helix structure in cell line 2247R agreed with the first instance curve fitting with low sensitivity (Fig. 4.14).

Protein analysis by FTIR does give the advantage that whole cells can be probed with very little sample preparation and with little disruption to the cell-biochemistry. It was noted, however, that underlying contributions from other biochemical molecules, such as nucleic acids can display absorbance peaks within the amide I wavenumber region [1]. It was recognised, therefore, that the analysis of protein secondary structure by this process is not ideal.

It was therefore important to investigate the cell line protein difference by other means in order to validate such a result. This led to further investigation by means of liquid chromatography mass spectrometry and Western blot.

4.3.6 │ Proteomic data analysis

The MS dataset of 58225 features, comprised of both identified peptides and non-identified molecular entities. A list of these features was generated in Progenesis with information regarding mass to charge (m/z), retention time (min) the retention time window (min), charge, significance (anova P value) and maximum fold change, the later describing the difference in feature abundance across all cell types. For example, if the abundance level of feature X was (20, 20, 40) in 2245R, 2246R and 2247R respectively, the fold change would be (1, 1, 2).

The normalized relative abundance for all feature data and each individual injection was subjected to PCA for visual protein discrimination between the cell lines. A duplicate lysate of one cell line (2245R) was prepared to demonstrate sample preparation and MS instrumentation robustness (Fig 4.16). In a similar stance to the FTIR data, 2247R was in contrast from cell lines 2245/6R in terms of feature abundance.
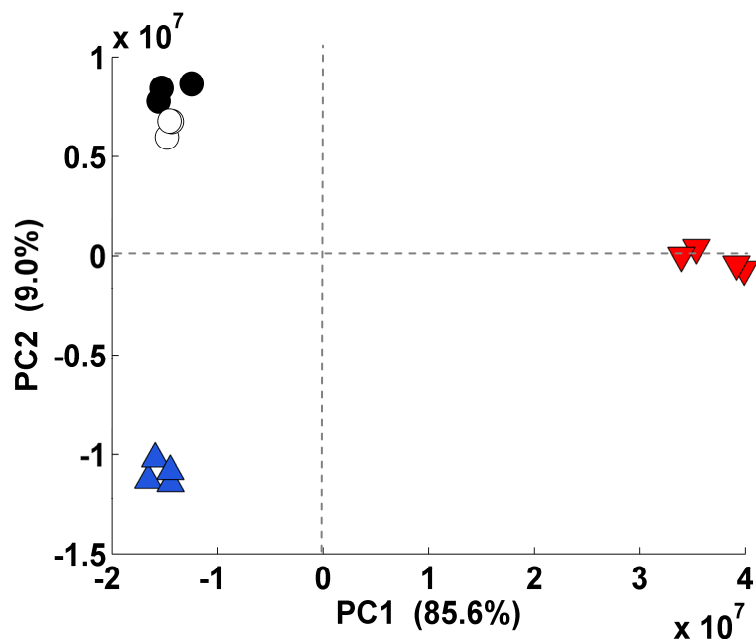
Figure 4.16 │ The PCA score plot of peptide feature abundance. Every data point represents a single LCMS injection for 2245R (black circle); 2246R (blue triangle) and 2247R (red down-triangle). A technical lysate replicate was made for cell line 2245R (white circle) showing good methodological consistency by grouping with its 2245R counterpart cluster. Cell line 2247R protein lysate injections are notably anti-correlated with the other two cell lines in PC1.

The normalised injection replicates of MS peptide feature abundance were averaged for each cell line and mean centred (Fig. 4.17). The data was analysed separately for each cell line for features that were 2.5×10^6 more/less abundant than the group mean. The reasoning behind this chosen threshold was to select for a snapshot of features that were most different from mean abundance and therefore most discriminatory. These features were matched to peptide sequence and then correlated to a protein from the feature data list (Table 4.2, Ch.9, Appendix).
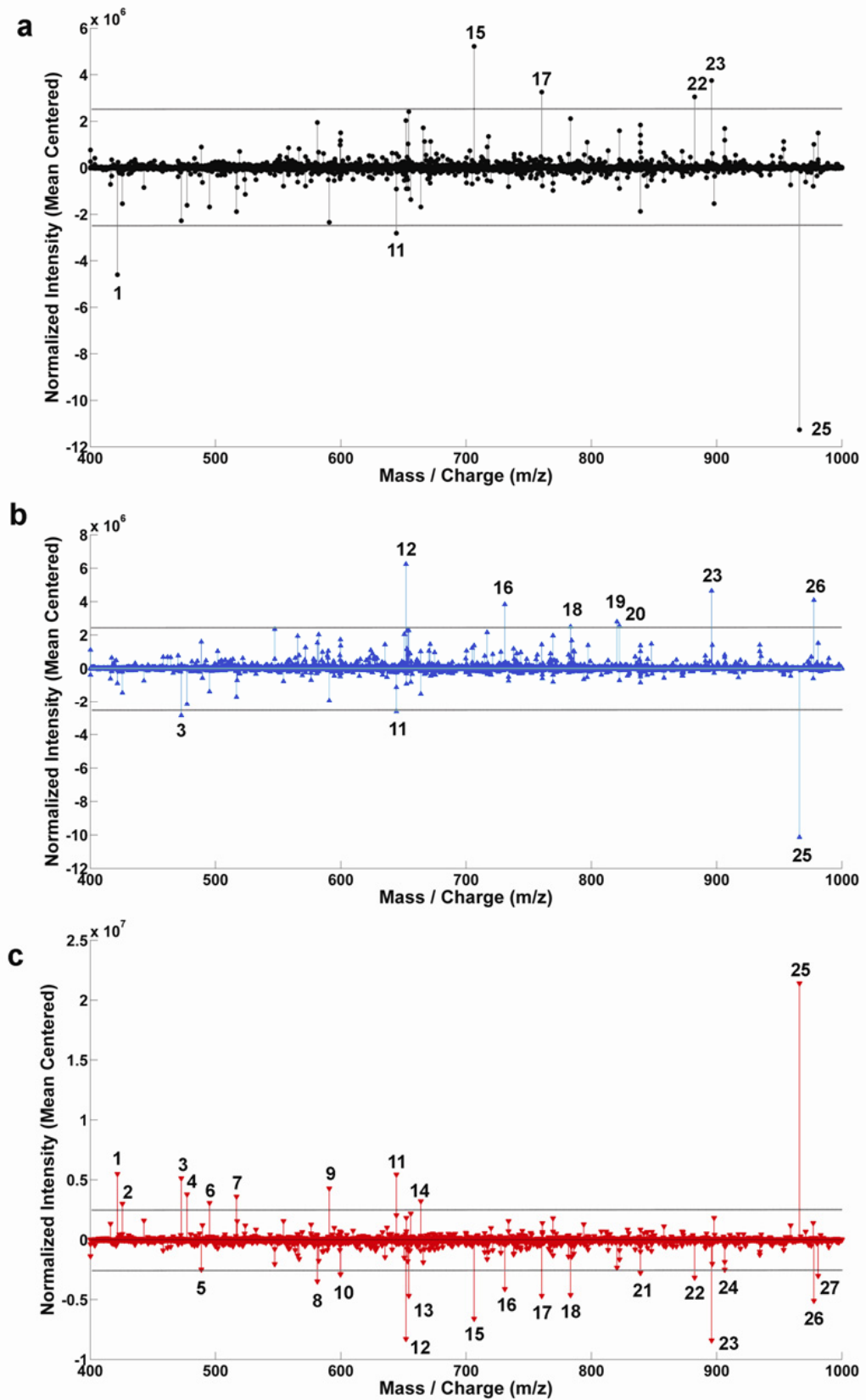
Figure 4.17 | Analysis of peptide feature abundance related to Figure 4.13. Each peptide feature was mean centred for 2245/6/7R (a, b, c respectively). As the dataset contained thousands of features, a set of 27 features which were furthest in distance from the mean were selected for closer evaluation.

The protein was cross-checked with the MSMS protein abundance fold change list and secondary structure contributions analysed (Table 4.3, Ch. 9, Appendix).

To create the MSMS protein abundance list, a total of 2700 proteins were identified from MSMS sequencing data of 10,000 peptide features. Using $p < 0.05$, protein abundance fold change ×2 or greater, the list was filtered down to 1485 proteins. 681 proteins in 2247R and 666 proteins in 2245/6R were higher in terms of mean protein abundance. 137 proteins were undefined using the strict selection rule that abundance of proteins in 2247R had to be lower or higher and not equal to the abundance in both 2245R and 2246R. This was to identify discriminatory proteins which may have caused the PCA profile of cell line 2247R to be distinctly different.

Databases containing protein secondary structure information were used to determine the level of alpha/beta structure. For proteins yet to be structurally characterised, secondary structure was predicted (Table 4.3, Ch. 9, Appendix). A total of 4 methods of structural identification were used to predict secondary structure with descending levels of confidence whereby the next best possible method would be used if protein secondary structure information was not available for the primary method.

The 'Primary' method with highest level of confidence information from protein structures found in the DSSP is derived directly from the Protein Data Bank (PDB). This data makes the assumption that the protein IMDH2_HUMAN for example will be homologous to IMDH2_MOUSE, i.e. they share a common ancestor protein, and their structure will be constrained because their function will depend on their structure. Differences may have been introduced but in a sequence that does not affect structure. Therefore where structure of the human protein is not available, structure from other species were used.

The 'Secondary' method used alternate identifiers which are results gained from a search using different identifiers of the same protein. The 'Tertiary' method uses available protein structure results from the Basic Local Alignment Search Tool (BLAST). BLAST searches find homologous proteins with more than 85% identity.

Finally the 'Quaternary' method uses results gained from trying to predict secondary structure 'de novo', i.e. predicted it from sequence alone.

Differences in peptide feature abundance analysed in Figure 4.17 are summarised below (Fig. 4.18). There were 27 peptide features that were notably different in abundance from the group mean. The most discriminatory feature was detected at 966.094 m/z ('Feature 25' in Fig. 4.17, 4.18). This feature was notably 'up' in 2247R expression and considerably 'down' in both 2245 and 2246R. The best peptide match score was 87.07% and was correlated to the protein B2R5B3_HUMAN (Table 4.2, Appendix). A protein search in the MSMS ID list for 2247R (Previously defined as proteins with twice or more as abundant in 2247R than 2245/6R) equated to a histone H2A fragment, and similarly other features of notable abundance also highlighted a large difference in subunits H2A (Table 4.3, Appendix).
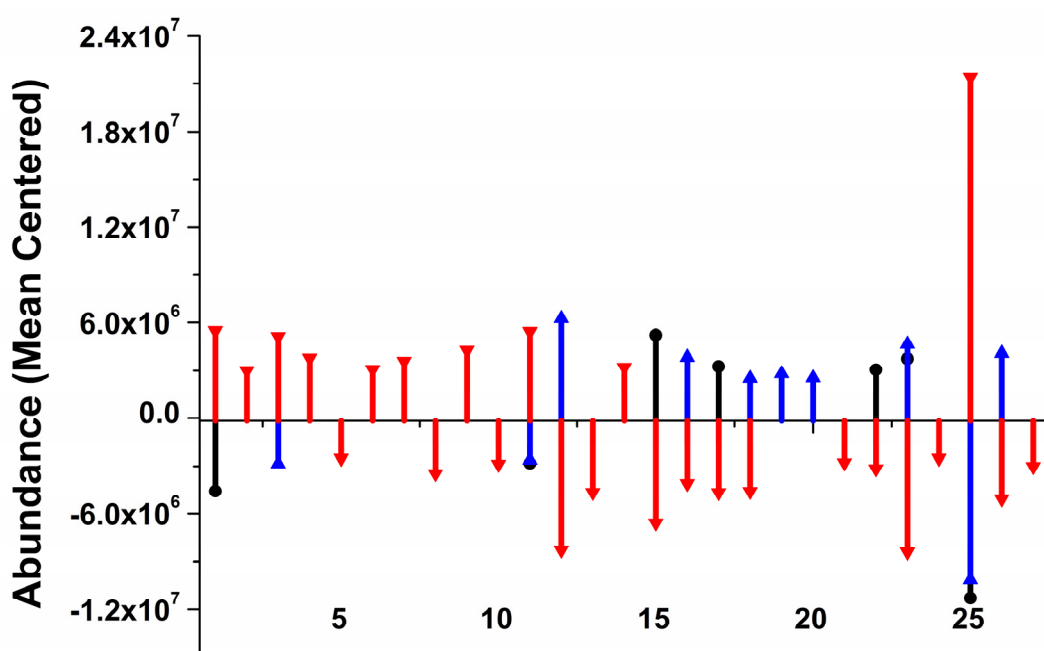
Figure 4.18 | Summary of MS peptide abundance for selected features furthest from the group mean. Details of each feature is listed in Table 4.2 in the Appendix (Ch. 9).

4.3.7 │ Proteomic validation by Western blot

To confirm the validity of the label-free method of quantisation and subsequent protein identification, a Western blot confirmed protein identification of histone H2A and its levels of expression in the cell lysates following standard protocol. This was performed in triplicate for differing protein loading levels. Results confirmed histone H2A expression in cell line 2247R to be extensively greater (Fig. 4.19).
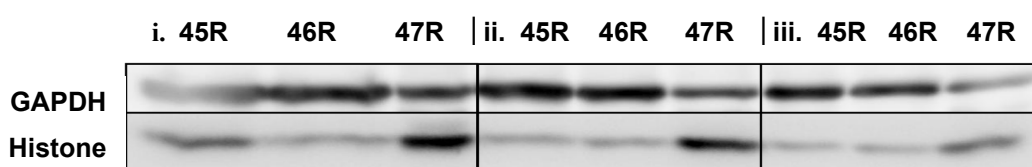
Figure 4.19 | Western Blot results: each cell line was loaded with (i) 27.2 μg (ii) 13.7 μg and (iii) 6.7 μg of protein. GADPH expression was used as a control to assess the consistency of the protein loading. Higher levels of histone H2A expression were observed for cell line 2247R.

4.3.8 │ Correlating FTIR and proteomic results

For all associated proteins that were identified with secondary structure available, the mean alpha and beta components were ratioed (Table 4.4, Ch. 9, Appendix).

The ratios of α/β contribution to overall secondary structure from the proteomic analysis were compared with the FTIR prediction (using the initial, low-sensitivity curve-fitting parameters). Both methodological approaches resulted in the same trend of estimated protein secondary structure contribution (Fig. 4.20).
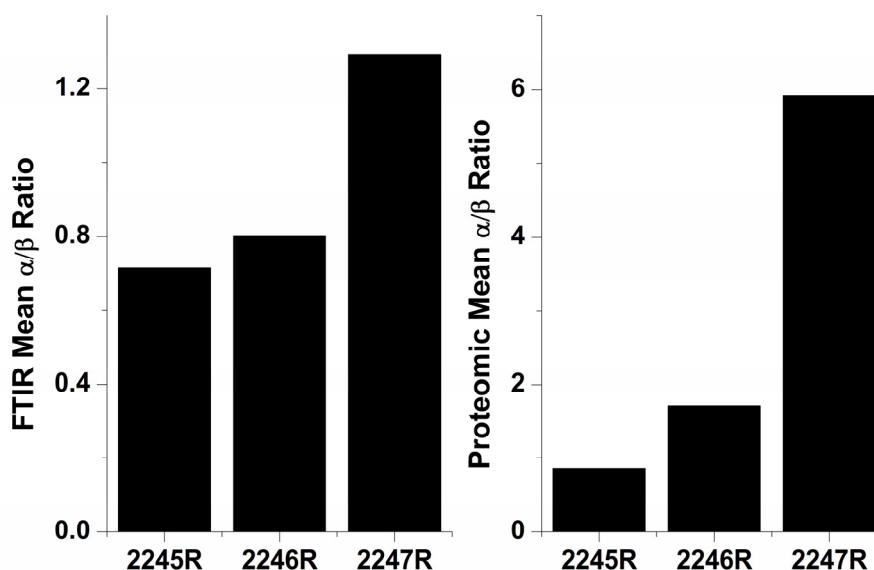
Figure 4.20 | Comparing the α/β ratio results of general protein secondary structure for FTIR (left) and LCMS (right).

4.4│Conclusions and Future Prospects

The main aims of the investigation were two-fold. The initial aim was to use FTIR spectroscopy as a tool to characterise cell lines, not by their physical scattering properties, but by true biochemical difference, an aim that was previously somewhat questionable. Secondly, it was important to link and validate conclusions from FTIR spectroscopy to other methods, to increase perspective.

With the use of chemometrics, the RMieS-EMSC corrected FTIR data, global cellular chemistry of three renal carcinoma cell lines were focused down to a notable protein difference in cell line 2247R, indicating protein secondary structure differences. Estimation of these differences predicted a larger abundance of alpha helical arrangements in cell line 2247R and comparatively larger β-structure in 2245/6R.

Using state of the art label free proteomics, whole cell lysates were analysed in a snapshot of most significant proteins were identified with protein secondary structure.

Despite both techniques using some degree of estimation, generally both approaches resulted in the same trend of alpha/beta ratio, estimated from protein secondary structure contributions.

The field of cancer biomarker detection has been likened to 'finding the needle in the haystack' [23]. It is therefore important to use multiple viable methodologies to fully understand complex processes, and in the same instance self-validate findings.

It is hoped that this investigation has helped to prove the legitimacy as FTIR spectroscopy as a tool that can be used to aid biomarker discovery in a clinical setting.

For example, it has been noted in this study that the glycogen/phosphate ratio is a good discrimination marker for different cell lines. Further work on this ratio may include progression to primary cell lines, resected from patients to correlate the ratio with increasing malignancy grade in renal carcinoma.

4.5│References

1. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

2. Gazi E, Dwyer J, Lockyer NP, Gardner P, Shanks JH, Roulson J, Hart CA, Clarke NW, Brown MD: **Biomolecular profiling of metastatic prostate cancer cells in bone marrow tissue using FTIR microspectroscopy: A pilot study**. *Analytical and Bioanalytical Chemistry* 2007, **387**(5):1621-1631.

3. Krishna CM, Sockalingum GD, Vadhiraja BM, Maheedhar K, Rao ACK, Rao L, Venteo L, Pluot M, Fernandes DJ, Vidyasagar MS et al.**: Vibrational spectroscopy studies of formalin-fixed cervix tissues.** *Biopolymers* 2007, **85**(3):214-221.

4. Chowdary MVP, Kalyan Kumar K, Mathew S, Rao L, Murali Krishna C, Kurien J: **Biochemical correlation of Raman spectra of normal, benign and malignant breast tissues: A spectral deconvolution study**. *Biopolymers* 2009, **91**(7):539-546.

5. Walsh MJ, Nasse MJ, Pounder FN, Macias V, Kajdacsy-Balla A, Hirschmugl C, Bhargava R: **Synchrotron FTIR imaging for the identification of cell types within human tissues**. In: *AIP Conference Proceedings:* 2009; Banff, AB; 2009: 105-107.

6. Peng Q, Xu Y, Li W, Wu J, Zhou X: **FTIR study on the normal and tumor gastrointestinal tissues.** *Guang pu xue yu guang pu fen xi = Guang pu* 1998, **18**(5):528-531.

7. Carmona P, Rodriguez-Casado A, Alvarez I, de Miguel E, Toledano A: **FTIR microspectroscopic analysis of the effects of certain drugs on oxidative stress and brain protein structure.** *Biopolymers* 2008, **89**(6):548-554.

8. Gasper R, Mijatovic T, Benard A, Derenne A, Kiss R, Goormaghtigh E: **FTIR spectral signature of the effect of cardiotonic steroids with**

antitumoral properties on a prostate cancer cell line. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 2010, **1802**(11):1087-1094.

9. Lamberti A, Sanges C, Arcari P: **FT-IR spectromicroscopy of mammalian cell cultures during necrosis and apoptosis induced by drugs**. *Spectroscopy*, **24**(5):535-546.

10. Draux F, Jeannesson P, Gobinet C, Sulé-Suso J, Pijanka J, Sandt C, Dumas P, Manfait M, Sockalingum GD: **IR spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells**. *Analytical and Bioanalytical Chemistry* 2009, **395**(7):2293-2301.

11. Gaudenzi S, Furfaro MG, Pozzi D, Silvestri I, Castellano AC: **Cell-metal interaction studied by cytotoxic and FT-IR spectroscopic methods**. *Environmental Toxicology and Pharmacology* 2003, **14**(1-2):51-59.

12. 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**(8):1586-1593.

13. Erukhimovitch V, Talyshinsky, M., Souprun, Y., Huleihel, M.: **Spectroscopic investigation of herpes simplex viruses infected cells and their response to antiviral therapy**. *Journal of Molecular Structure* 2006, **792-793**(3):99-103.

14. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, Gazi E, Brown M, Clarke N, Gardner P: **Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples**. *Analyst* 2010, **135**(2):268-277

15. Harvey TJ, Gazi E, Henderson A, Snook RD, Clarke NW, Brown M, Gardner P: **Factors influencing the discrimination and classification of prostate cancer cell lines by FTIR microspectroscopy**. *Analyst* 2009, **134**(6):1083-1091.

16. Mordechai S, Sahu, R.K., Hammody, Z., Mark, S., Kantarovich, K., Guterman, H., Podshyvalov, A., Goldstein, J., Argov, S.: **Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma**. *Journal of Microscopy* 2004, **215**(1):86-91.

17. Harvey TJ: **The Development of Vibrational Spectroscopic Cytology for Prostate Cancer Diagnosis**. *The University of Manchester*; 2008.

18. Dumas P, Polack F, Lagarde B, Chubar O, Giorgetta JL, Lefrancois S: **Synchrotron infrared microscopy at the French Synchrotron Facility SOLEIL**. *Infrared Physics and Technology* 2006, **49**(1-2):152-160.

19. Gazi E, Dwyer, J., Gardner, P., Ghanbari-Siahkali, A., Wade, A.P, Miyan, J., Lockyer, N.P., Vickermann, J.C., Clarke, N.W., Shanks, J.H., Scott., L.J., Hart, C.A., Brown, M.: **Applications of Fourier transform infrared microspectoscopy in studies of benign prostate and prostate cancer. A pilot study.** *Journal of Pathology* 2003, **201**(1):99-108.

20. Stuart B: **Biological Applications of Infrared Spectroscopy**: *John Wiley & Sons*; 1997: p119

21. Surewicz WK, Mantsch HH, Chapman D: **Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment.** *Biochemistry* 1993, **32**(2):389-394.

22. Grdadolnik J: A **FTIR Investifation of Protein Conformation**. *Bulletin of the Chemists and Technologists of Macedonia* 2002, **21**(1):23-34.

23. Danovi SA: **Biomarkers: Finding the needle in the haystack.** *Nature Review Cancer* 2008, **8**(September):659.

Chapter 5 │ Discrimination of Renal Cancer cells Displaying Stem Cell Characteristics

5.1 │Introduction and Aims

As previously discussed, there is a hypothesis that cells displaying cancer stem cell properties exist (Ch. 1.3.2). Adult stem cells and putative cancer stem cells are rare and due to the lack of specific stem cell markers, making isolating and subsequent characterisation of these cells challenging.

One method to enrich for cells which possess many of the characteristics of stem cells [1-3] is the Hoechst 33342 dye efflux assay developed by Goodell et al. in the haematopoietic system and has been adapted to solid tissue tumours and cell lines in the isolation of a 'side population' (SP).

Uptake of the DNA-binding dye Hoechst 33342 is universal in a cell population, however, the ability to efflux the dye is restricted to a sub-population. The efflux process can be blocked with Verapamil, a calcium channel blocker which inhibits the ABCG2 transporter system. Without Verapamil, effluxing-able cells can pump out dye, reducing the fluorescent signal. The stem-like cells are called the 'Side Population' (SP) cells as they appear on the 'side' of the main cellular population in the FACS profile (Fig 5.1).
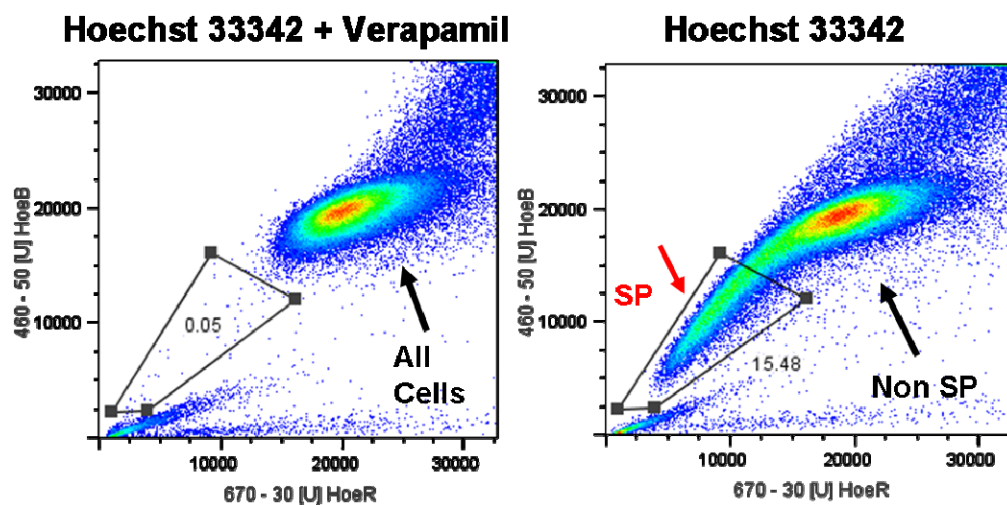
Figure 5.1│ Hoechst 33342 profile of the 2245R renal clear-cell carcinoma epithelial cell line. A typical Hoechst 33343 red/blue profiles of 2245R cells stained with Hoechst 33342 prior to sorting on a FACS Vantage cell sorter. SP gate was defined using the SP inhibitor Verapamil. The SP can be sub-sorted into proximal (PSP) and distal (DSP) side populations.

The SP itself is enriched for these stem-like cells, but is still heterogeneous. A further sub-fraction of cells with greater efflux-ability can be isolated. These appear further down the tail of the SP FACS profile. The 'distal' SP cells (DSP) are said to be more primitive, hence more closely resembling true stem cells [4].

Characterisation of the SP subpopulations is problematic enough due to the very low numbers of cells within each subpopulation, often below the level required for traditional cellular and '-omic' techniques. Characterising these sub- side population cells is a further challenge, yielding even lower quantities. The use of SR-FTIR provided a way to analyse small numbers of single cells. The aim was to generate spectral chemical profiles of these small cellular sub-populations which could possibly provide more insight into the nature of these rare cells.

5.2 │ Experimental

5.2.1 │ FACS Sorting of cell populations

The FACS sorting and isolation of the cell populations and sub-populations was kindly performed by M. Liew in the Genito-Urinary Research Group at the Paterson Institute of Cancer Research, University of Manchester.

The 2245R renal carcinoma cell line was cultured in DMEM with FCS (10%) and L-glutamine (1 %) at 37˚C in a humidified atmosphere (5% CO_2) until 80% confluent. Hoechst 33342 staining was conducted according to the method of Addla et al. [40]. Briefly, $1x10^6$ 2245R cells in 1 ml Hoechst buffer (Hank's buffered salt solution, 10% FCS, 20 mM HEPES, and 1% D-glucose) were stained with 5 mM Hoechst 33342 for 90 min at 37°C with continuous agitation (5 Hz). Cells were washed and re-suspended in ice-cold Hoechst buffer prior to FACS analysis. Hoechst 33342 dye efflux SP was defined by 50 mM Verapamil hydrochloride blockade of Hoechst 33342 dye efflux. Hoechst 33342 staining was detected using a Becton Dickinson FACS Vantage SE flow cytometer (FACS), exciting at 357 nm and detecting Hoechst Blue with a 424/44 broad pass (BP) filter and Hoechst Red

with a 675/20 BP filter (Omega Optical, Brattleboro VT) SP, as well as distal- and proximal-SP, and non-SP cells were 'sorted' based on pre-defined profiles.

5.2.2 │ Sample preparation for spectroscopy

Following cell sorting, the isolated cell populations were fixed in formalin (4 % in PBS). The cells were subsequently spun and re-suspended in 500 μL of PBS. The cells were then mixed to form a homogeneous suspension and 150 μL aliquots were used for deposition with a cell cytospinner onto CaF_2 substrates, as described in Ch. 2.4.4.2. The cells were initially spun at a gentle 400 g for 5 minutes, then additionally at 800 g for 5 minutes. The substrates were left overnight to dry before washing with distilled water to remove residual salt, originating from the PBS.

5.2.3 │ Data acquisition

SR-FTIR transmission spectra were acquired on the SMIS beamline of the French Synchrotron Facility SOLEIL (Saint-Aubin, France) [5] using the instrumentation setup described in Ch. 2.4.5.2. Spectra were recorded at 4 cm^{-1} resolution with 256 co-scans. A total of 25 spectra were recorded for each cell type.

5.2.4 │ Data analysis

The raw spectra were corrected with an ATR Matrigel reference spectrum and 100 iterations of the RMieS-EMSC algorithm [6-7]. The corrected spectra were transformed to the second derivative with 7 point Savitsky-Golay smoothing with a polynomial order of 3 in Omnic 7.2 (Thermo Electron Corporation).

Analysis was performed on the second derivative corrected spectra with a range of 1475-1000 cm^{-1} to achieve maximum chemometric separation between the Non-SP and the SP cells. The corrected spectra were then vector normalised and mean-centred before principal component analysis (PCA), performed in Matlab (The Mathworks Inc.). PC-LDA models were generated using SPSS (SPSS Inc.).

5.3 │ Results and Discussion

5.3.1 │ Sample assessment

The Hoeschst dye efflux model was used to enrich populations of stem-like cells prior to further characterisation and analysis. The dye is known to be effluxed out of the cell after several hours or days. At the 1-2 hour time point, however, cells would still contain certain levels of dye allowing them to be sorted by their fluorescent signals.

The FACS blot (Fig. 5.1) shows an SP 'Tail' containing cells with various concentrations of dye within. If Hoechst 33342 dye was no longer present their signal would fall within the bottom left hand corner of the plot of the cell FACS profile threshold and would not be collected for the study.

The Hoechst 33342 red blue FACS profile generated by 2245R displayed a Verapamil sensitive SP containing up to 16% (average 6%) of the renal epithelial cells. This enabled the isolation of 3 subpopulations for analysis; the non-SP which contained differentiated cells and two SP sub-populations. The subpopulations were proximal side population (PSP), containing putative transit amplifying cells and the distal side population (DSP) which is thought to contain the most primitive (stem) cell types.
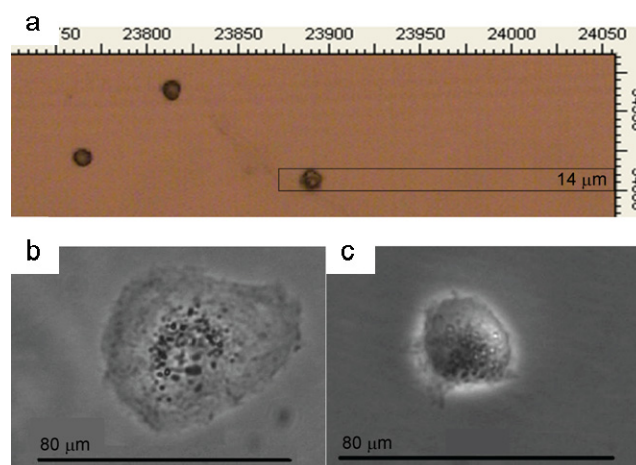
Fig. 5.2 │ (a) Isolated FACS sorted cells, chemically fixed immediately after sorting for Synchrotron experiment, spun onto CaF$_2$ slides (b) in vitro photomicrographs of NSP and (c) SP cells after 24 hours of cell culture post FACS sorting.

Figure 5.2 compares centrifuge cytospun Non-SP cells onto CaF$_2$ slides and photomicrographs of the cell types in vitro, after 24 hours of cell sorting. Cellular differentiation is illustrated by morphological changes that may occur if the cells are not fixed and instead cultured. The Non-SP cells (Fig. 5.2b) begin to extend their cytoplasmic boundaries, a characteristic morphology of the clear-cell renal carcinoma phenotype. The SP cells remain spherical in morphology more in keeping with a stem cell phenotype (Fig. 5.2c).

The cells were chemically fixed in formalin (4%) immediately after FACS (Fig. 5.2a) for the FTIR experiments. By doing so, the cellular content at the time of sorting was preserved and no further differentiation could occur before the spectral fingerprint acquisition. The cells are analysed in a 'snapshot' of time, representing a controlled and fair comparison. All cell types analysed were typically of the same size and appeared spherical by plating the cells onto a CaF$_2$ substrate at a predetermined force so that the cells remain intact whilst attaching to the substrate.

5.3.2 │ SR-FTIR assessment

It is also understood that even small differences in cell size may dominate chemometric discrimination, as different sizes of cells would correspond to different resonant Mie scattering profiles which would be detected in chemometric discrimination. This concern, however, was reduced with use of the RMieS-EMSC correction algorithm [6-7].

Figure 5.3 presents the mean infrared spectrum with standard deviation of each type of sorted cell. The non-side population displayed the greatest spectral variance, particularly in the 1250-1000 cm^{-1} region. This is understood to represent the heterogeneity of the non-side population, with single cells more likely than the SPs to be present at different stages of the cell cycle at the time of FACS sorting.
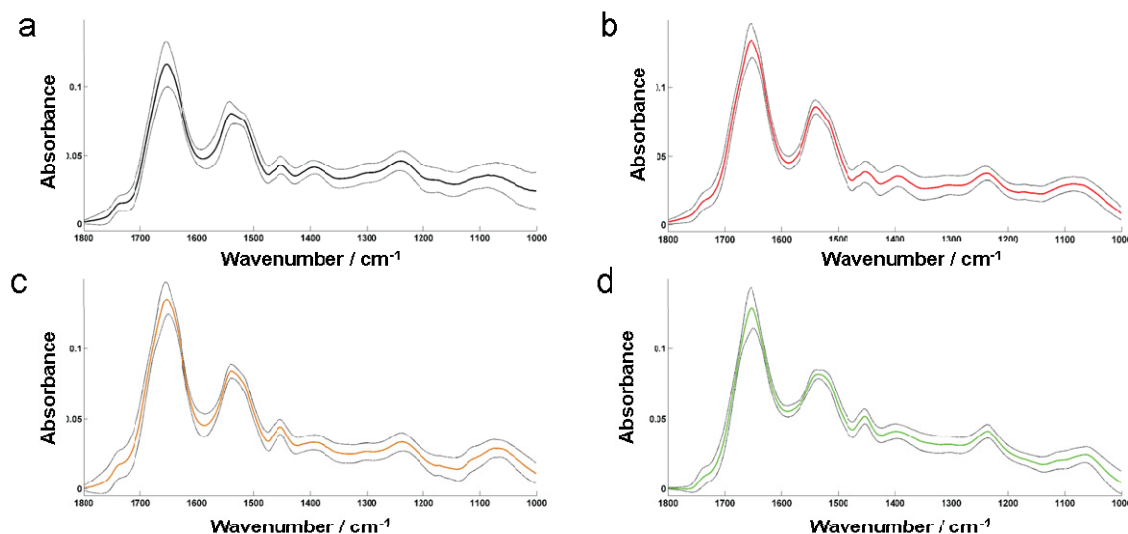
Fig. 5.3 │ **SR-FTIR mean fingerprint spectra +/- standard deviation from single cells of (a) non-side population (b) side population, (c) sub-side population; proximal, (d) sub-side population; distal. For multivariate analysis the range was reduced to 1475-1000 cm^{-1} and the spectra transformed to the second derivative.**

5.3.3 │ PCA of the total dataset

Changes in cell cycle phase can influence the overall chemical signature of the individual cell [8-10]. This is also apparent in the PCA score plot of the total dataset (Fig. 5.4a) where the non-SP cell spectra are more dispersed than the side population cells of DSP, PSP and SP. PCA separations of different cell lines is more straightforward than discriminating differences in a single cell line as there is partial overlap between the non-SP and SP spectra, as would be expected.

It is apparent, however, that there is some degree of separation between non-SP and SP spectra in PC3. The SP contained a mixture of DSP and PSP cells and therefore it was expected that the SP cluster would fall on the same trajectory as the DSP and PSP spectra. The associated loading plot for PC3 is displayed in Fig. 5.4b.

As second derivative peak minima are directly aligned to the centre of the absorbance peak of the original spectra, the loadings can be interpreted as biological references. There are many features that contribute to the differences between the non and side-populations.
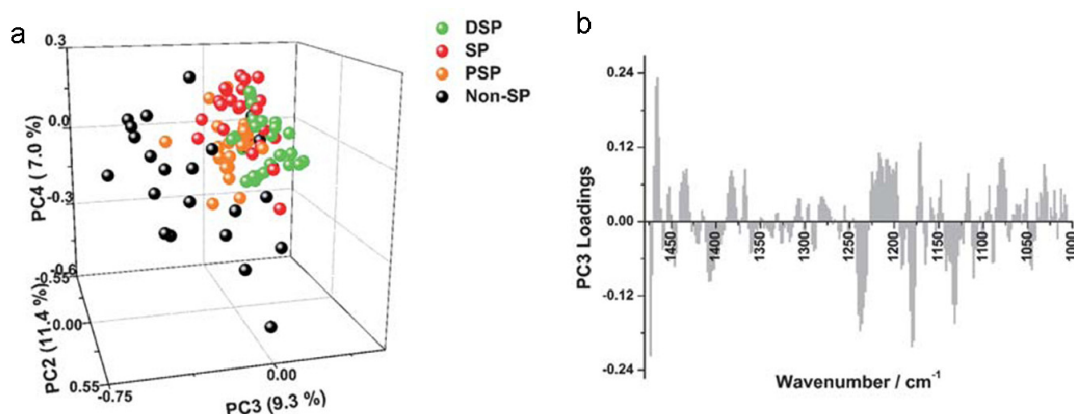
Fig.5. 4│Total dataset analysis (a) PCA plot of PC2, PC3 and PC4 (b) The loadings of PC3. Discrimination of the Non-SP cells can be seen through principal component 3.

These include: $\delta_{as}(CH_2)$ scissoring vibration of lipids at ~1468 cm^{-1}; $\nu_sC=O$ (of COO-) at ~1400 cm^{-1}; $\delta_s(CH_3)$ at ~1380 cm^{-1}; $\nu_{as}P=O$ (of PO$_2^-$) of phosphodiester group of phospholipids, phosphorylated proteins and nucleic acids (arising from the phosphate sugar backbone of DNA and RNA) at ~1237 cm^{-1}; Amide III band protein components at ~1240-1310 cm^{-1}, ν_{as}(CO-O-C) ring vibrations of carbohydrates at ~1180 cm^{-1}; $\nu_sP=O$ (of PO$_2^-$) of phosphodiesters at ~1080 cm^{-1}, ν_s(CO-O-C) of carbohydrates at ~1050 cm^{-1} and ν_s(C-O) of carbohydrates at ~1030 cm^{-1}.

For clarity, principal component analysis was repeated for the three specific cell types in the dataset, namely, DSP, PSP and Non-SP. Fig. 5A illustrates the resulting PC score and plot for PCs 1, 2 and 3, and the loading plot of PC1 corresponding to the discrimination of the Non-SP from the SP spectra (Fig 5.5b).

The lipid signals, mentioned in the previous PC analysis, dominate the loadings plot of PC1. However, differences in the phosphodiester stretching absorptions also remain apparent in the loadings of PC2, along with carbohydrate vibrations (Fig. 5.5c).

Gault et al. looked into the effect of the Hoechst 33342 dye on cell FTIR spectra as a negative control when investigating alpha particle cellular irradiation. They found that Hoechst 33342 treatment of cells did not damage DNA.

They also found the binding of Hoechst 33342 did not induce intensity or band shift changes in the IR regions characteristic of proteins and lipids and there was only a subtle change in DNA conformation. (There was a 4 cm^{-1} shift in ν_{as}P=O (of PO_2^-) centred at 1238 cm^{-1} to 1242 cm^{-1} however the ν_sP=O (of PO_2^-) centred at 1086 cm^{-1} remained unchanged) .
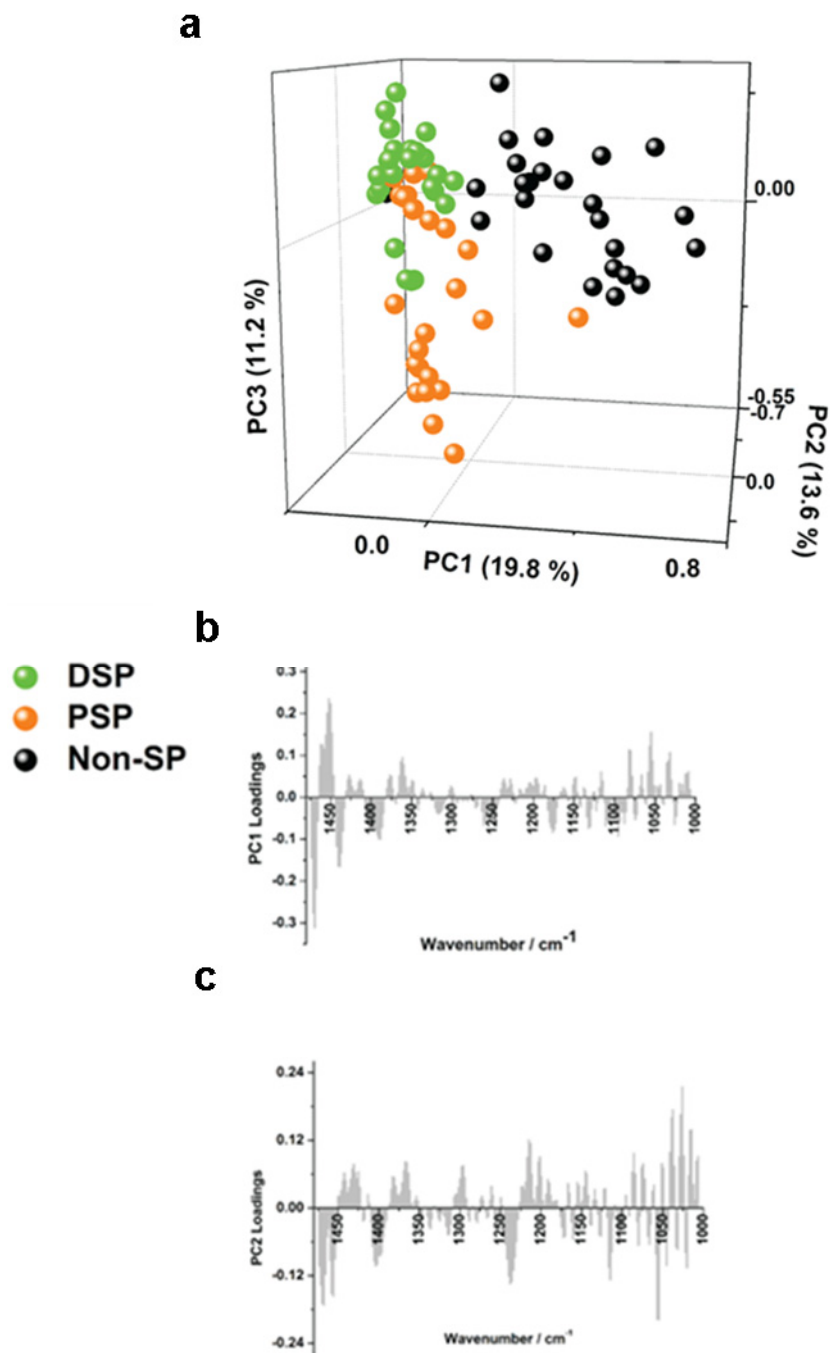
Fig. 5.5 │ PC analysis of the sub-side populations DSP and PSP versus non-SP cell spectra (a) the score plot of PC1, PC2 and PC3 (b) The corresponding loadings of PC1 and (c) PC2. Key biochemical differences are outlined in lipid, phosphodiester and carbohydrate absorption bands

5.3.4 │Qualitative spectral analysis

As a comparison, the mean second derivative spectra for Non-SP, PSP and DSP were qualitatively assessed (Fig. 5.6). There is however, always a danger when comparing averaged spectra for a cell type, particularly if one population is more heterogeneous, and the sample population is small. This approach was therefore more subjective than analysing PCA loadings which take account all individual cell spectra.
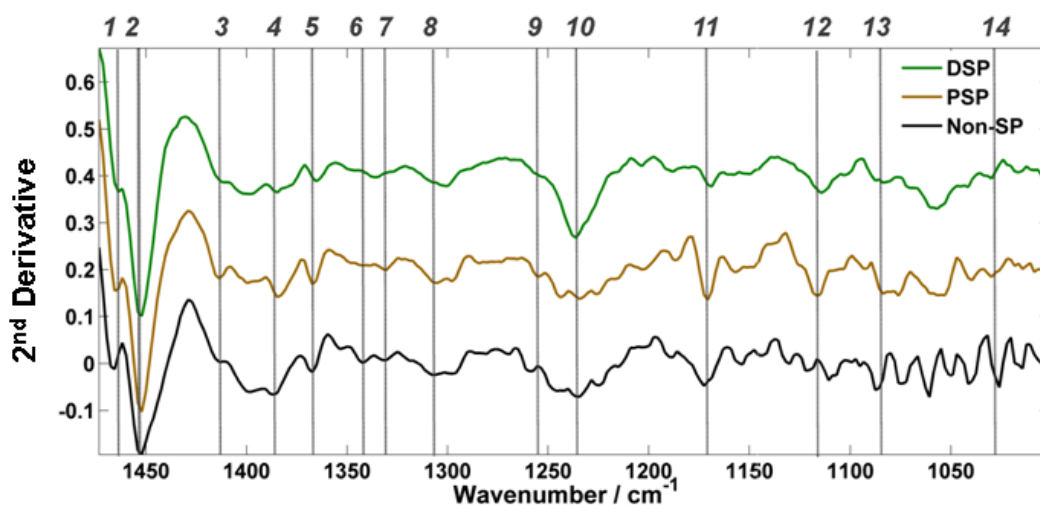
Fig. 5.6 │ Qualitative evaluation of the mean second derivative spectrum for DSP, PSP and Non-SP. Differences in spectral shifts are minimal whereas absorbance differences are most distinctive.

The assignments [11] of notable differences in the second derivative spectra were; *(1)* $\delta_{as}(CH_2)$ shift from 1464 to 1468 cm^{-1} between DSP to NSP and an increase peak height; *(2)* Increased peak height for DSP/PSP at 1452 cm^{-1} (C-H bending); *(3)* Increase in peak height for PSP at 1414 cm^{-1}; *(4)* $\delta_s(CH_3)$ shift from 1384 cm^{-1} to 1387 cm^{-1} between DSP to NSP; *(5)* DSP is shifted to 1365 cm^{-1} relative to PSP/NSP at 1367 cm^{-1}; *(6, 7)* Notable absorption bands at 1341 cm^{-1} and 1331 cm^{-1} for NSP; *(8)* Prominent band in PSP/NSP at 1308 cm^{-1}; *(9)* Carbohydrate marker at 1255 cm^{-1} for DSP/PSP, increasing in peak height and shift to 1258 cm^{-1} in NSP; *(10)* Increased peak height of $\nu_{as}P{=}O$ (of PO_2^-) at 1236 cm^{-1} in DSP relative to PSP/NSP at 1234 cm^{-1}; *(11)* Band at 1169 cm^{-1} shifting to 1173 from DSP to NSP, band most prominent in PSP; *(12)*Possible RNA band at 1115 cm^{-1} shifting to 1119 cm^{-1} from DSP to NSP; *(13)* $\nu_sP{=}O$ (of PO_2^-) AT 1085 cm^{-1} in DSP/PSP shifted to

1087 cm^{-1} in NSP; *(14)* Glycogen band at 1030 cm^{-1} for DSP/PSP shifted to 1026 cm^{-1} with an increase in peak height for DSP.

In the assessment into the effect of Hoechst 33342 on the cell spectra, Gault et al. observed a small shift of 4 cm^{-1} in the anti-symmetric v_{as}P=O phosphodiester band in mean spectra between stained and unstained cells. For the same band in the experimental renal data, there was no shift greater than 2 cm^{-1} between the all cell types.

The DSP has a larger v_{as}P=O peak height relative to the PSP/Non-SP. Gault et al. also observed that the v_sP=O at 1086 cm^{-1} remained unchanged [12].

In the renal data of this experiment, a 2 cm^{-1} shift and an increased absorbance for NSP was found relative to the SP subtypes, whereas Gault et al. found no striking changes in the lipid and protein signals between stained and unstained cells [12]. In the renal data analysis the differences in lipid signals studied in the wavenumber range appeared to significantly contribute to chemometric separation.

5.3.5 │ Linear discriminant analysis

To confirm the residing differences between the three cell types, principal component–linear discriminant analysis was performed (PC-LDA). LDA optimises inter-group separation whilst minimises intra-group separation [13]. Cross-validation was performed as is highly recommended when classifying a limited-size dataset. The model was validated by using a training data set where the algorithm is supervised with a priori knowledge of each data type, followed by the use of a test data set, introduced to the model as an independent validation.

As it is common practice to input into the model data of reduced dimensionality the principal components from PCA were invoked. Reduced data is used to overcome issues of over-fitting whereby good classification but poor significance can result. It is reported that over-fitting can be seen when using *'(n-g)/3'* components or more, where *g* is the number of groups and *n* is the total number of data points [14].

For the LDA classifier variables, the principal components were calculated in PCA for the DSP, PSP and Non-SP dataset (g = 3, *n = 75*). A total 8 principal components were used, a number securely far from the over-fitting threshold (24 PCs using the equation *'(n-g)/3'*). 8 PCs accounted for 71 % of the total variance. It is generally recommended not to use PCs that explain 100 % of variance as this will introduce PCs calculated on noise and will cause over-fitting of the LDA data. A training set comprising randomly selected data points from the total data set was constructed. 15 spectra per cell type were used for the training set (60 %).

Figure 5.7a displays the PC-LDA plot for the training set consisting of 45 spectra in total. Relative to the original PCA, the clustering and separation between cell types has greatly improved. Results show that the non side population cells are clearly differentiated from the side population cell types in discriminant function 1 (DF1). The proximal 'PSP' cells clustered relatively closer to the non-SP population than the distal DSP cells. This is accountable to the fact that the PSP cells share more chemical similarities to the Non-SP than the DSP cells share; the PSP cells are on the frontier of the SP cell gate in the FACS profile.

The DSP however are most different, separating from the non-SP in DF1 and even the PSP in DF2. It has been shown previously that the renal SP from human primary cells is a heterogeneous population [15] and that it is possible to sub-fractionate the populations based on the dye efflux status of the cells to generate sub-populations which are increasingly primitive. The results seem to confirm the existence of chemical difference within these sub-populations which can be used to further characterise or be utilised as a spectroscopic markers for these cells.
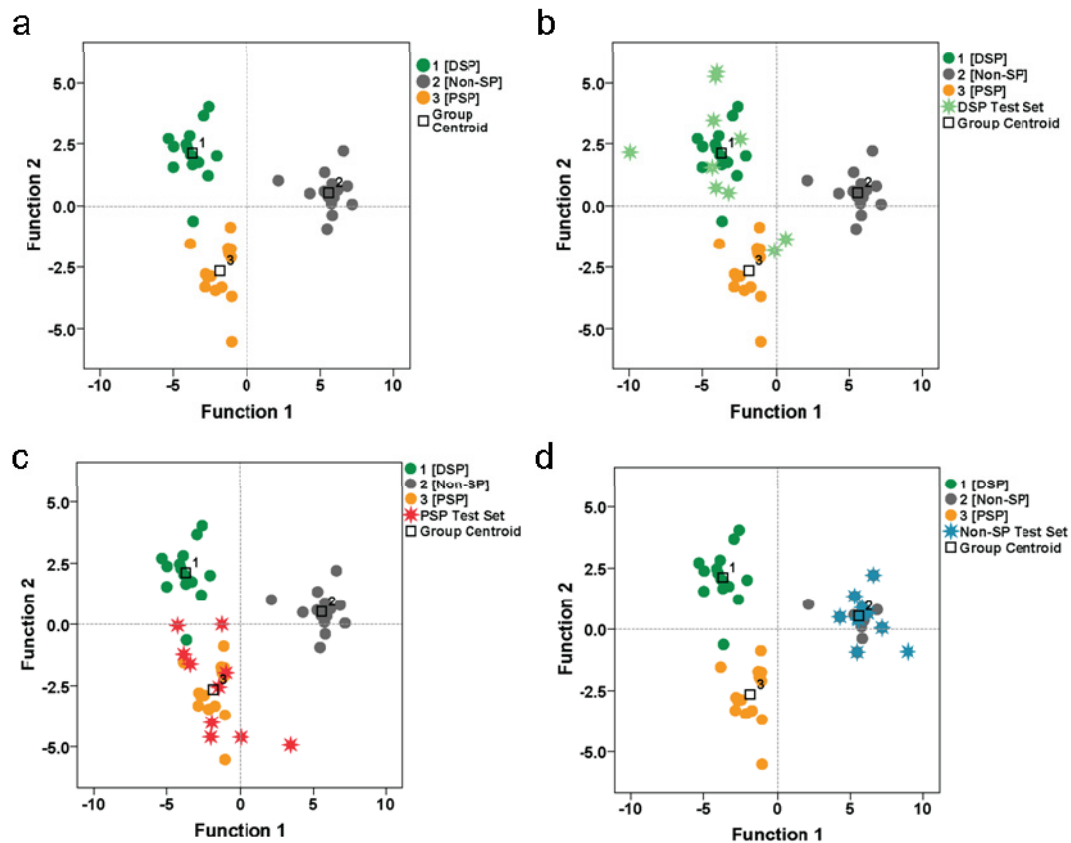
Fig. 5. 7 | PC-LDA of DSP and PSP versus Non-SP spectra (a) The training set, n=45 (b) DSP test set (c) Non-SP test set (d) PSP test set

In terms of model classification 93.3 % of training spectra correctly classified. Of the total dataset only a 3 spectra were misclassified. 6.7 % of DSP and 13.3 % of Non-SP were misclassified as PSP. In cross-validation using the 'leave-one-out' method, 84.4 % of spectra correctly classified. 13.3 % of DSP and 20 % of Non-SP were misclassified as PSP. 13.3 % of PSP were misclassified as DSP.

Figures 5.7b, c and d display the same PC-LDA plot but also include the independent validation DSP, PSP and Non-SP datasets respectively. 10 spectra per cell type were used to test the algorithm. The correct classification scores of the independent validation DSP, PSP and Non-SP dataset are 80 %, 90 % and 100 % respectively. In general, the resulting plots and relative scores of each independent validation set reveal that the spectra have been correctly assigned. The similarity in the phenotype of the sub-side populations is a probable explanation where a small percentage of data points have been misclassified.

5.3.6 │ Further assessment of the distal side population

PC analysis was performed using only the DSP versus the Non-SP to get a clearer picture of the distinction of the non-side population and stem-like cells relative to the most separable sub-side population. The DSP are the greatest distance away from the non-side population in the Hoechst FACS profile. The chemistry of the DSP should be the most distinctive due to the more primitive nature of the cells.

The PCA score plot (Fig. 5.8a) of PC1 versus PC2 accounts for 40 % of total variance. The distal side population clearly tightly clusters separately from the non-side population. The clustering of the DSP suggests that the cells within the DSP are more chemically and spectrally homogeneous than the non-SP which appears to form two clusters, separated in PC2.

The non-SP variability may be due to actively proliferating cells at various stages of differentiation within the overall non-SP population. In contrast the DSP is much more likely to be quiescent and more clonal in nature leading to a more uniform spectral profile, as it is hypothesised to contain the most primitive stem like cells. The noteworthy differences between the distal and non-side populations are shown in the PC1 loading plot (Fig. 5.8b).

There are significant phosphodiester and carbohydrate signals but these are secondary to the biochemical lipids signatures that dominate the loading plot at $\delta_{as}(CH_2)$ ~1468 cm^{-1} and $\delta_s(CH_3)$ at ~1380 cm^{-1}.
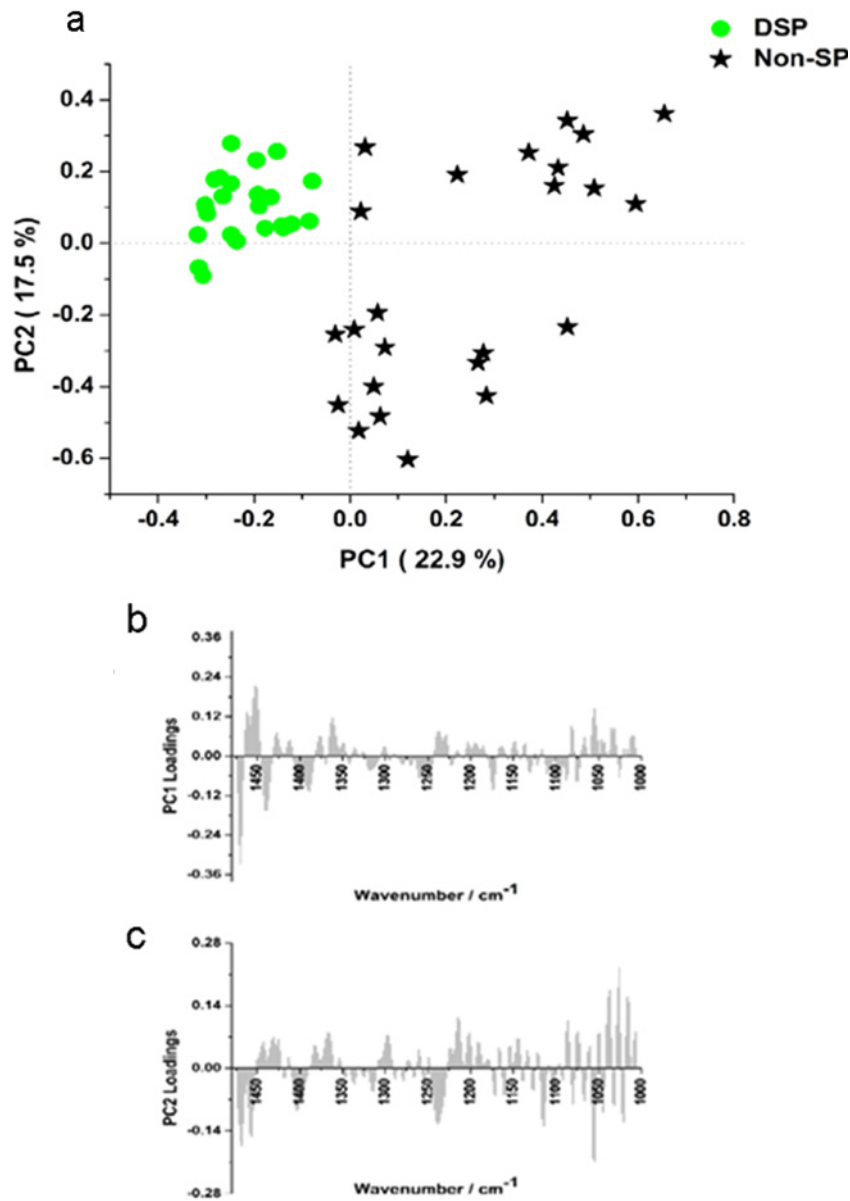
Fig. 5.8 │DSP versus Non-SPs (a) The PCA plot of PC1 and PC2. PC1 separates the DSP and NSP populations, whereas PC2 shows clustering in the NSP, indicative of heterogeneity in the population (b) the corresponding loadings for PC1 and PC2.

Carbohydrate bands are highly loaded in PC2 where the Non-SP separate into two clusters. Coupled with the symmetric and anti-symmetric phosphodiester stretching bands this suggests differences in metabolic activity, possibly due to cell cycle effects. The glycogen: phosphate ratio at 1030/1080 cm^{-1}, a known marker of

metabolic activity [16]. This is thought to correspond to increased metabolic activity, as glycogen stores are broken down at a greater rate. The PC2 loading plot displays the re-occurring symmetric and anti-symmetric phosphodiester band features also seen in Fig. 4b and Fig 5c.

For a final evaluation, the differences in cell chemistry between the DSP and PSP cells were analysed (Fig 5.9). Clear separation can be seen between the cell types in PC1 shown in the score plot. In the comparison of DSP and non-SP spectra, the main differences were highlighted by lipid chemistry. Also highlighted are the phosphodiester stretching bands as a significant discriminatory feature.
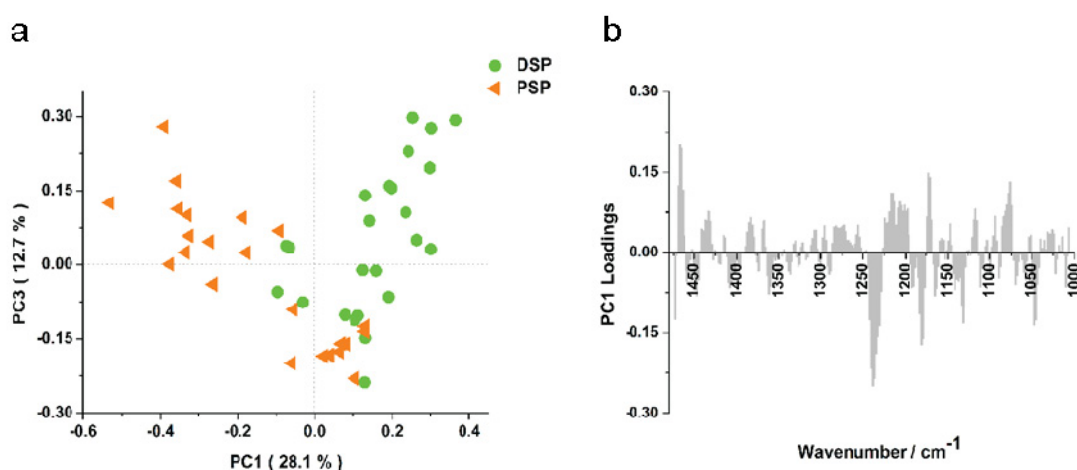
Fig. 5.9 │ The PCA score plot of distal (DSP) versus proximal (PSP) second derivative spectra. Sub-side population discrimination is shown in the first principal component (28.1%). The loading plot highlights the ν_{as}P=O stretch a distinct marker, which has also been present in the previous loadings comparing NSP to the SPs.

A possible explanation as to why the ν_{as}P=O (of PO_2^-) is most apparent in this comparison is that the DSP and PSP cells have greater similarity in terms of their proliferative and morphological phenotype. When possible differences due to cell cycle or overall cellular heterogeneity are removed, it is apparent that the phosphodiester stretching bands may be a prominent marker in the distinction of these stem-like cells. Differences in the symmetric and asymmetric phosphodiester

stretching bands continually re-occur throughout the analysis and are always anti-correlated in the loadings plots.

This may be indicative of reported DNA conformational changes where the symmetric phosphodiester modifications ~1080 cm^{-1} were detected as a significant marker of putative stem cells in human intestinal crypts [17]. The symmetric and anti-symmetric phosphodiester stretching vibrations refer to the central phosphorous atom in a phosphodiester linkage (--O--PO$_2$---O--), seen in DNA, RNA and some phospholipids. In DNA, the negatively charged phosphorous atom is countered by Na^{+} ions [18].

A recent study [19] of other stem-like cell line populations supports this discriminatory band along with a carbohydrate marker at 1054 cm^{-1}. However, rather than looking for a univocal specific maker of stem-like cell discrimination, the authors propose using a change in a global absorption pattern as a better approach, particularly monitoring the carbohydrate rich vibrations at 1000-1200 cm^{-1}, which they report to be a particular stem-cell like discriminator range. Yet spectral changes in carbohydrate bands may be more indicative of cell cycle effects. This would contribute greatly to chemometric discrimination and therefore it is important to bear this in mind when interpreting cell characterisation. The later research characterises the stem-like cells during a live cell suspension.

Throughout the time of data acquisition there are a range of variables that may affect the absorption pattern of a single cell. For instance, cell viability may be reduced and dying cells have previously been reported to have different spectral features to viable cells [9]. The spectra may be recording the independent cellular response to the non-ideal conditions of their environment for the respective cell types. Also considerations would have to be taken into account for the cell suspension fluid and the contributions to the spectra. However, if these factors can be monitored and perhaps controlled, a live cell approach would be a very useful application when monitoring side population cells after subsequent passage, or their chemical response to anti-cancerous agents for instance.

PCA and LDA were used as part of a data-reduction analysis to elucidate chemical difference between the SP and Non-SP cell types. All chemometric analysis was performed on the second derivative corrected spectra with a range of 1475-1000 cm^{-1} to achieve maximum separation between the Non-SP and the SP cells. The full fingerprint region including the amide I and II bands reduced the clarity of separation (Fig. 5.10). For example, use of the full fingerprint range 1800-1000 cm^{-1} produced a poorer discrimination in PCA in scores for DSP/PSP versus NSP (Fig. 5.4). The percentages explained by PC1, PC2 and PC3 were 19.8, 13.6 and 11.2% respectively for range (44.6% of total variance explained in 3 PCs). If the full fingerprint range was used, the values decreased. (15.1%, 10.9% and 9.9% respectively, 35.9% in total).

Figure 5.10a is the score plot for PC1 versus PC2 for the full fingerprint range where PC1 displays no separation of the different subtypes. In the loadings plot the cause of the spread of scores is predominantly due to the amide region (Fig. 5.10b). PC2 (Fig. 5.10c, d) displays some separation and the PC2 loadings show distinct phosphate and lipid signals are present, along with the amide I and II which still appear to worsen the separation (relative to omitting the amide region). In PC3 (Fig. 5.10e) again there is no separation between the SP and the Non-SP sub types, however the Non-SP clusters into two groups. In the PC3 loadings it is the amide I and II regions are again largely contributing, along with the carbohydrate absorptions, as noted in the comparison of DSP and Non-SP (Fig. 5.8).
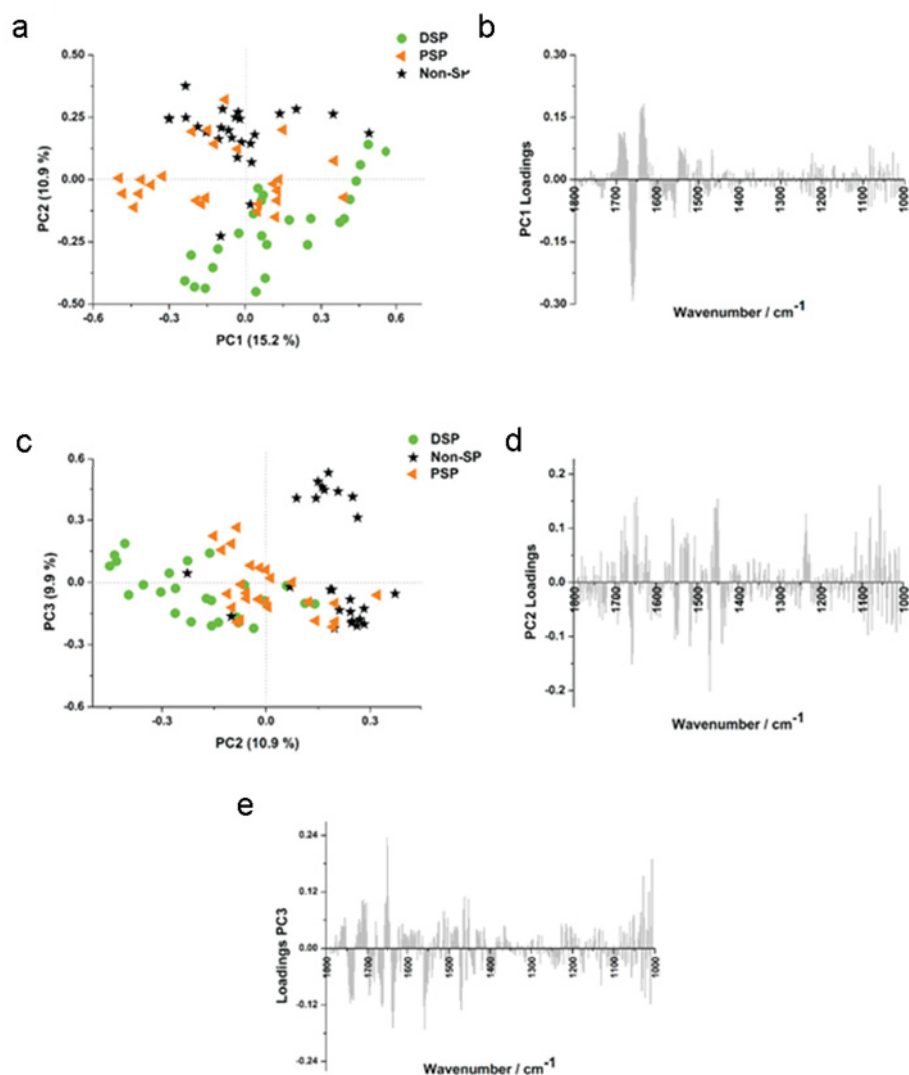
Fig. 5.10 │The PCA analysis of extended spectral range (1800-1000 cm^{-1}) resulted in poorer cell type separation of distal (DSP) versus proximal (PSP) and Non-SP cells using second derivative spectra; (a) score plot of PC1/PC2 and (c) PC2/PC3 and corresponding loading plots of (b) PC1 (d) PC2 and (e) PC3. Prevalent phosphodiester and lipid peaks evaluated previously are apparent in PC2 whereas PC1 and PC3 denote intra-cellular heterogeneity

5.4│Conclusions and future prospects

It has been successfully demonstrated that it is possible to isolate a side population from a human renal cancer cell line as well as from solid tissue. However these 'side population' cells only account for a small proportion of the total cell population (typically <5%), creating difficulties for analysis, which increases with

the further sub-fractionation required for increasing primitive cell isolation. It is therefore difficult to characterise cell populations of this type by conventional biological methods.

High brilliance synchrotron FTIR- spectroscopy made it possible to record the spectral fingerprint of these different cell types after FACS. PC and PC-LDA analysis confirmed the biochemical differences of these cell types, providing further support that the side population contains a different subpopulation of cells and that this side population can be further sub-fractionated to give distinct populations based on their cellular biochemistry.

Interpreting the cause of discrimination of the cell types revealed variances within lipid and phosphodiester signals, whereas carbohydrate and protein signals in the loadings plots suggested heterogeneity within the non-SP cells.

It is perhaps the phosphodiester signals that were most revealing, as they may represent conformational changes in the DNA/RNA, as well as some phospholipids. These are indicated in the loadings plots in Figures 4b, 5c, 8c and most prominently in 9b. The anti-symmetric phosphodiester vibrations at ~1237 cm^{-1} are partially overlapped by complex amide III vibrations, whilst the symmetric phosphodiester vibration at ~1080 cm^{-1} can be masked by carbohydrate vibrations. However, with the use of the second derivative combined with chemometric data reduction techniques, the phosphodiester bands are distinguishable and seem to be a significant cause of variance in the biochemical fingerprints of these side population cells relative to the general cell population.

Another key factor of cellular distinction in the loadings (Fig. 5.4b 5.5b, 5.8b, 5.9b, 5.10d) and in the qualitative mean spectral analysis (Fig. 5.6) were shifts and a higher absorbance for Non-SP in $\delta_s(CH_3)$ and $\delta_{as}(CH_2)$ and a higher absorbance in for DSP/PSP at 1452 cm^{-1} (C-H bending).

At this point, the impact of the differential chemistry results surrounding the SP cells can only be speculated. The concluding analysis suggests possible DNA and

lipid differences. There may be a link between these features and it is known that DNA and phospholipids share a common motif recognizable by proteins [20]. Also, it is generally accepted that lipids play both functional and structural roles in the nucleus as well as in the plasma membrane and that lipid metabolism in the nucleus may be regulated separately from that of the plasma membrane [21]. Perhaps this is a possible route of further investigation into DSP evaluation.

In conclusion this novel approach offers considerable potential as a tool to distinguish small cell numbers associated with potential cancer stem cells and the SP phenotype and may be of particular use in facilitating the understanding of cancer stem cell biology.

Finally, it is possible after FACS to obtain live cells which have shown limited cell culture of the side population phenotype. Further studies could involve time-course experiments, monitoring the rate of differentiation. These cells are also under investigation with chemotherapeutic agents and perhaps FTIR spectroscopy could shed some light on the responses of these cells to such toxins.

5.5 │References

1. Bhatt RI, Brown MD, Hart CA, Gilmore P, Ramani VAC, George NJ, Clarke NW: **Novel Method for the Isolation and Characterisation of the Putative Prostatic Stem Cell**. *Cytometry* 2003, *54*(2):89-99.

2. Brown MD, Gilmore PE, Hart CA, Samuel JD, Ramani VAC, George NJ, Clarke NW: **Characterization of benign and malignant prostate epithelial Hoechst 33342 side populations**. *Prostate* 2007, **67**(13):1384-1396.

3. Addla SK, Brown MD, Hart CA, Ramani VAC, Clarke NW: **Characterization of the Hoechst 33342 side population from normal and malignant human renal epithelial cells.** *American Journal of Physiology - Renal Physiology* 2008, **295**(3).

4. Brown M, Grey B, Oates J, Clarke NW.: **Cancer Stem Cells:** *Transworld Research Network*; 2010, p.107-129.

5. Dumas P, Polack F, Lagarde B, Chubar O, Giorgetta JL, Lefrancois S: **Synchrotron infrared microscopy at the French Synchrotron Facility Soleil**. *Infrared Physics and Technology* 2006, **49**(1-2):152-160.

6. 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**(8):1586-1593.

7. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, Gazi E, Brown M, Clarke N, Gardner P: **Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples**. *Analyst* 2010, **135**(2):268-277.

8. Goormaghtigh E, Derenne A, Benard A, Gasper R, Raussens V: **Data processing in FTIR imaging of cells: Towards protein secondary structure imaging**. *Spectroscopy* 2010, **24**(1-2):51-54.

9. Holman HYN, Martin MC, Blakely EA, Bjornstad K, McKinney WR: **IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based Fourier transform IR spectromicroscopy**. *Biopolymers-Biospectroscopy Section* 2000, **57**(6):329-335

10. Swain RJ, Stevens MM: **Raman microspectroscopy for non-invasive biochemical analysis of single cells**. *Biochemical Society Transactions* 2007, **35**(3):544-549.

11. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

12. Gault N, Rigaud, O., Poncy, J-L., Lefaix, J-L.: **Biochemical Alterations in Human Cells Irradiated with a Particles Delivered by Macro- or Microbeams**. *Radiation Research* 2007, **167**:551–562.

13. Fearn T: **Handbook of Vibrational Spectroscopy**. *John Wiley & Sons* 2002, 3:2086–2093.

14. Defernez M, Kemsley, E.K.: **The use and misuse of chemometrics for treating classification problems**. *Trends in Analytical Chemistry* 1997, **16**:216-221.

15. Addla SK, Brown MD, Hart CA, Ramani VAC, Clarke NW: **Characterization of the Hoechst 33342 side population from normal and malignant human renal epithelial cells.** *American Journal of Physiology - Renal Physiology* 2008, **295**(3).

16. Gazi E, Dwyer J, Lockyer NP, Gardner P, Shanks JH, Roulson J, Hart CA, Clarke NW, Brown MD: **Biomolecular profiling of metastatic prostate cancer cells in bone marrow tissue using FTIR microspectroscopy: A pilot study**. *Analytical and Bioanalytical Chemistry* 2007, **387**(5):1621-1631.

17. Walsh MJ, Fellous TG, Hammiche A, Lin WR, Fullwood NJ, Grude O, Bahrami F, Nicholson JM, Cotte M, Susini J et al.**: Fourier transform infrared microspectroscopy identifies symmetric PO_2^- modifications as a marker of the putative stem cell region of human intestinal crypts**. *Stem Cells* 2008, **26**(1):108-118.

18. Salzer R, Siesler, H.W.: **Infrared and Raman Spectroscopic Imaging**: *Wiley-VCH*; 2009: p183

19. Zhao R, Quaroni L, Casson AG: **Fourier transform infrared (FTIR) spectromicroscopic characterization of stem-like cell populations in**

human esophageal normal and adenocarcinoma cell lines. *Analyst* 2010, **135**(1):53-61.

20. Sekimizu K: **Interactions between DNA replication-related proteins and phospholipid vesicles in vitro**. *Chemistry and Physics of Lipids* 1994, **73**(1-2):223-230.

21. Tamiya-Koizumi K: **Nuclear lipid metabolism and signalling**. *Journal of Biochemistry* 2002, **132**(1):13-22.

Chapter 6 │ Investigating Cellular Response to Chemotherapeutics

6.1 │ Introduction and Aims

In the previous chapters, spectral analysis focused on cell characterisation. Discriminating the FTIR signatures of different cell lines can be relatively straightforward. Comparing differences in cells from the same population, however, can be more challenging, as described in Chapter 5.

Cells are said to have specific infrared signatures in terms of cell cycle, viability, apoptosis and stress [1-7]. If a chemical spectroscopic signature, unique to a particular mode of drug action, was discovered it would be very beneficial in the early stages of the drug development process as novel chemotherapeutic agents with potential regularly fail during their progression to clinical trial testing. Currently, high throughput procedures used in the development of a novel compounds include cytotoxicity multi-cell line panel testing and the assessment of agent interaction and involvement in parts of numerous biological pathways [8].

Implementing FTIR spectroscopy as a tool in this area would only be appropriate for the assessment of global changes in cellular chemistry. Attempting to understand the interaction drug activity upon cellular processes is a complex area. Despite this fact, there are a number of key questions that may be answered:-

- Is it possible that cellular responses to drugs can be detected in an infrared spectral fingerprint and is there a biomarker to monitor this?
- Is it possible to discriminate between cells that have been treated with different drugs that act through different mechanisms / or have different modes of action?
- Is it possible to detect cellular response to drugs in the infrared that is not apparent early on by conventional testing?
- Is it possible to deduce common biomarkers to show cellular response that may help to predict drug efficacy?

The cell line model comprised of an established renal carcinoma cell line, CAKI-2, dosed with a drug of known mode of action as well as a set of novel gold-based

compounds with unknown efficacy. The novel compounds studies were gold analogues, KF0101, KF0113 and KF0501, (which will be referred to as KF1, KF13 and KF5 respectively).

As discussed in Chapter 1, the study of multidrug resistance (MDR) in tumour cell lines has led to the discovery of the plasma membrane P-glycoprotein (P-gp) molecule. The protein functions as an efflux pump to remove foreign compounds such as drugs. P-gp-mediated MDR tumour cells have been found in human cancers and are highly likely be a determining factor in negative response to treatments in patients [9].

Anti-tumour agents from the platinum family such as cisplatin and carboplatin have been widely used in clinical therapy. They have DNA cross-linking activity, inhibiting further replication or transcription to inhibit tumour proliferation and have received special interest due to their different kinetics and geometries. Synergistic treatments can enhance apoptosis, such as combinations of Cisplatin with Lexatumumab, a human agonistic antibody that targets a tumour necrosis factor [10].

Cisplatin treatment alone, however, performs poorly in RCC due to resistance. Altered DNA process and changes involving glutathione and metallothioneins are some examples where resistance has developed. This has prompted a recent surge in new non platinum drugs and old species have received much attention due to their strong tumour cell growth inhibiting effects [11]. The mechanism of action of anti-proliferative gold complexes is still speculative and due to a vast library of gold species with different oxidation states and ligands it is unlikely for there to be a unique mode of action to account for them all.

The novel gold compounds of the KF series comprise of two covalently bound Au(I) atoms within an aromatic ring with triphenylphosphine ligands. They exhibit a high potency in cell toxicity studies and DNA cross-linking assays that indicate these compounds should show efficacy in the treatment of carcinomas [12]. It is supposed that the highly stable gold(I)-carbon covalent bonds, due to the similar

electronegativity of Au(I) and carbon, facilitate the DNA cross linking. This increased facilitation may improve activity of these novel compounds in cell lines that are perhaps platinum resistant [12].

The use of FTIR as a tool for analysing drug response in cells is relatively new, although there are already a number of studies that have investigated this which have shown promise [13-15]. In the following experiment, complimentary techniques such as proliferation and cytotoxicity assays were employed in parallel to infrared spectroscopy to corroborate any possible findings.

6.2 │ Experimental

6.2.1 │ Determination of inhibitory concentrations

To deduce inhibitory concentrations (IC) for the drugs against the Caki-2 cell line, the widely used sulforhodamine B (SRB) cytotoxicity assay was implemented according to the Nature protocol [16] described in Ch. 24.6. The data was exported to Graphpad Prism where it was normalised and plotted in the classic sinusoidal logarithmic form to calculate the IC_{50} values for each agent (Fig. 6.1).

6.2.2 │ Proliferation assay

Caki-2 cells were added to well plates at a density of $1x10^5$ cells/well and left to establish. 72 wells were setup to allow for triplicate measurements over a 6 day period. After 24 hours, agents KF1, KF13 KF5 and 5FU were introduced at the IC_{50} level determined by a 3-Day SRB assay (3.51, 3.52, 5.32 and 2.39 μM respectively). For each subsequent day (1–6 days after drug introduction) viable cells were identified using Trypan blue dye and then counted using a haemocytometer.

6.2.3 │ Spectral sample preparation

The samples were prepared as described in the FTIR single-point spectroscopy protocol for sample preparation (Ch. 2.4.2.1). In brief, Caki-2 cells were added to well plates at a density of $1x10^5$ cells/well and left to establish. After 24 hours agents KF0101, KF0113 and 5FU were introduced at the IC_{50} level, as determined by the SRB assay. At the appropriate time point the cell cultured slides were washed twice in PBS and fixed in formalin (4%) for 40 minutes. The slides were then dripped in distilled water to remove salt residues.

6.2.4 │ Data acquisition

6.2.4.1 │ Single cell measurements

SR-FTIR transflection spectra were acquired on the B22 beamline of the British Synchrotron Facility Diamond (Oxford) using a Bruker80 V spectrometer equipped with an MCT detector and coupled with a Hyperion 3000 microscope. Spectra were recorded at 4 cm^{-1} resolution with 256 co-scans and the size of the aperture was adjusted to match the diameter of the cell such that it was fully illuminated, typically $15\mu m^2$.

6.2.4.2 │ Monolayer measurements

Imaging measurements were taken using a Varian 670-IR FTIR spectrometer as described in Ch. 2.4.5.3. Hyperspectral images of 700 x700 μm were recorded using a spectral resolution of 4 cm^{-1} resolution with 128 co-scans.

Single point measurements were acquired as described in Ch. 2.4.3.1 with 128 co-scans using an aperture of 60 x 60 μm.

6.2.5 │ Spectral pre-processing

All datasets were subject to the RMieS-EMSC correction algorithm at 150 iterations for the SR-FTIR and single-point monolayer data and 80 iterations for the hyperspectral images. A Matrigel transflection spectrum was used as the reference.

The corrected spectra were transformed to the second derivative with 7 point Savitsky-Golay smoothing with a polynomial order of 3 in Matlab. Analysis was performed using the fingerprint range of 900-1825 cm^{-1}. The corrected spectra were vector normalised and mean-centred before principal component analysis (PCA) and discriminant function analysis (PC-LDA) also in Matlab.

6.3│Results and Discussion

6.3.1 │Cell cytotoxicity

Gold (I) complexes with labile ligands such as chlorine interact readily with DNA. Cisplatin for example, a drug with a cytostatic effect, is an effective TrxR inhibitor. However, complexes with tightly binding ligands such as Auranofin, will not bind to DNA [11].

A gold compound with multiple phosphine ligands is [Au(dppe)2]+, a tetrahedral bi-chelated Au complex, has shown antitumor activity in vitro and in vivo and induced DNA protein cross linking and DNA strand breaks in cells[11] .

A similar complex, [Au(dppp)(PPh$_3$)-Cl] was tested at the N.C.I., U.S.A. 29 cell lines of a 60 cell line panel and shown to have IC$_{50}$ values in the lower μM range where the cells displayed an apoptotic response with decreased expression of anti-apoptotic proteins . Gold triphenylphoshine complexes have also shown to have IC$_{50}$ values in the lower μM range [17].

The KF series for the drugs tested displayed similar IC$_{50}$ concentrations to the above compounds, with KF1, KF13 and KF5 values of [3.51 μM], [3.52 μM] and [5.32 μM] respectively. The KF5 curve, however, did not fit as accurately as the other analogues therefore the value obtained was not taken with confidence and more repetitions would be needed. In contrast the IC$_{50}$ value for 5FU was [2.39 μM], almost one and a half times less-concentrated than the KF series (Fig. 6.1).

5FU is typically dosed at 12 mg/(kg day) to cancer patients, giving a plasma concentration of [0.75 mM] [18]. It is difficult, however, to relate IC_{50} values obtained in vitro to human in vivo pharmacology. There are many factors that may explain why such a concentration is used in the human body including bioavailability and pharmacokinetics. The disposition of a pharmaceutical compound within an organism can be described by four principal criteria: absorption, distribution, metabolism and excretion of a drug (ADME) [19]. They influence the drug levels and kinetics of drug exposure to the human body and hence influence the performance and pharmacological activity of the compound as a drug. For example in drug discovery clinical trials, failures of drug progression are often down to clinical ADME. Adverse events or poor pharmacokinetics may be observed that are not predicted by animal studies. Drugs may also fail to demonstrate the efficacy expected from animal studies or results from early in vitro studies [20].
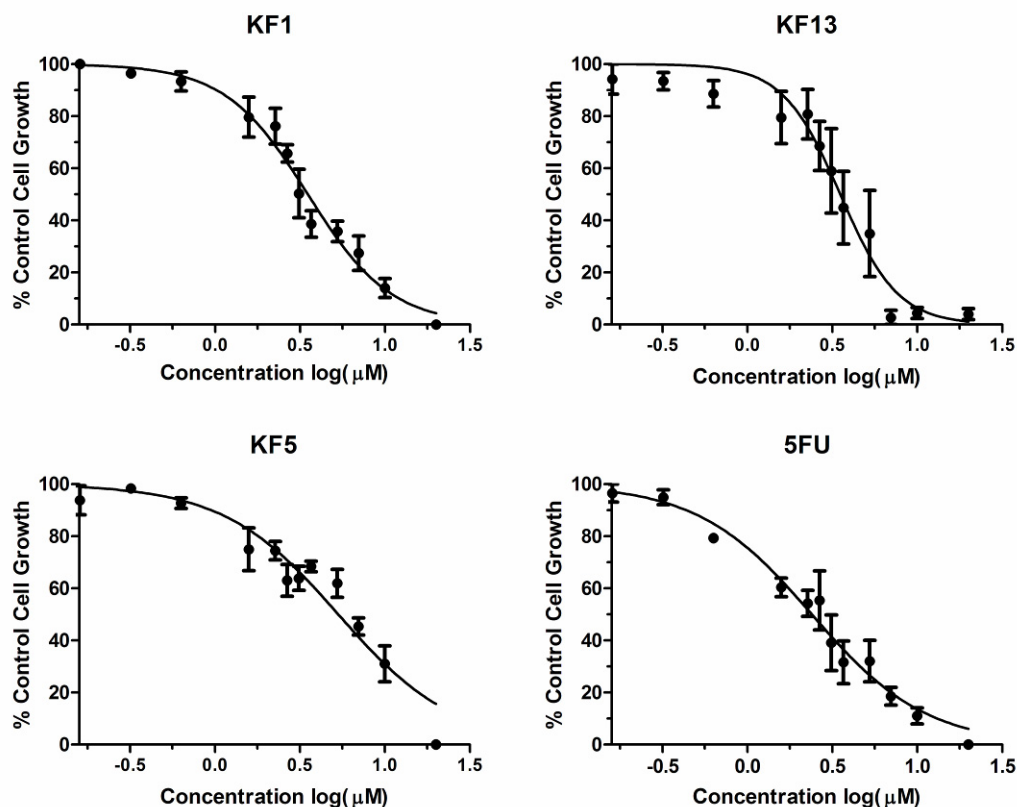
Figure 6.1 │ Dose response curves to determine IC$_{50}$ values for the Caki-2 cell line from the SRB assay data (repeated in triplicate). Graphs were created in Graphpad Prism using mean values ± standard error.

6.3.2 │ Single cell SR-FTIR analysis after 24 hours exposure to compounds

6.3.2.1 │ Principal component analysis (PCA)

One of the aims of the study was to determine if there was an obvious, general global infrared maker distinguishing cell response to chemotherapeutic agents. Spectra of drug-treated cells after 24 hours of drug exposure were assessed against the (non-drug treated) control group (Fig. 6.2). There was little difference to be observed in the mean spectra for each cell type in general (Fig. 6.2a), although subtle differences in absorbance of peaks associated with $\delta_{as}(CH)_2$ at ~1480 cm^{-1}, v_{as} (PO)$_2$ at ~1240 cm^{-1} and v_s (PO$_2$) at 1080 cm^{-1} were noted for KF1 and KF13-treated cells.
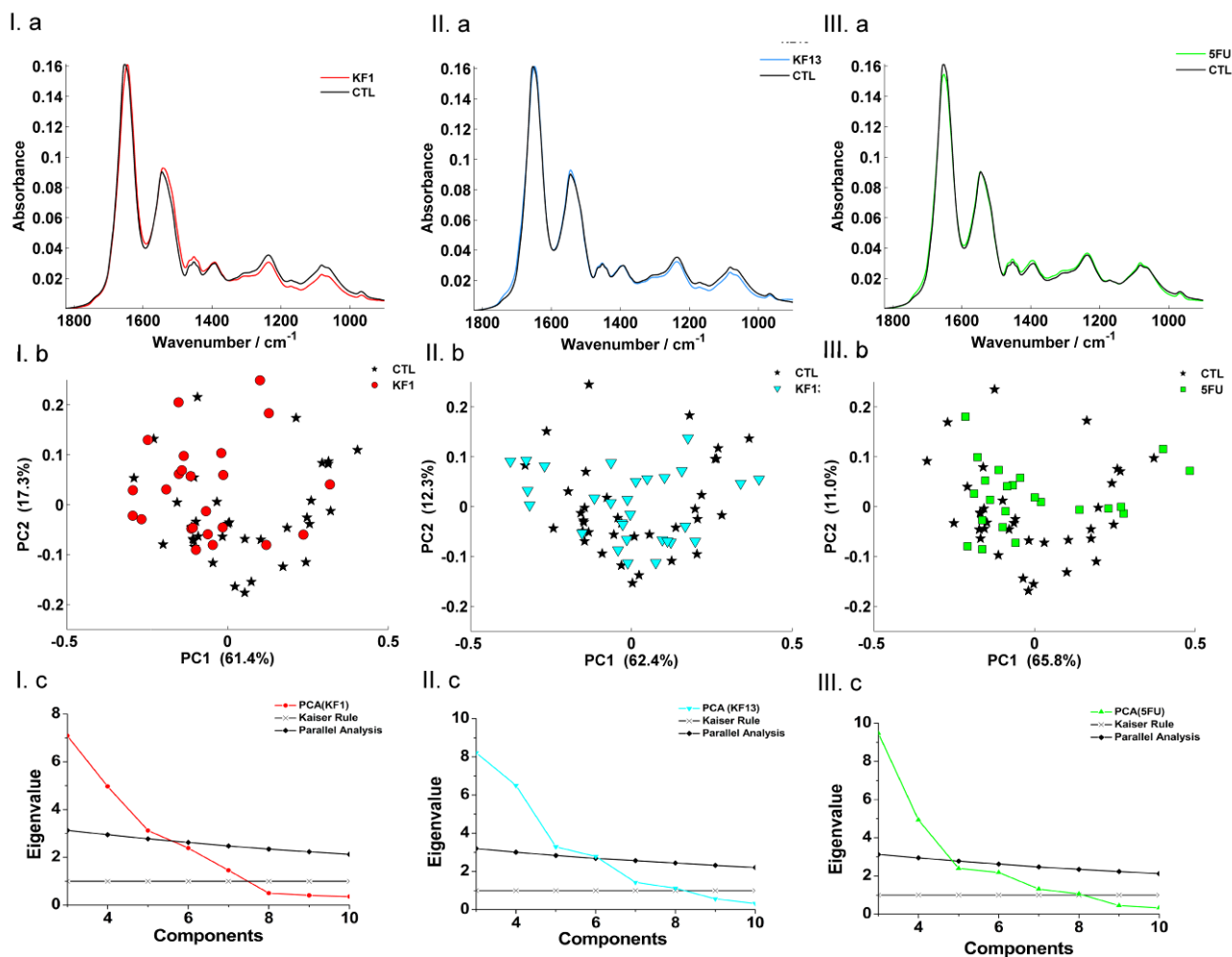
Figure 6.2 │ **(a) Mean spectra for drug treated cells versus control, (b) PCA score plots and (c) Eigenvalue scree plots for determining the number of PCs to use in LDA for (I) KF1, (II) KF13 and (III) 5FU respectively**

It was difficult to make any distinction between control cells and drug treated cells using principal component analysis (PCA) alone (Fig. 6.2b). This may due to the intra-sample heterogeneity observed in the cell populations. Cells exhibit specific chemistry in a certain phase of the cell cycle and therefore it may difficult to see any chemistry purely associated with drug response. Another complicating factor may be the influence of the agents on the cell cycle. It is known that 5FU does affect cell cycle in the S phase; however there is no information on whether the KF series will also have an influence.

It is also this reason that analysing these variables by using mean spectra alone may be misleading. The mean spectra represent a small portion of a complete population

of cells studied. There may be bias as more spectra from cells in one phase in the cell cycle may have been sampled in one group more than group due to chance and may not be completely representative of the true cell population.

6.3.2.2 │ *Principal component linear discriminant analysis (PC-LDA)*

In the previous chapter, PC-LDA was used to classify the cellular sub-types in a population. SPSS software was used, enabling results to be cross-validated in terms of categorising cell spectra into distinct groups. A slightly different approach was used to assess cellular responses to chemotherapeutic agents. The use of IC_{50} values, by their definition, will equate to half the population of cells being affected at that specific concentration. The implication for single cell analysis will mean that the agent may elicit a cellular response in a given period of time for some cells, but not others. It was therefore unpractical to assume that the cell spectra would fall into rigid clusters and it was not the aim to categorise the spectra in such a way. Subsequent model testing by the input of 'blind' spectra as a means of cross-validate to the robustness of such distinct classified cell spectra did not seem relevant to the samples in question.

PC-LDA was useful, however, as a measure to maximise overall differences in the cells treated with different agents in an attempt to detect general trends. As previously described (Ch 5.3), there is no strict formula to know how many principal components should be used to for PC-LDA. As sample numbers were small, further methods were used in an attempt to assess the number of principal components to use in LDA (Fig. 6.2c), rather than simply following the (n-g)/3 over-fitting rule [21]. The Kaiser rule recommends discarding principal components after the Eignenvlaues are no longer equal to or above 1.0 [22]. Another approach is to use the number of PCs that explain 95% of total variance. A fourth method commonly used is Cattell's scree plot analysis, whereby the number of components taken should equal the point before the plot reaches the 'elbow' [23]. Finally, Horn's parallel analysis was tested. M. Watkins suggested Monte Carlo PCA for parallel analysis whereby PCA is performed to obtain Eigenvalues for a correlation matrix constructed by random numbers, equal in size to the experimental data [24].

The scree plots for simulated and experimental data are compared and the threshold is the PC before the experimental plot crosses over the simulated plot [25]. The results can be seen in Figure 6.2c where it is apparent that the lowest number of acceptable components is four.

The LDA score plots shown in figure 6.2c resulted from four principal components that explained 91%, 91% and 89% and of variance in 5FU, KF1 and KF13 respectively (Fig 6.3a).
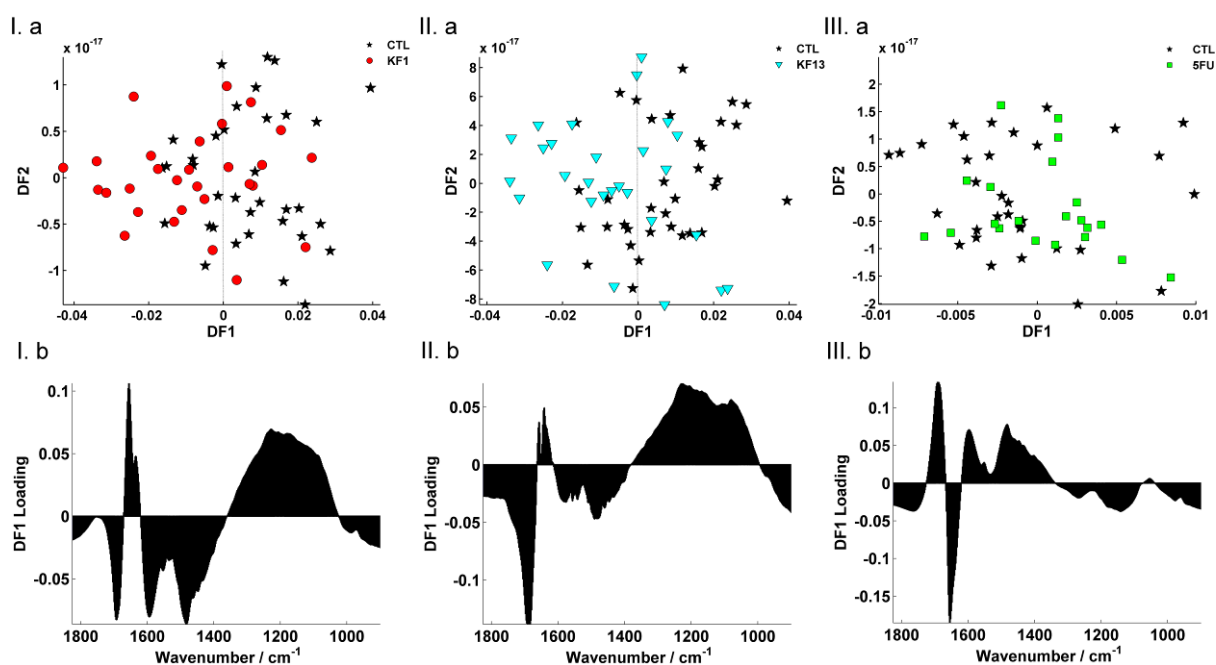
Figure 6.3 | (a) LDA score plot using 4 PCs for KF1 (I), KF13 (II) and 5FU (III). (b) LDA loadings are displayed for DF1 respectively.

As expected, there was still an overlap of drug-treated and control spectra. Drug treated cell spectra were primarily distributed in negative DF1 space and control spectra in positive DF1 space for KF1 (Fig. 6.3Ia) and KF13 (Fig. 6.3IIIa). No distinction could be seen for 5FU treated cells at this point. It could be argued that the cells are showing a greater response to the gold analogues than 5FU at this 24 hour check point, as separation in DF1 is marginally better. The weighting in the KF1 and KF13 loading plots of the 1300-900 cm^{-1} region, associated with

carbohydrates and nucleic acids, appears greater than in the 5FU loading plot, where the region weighting is considerably less (Fig. 6.3b).

6.3.2.3 │ *Use of second derivative spectra*

To determine if discrimination between the treated and non-treated cell spectra could be enhanced PCA and subsequent LDA was repeated for the second derivative-equivalent of the dataset. Generally it was noted that more PCs were required to explain the same amount of variance in the 2nd derivative dataset (Fig 6.4).
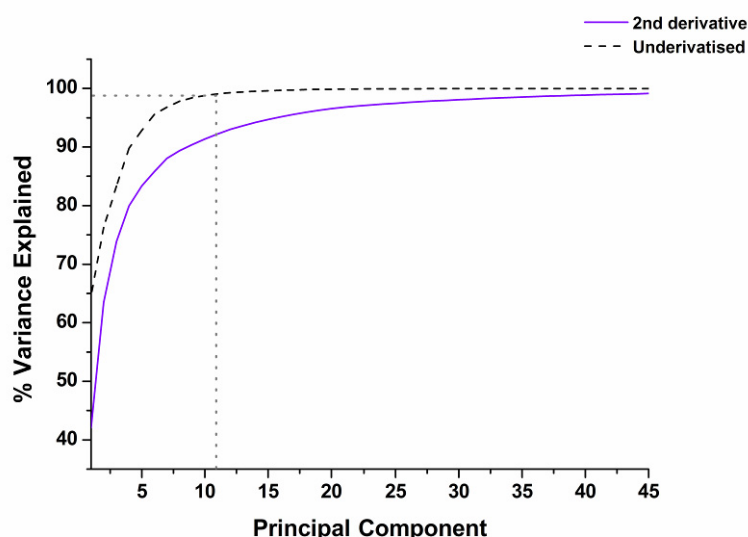
Figure 6.4│Cumulative variance explained by principal components for the un-derivatised and equivalent 2nd derivative dataset

A possible suggestion is that more PCs may be required to explain the equivalent variance. Derivatives assess the rate of change in absorbance with respect to wavenumber and consequently the strong amide I and II signals dominate derivative spectra and subsequent loading plots. Information from other regions of the spectrum with small derivative peaks do not become significant until later components are used. As more and more PCs are used, however, more noise contributes to the signal and a fine balance must be made.

6.3.2.4 │ Second derivative spectra for LDA using 90% PC variance explained

To obtain an equivalent 'variance explained' value as the non-derivative dataset at 90%, 8 PCs were used as opposed to 4 PCs. The score plot separation (not displayed) was only slightly better at distinguishing the control from the drug-treated cell spectra.

Loading plot analysis suggested a possible reason for this result. Figure. 6.5a displays the DF1 loading plots for the second derivative forms using 8 PCs. There were notable peaks for amide I (~1657 cm^{-1}) and amide II (~1547-62 cm^{-1}) regions in the loading plot. These signals, however, were indistinguishable from the signals in the DF2 loading plot (not pictured). Discriminant analysis (DA), like PCA is a form of eigen-analysis and for k groups, DA finds the k–1 discriminant axes that maximally separate the k groups. In a two group case of LDA, there will be only one discriminant function that will be meaningful, i.e. one discriminant axis for the two groups. This indicates that these regions covering the amide bands are not that informative in terms of describing different chemical features between the control and the drug treated cells at this 24 time point.

6.3.2.5 │ Second derivative spectra for LDA using 95% PC variance explained

To determine if any discriminative features were present in the loadings, LDA was re-analysed using 14 PCs, accounting for 95% of variance in the 2nd derivative dataset (Fig 6.5b-d). The LDA score plots in all cases showed distinction, separating control cell spectra from those treated with the agents (Fig 6.5b).

For the loadings observed for DF1 (Fig 6.5c), there are two common markers that stand out that are not present in the loadings of DF2, suggesting that the markers are informative with regard to the group separation. The first is at 966 cm^{-1}, associated with C-O stretching of the phosphate monoester group of phospholipids, phosphorylated proteins and nucleic acids [26]. This peak in the loading positively correlates with 1084 cm^{-1} peak in KF13 and 5FU and 1051 cm^{-1} in KF1. The assignment of ~1084 cm^{-1} is mainly linked to a symmetric stretch $\nu_s(PO_2^-)$ of the phosphodiester group of nucleic acids and membrane phospholipids, and partially

protein (amide III) . The absorption at ~1051 cm^{-1} is associated with C-O-C stretching of DNA and RNA [26].

The second universal marker of drug-treated cell spectra versus the control is found at ~1470 cm^{-1}, connected to δCH_2 bending vibration of lipids and proteins and scissoring vibrations of the acyl chains in phospholipids [26]. KF1 appeared to have a unique marker at 1387 cm^{-1} associated with $\delta(CH_3)$ bending modes of lipids.
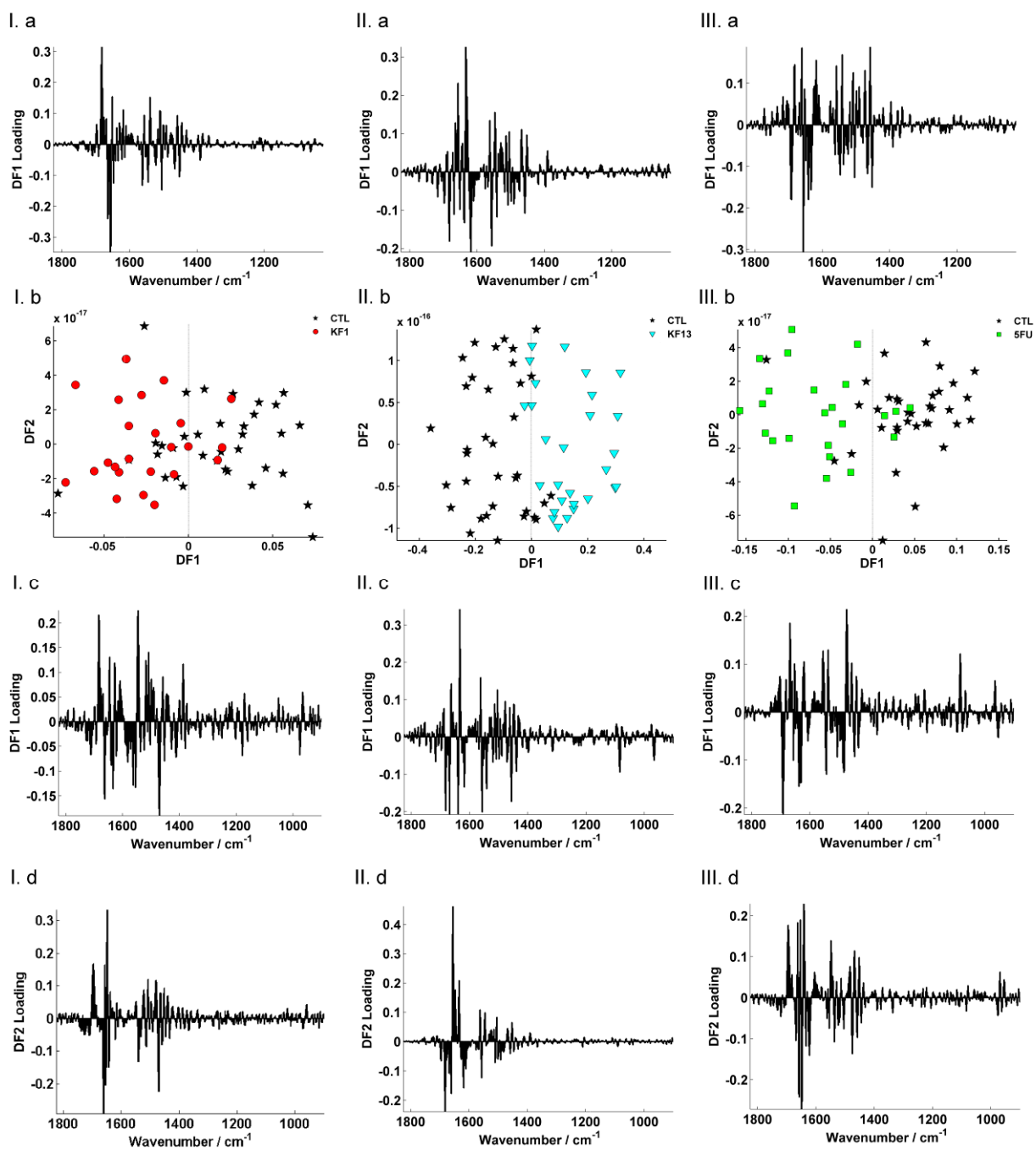
Figure 6.5 │ KF1 (I), KF13 (II) and 5FU (III) displaying LDA loadings (a) using 90% dataset variance from PCA (b) using 95% variance scores and subsequent LDA loadings in DF1 (c) and DF2 (d) for the 2nd derivative dataset.

6.3.2.6 │ Second derivative spectra for multiclass LDA

Data groups CTL, 5FU, KF1 and KF13 were combined and analysed as a whole dataset. Figure 6.6 (I a-c) displays the LDA results of the non-derivative version of the dataset with 4 PCs, where no distinction or interpretation of the loadings could be made, as found in the two class case previously. 4 PCs explained 90% of variance for non-derivative data but only 79% in the 2nd derivative data. Therefore an equivalent 9 PCs was used for the 2nd derivative data and there appeared to be a vague separation of 5FU spectra in DF2 (Fig. 6.6 (II)).

The variance explained for 95% in the 2nd derivative data equated to 16 PCs (Fig 6.6 (III)). The KF13 cluster, being furthest from CTL in DF1, suggests that the cells are responding, or have been affected most to KF13 at this 24 time point, This was also proposed in the single–drug analysis (6.3a). 5FU clustering appeared more distinct in DF2. This is promising as it may suggest that the cells are responding differently, or that the drugs act in a different mode to 5FU.

The use of 95 % variance explained and was tested to further resolve differences in the highly complex dataset. The markers at 966 cm^{-1} and 1084 cm^{-1}, seen before for the single group analysis were clearly resolved. The extra level of noise introduced by adding more PCs did not have an impact from 9 PCs to 16 PCs.
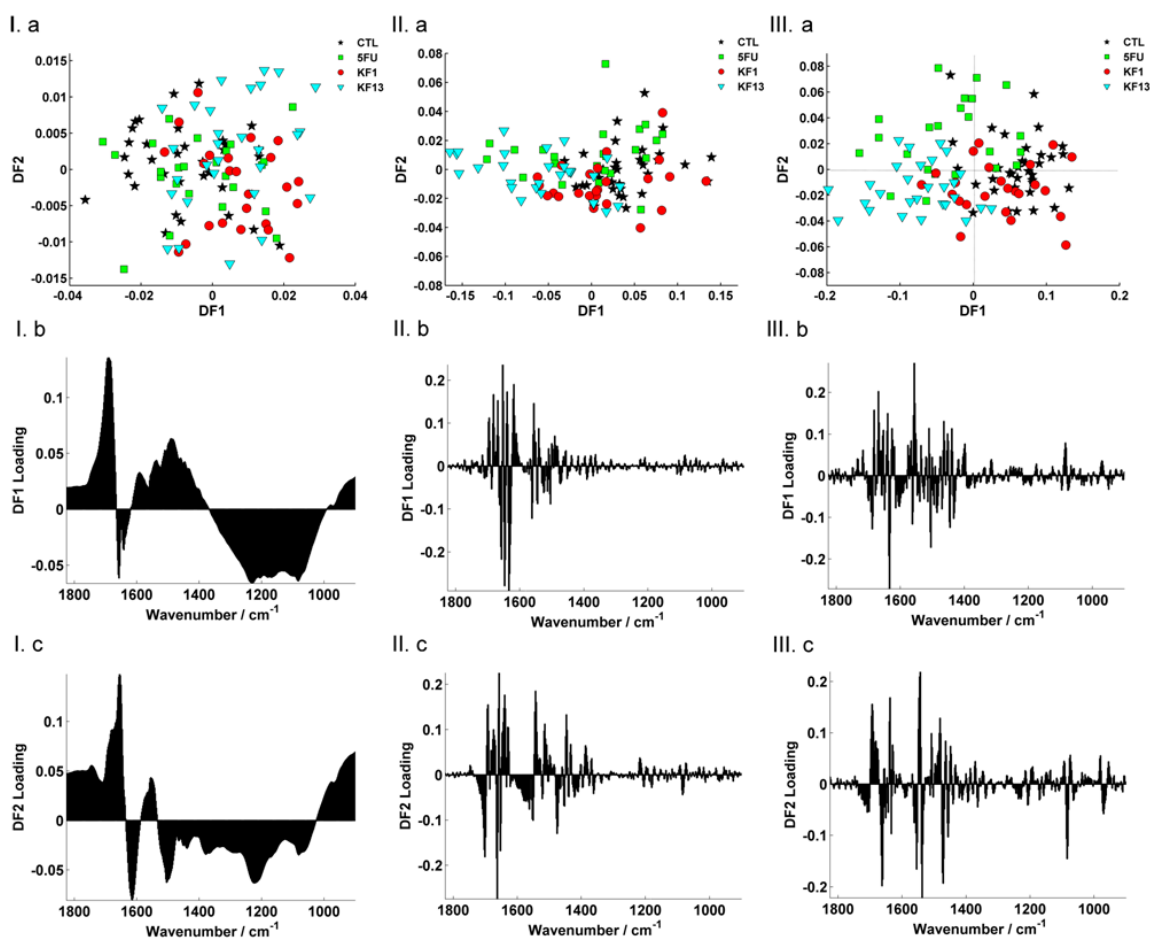
Figure 6.6 │ LDA score plot (a) and loadings (b, c) for non-derivative format for (I) 4 PCs (90% VE) (II) 2nd derivative equivalent (90% VE) and (III)2nd derivative using 16 PCs (95% VE)

There has been some research into the relation of signal to noise contribution of PC loads to determine the number of PCs to use in discriminant analysis and there is no definite threshold [27]. It has been said that determining the PC cut-off by increasing contributions of noise serves as a practical way to deduce a threshold. Ultimately, adding more and more PCs will introduce more noise as exampled in Figure 6.7; where the first PC to introduce 99% VE was 47. This is definitely a case of over-fitting the data, as both the clustering in the DF score plot is perfect, and no markers can be resolved above the nose in the loading plot.
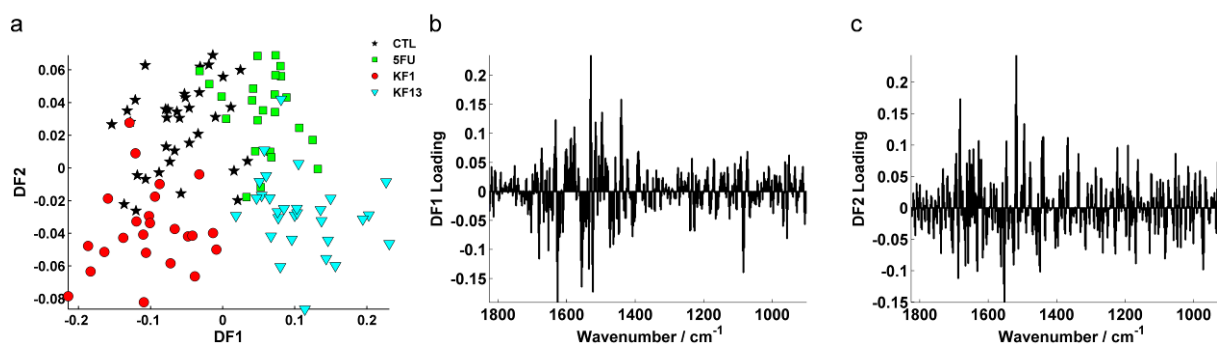
Figure 6.7 │ Example of over-fitting: too many PCs using 2nd derivative data (99% VE) by LDA score plot (a) and the loadings for DF1 and DF2 (b, c)

6.3.2.7 │ Qualitative assessment of mean derivative spectra

In further assessment of the mean second derivative spectra for each group (CTL, 5FU, KF1, and KF13; number of spectra, 'n', was 35, 24, 24, 29 respectively), there no apparent shift in peak minima (Fig. 6.8a). There were, however, five areas where variation was apparent when the derivative mean spectra were overlaid (Fig. 6.8b). The peak minima at these positions were at (1)1171; (2) 1155; (3) 1084; (4) 1061; (5) 968 cm^{-1}. These spectral regions are said to be associated with (1) ν_{as} (CO-O-C); (2) ν_s(C-O); (3) ν_s (PO$_2^-$); (4) ν_s (C-O) and (5) ν_s (C-O).

There is said to be a triad of infrared peaks which are characteristic of nucleic acid vibrations at 1031 cm^{-1}, 1060 cm^{-1}, and 1081 cm^{-1} [28]. Markers at points 3 and 4 may highlight differences in two of these associated triad peaks. Marker 3 is associated with the phosphodiester bonds of DNA whilst marker 4 is due to stretching C-O of deoxyribose. It is likely there are differences in the third referenced nucleic acid peak at ~1031 cm^{-1}, however this peak is situated in a region of noise and therefore it is not possible to make definitive observation for this peak (Fig. 6.8b).

Figure 6.8c displays the un-derivatised, equivalent mean spectra (vector normalised). Throughout the entire region of 1200-900 cm^{-1} there is a considerable decrease in absorption in the KF-treated cell spectra. The predominant peaks around ~1080 cm^{-1} and ~968 cm^{-1} (markers 3, 5 in Fig. 6.8b) also highlight differences in absorbance. Without making assumptions, it may be possible that KF drug activity

with DNA, is detectable, possibly causing DNA damage at this time point. As 5FU is cell cycle specific, it may be that the majority of cells are yet to progress to the next phase at this time point. Whatever the reason for the difference, it does seem apparent that the KF analogues are acting in a different mode, relative to 5FU.
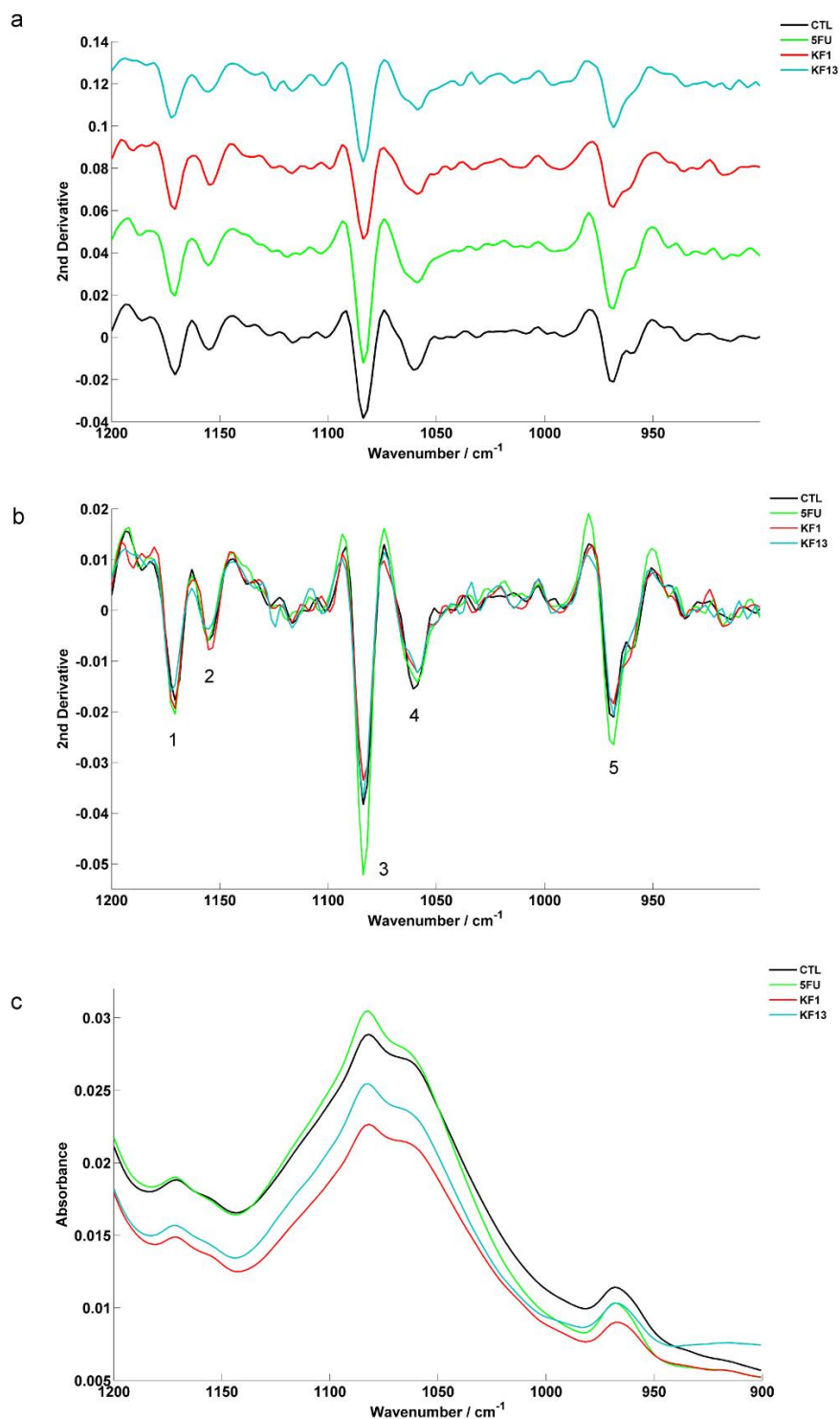
Figure 6.8 │ (a) Mean 2nd derivatives, stacked for each group (CTL, 5FU KF1 and KF13-treated cell spectra). (b) The same mean derivatives, overlaid (c) The non-derivative, normalised equivalents.

6.3.3 │ FPA-FTIR-imaging analysis after 24 hours exposure to compounds

FTIR hyperspectral images were acquired using same samples as for SR-FTIR. As another method of analysing drug-treated cell spectra, these images included data of multiple cells, providing thousands of spectra to manipulate. The aim was to determine of the same trends as seen in the single cell data could be reproduced using LDA, and also to ascertain if clear biomarkers could be determined by using mean spectra and PCA.

Infrared images from the same samples were taken in triplicate per group (CTL, 5FU or KF1) from areas of similar confluence. There were, however, substrate regions with spaces where the cells had not grown (Fig. 6.8a). It is known that cells secrete chemicals into the surrounding area for a number of purposes. Rather than quality testing these 'vacant' areas, removing them from spectral analysis, an alternative approach was taken.

The individual spectral images were corrected using the RMieS-EMSC algorithm. Subsequently, one image each for CTL, 5FU and KF1 were stitched as one mosaic, to analyse the images as a whole. K-means clustering was performed to obtain clusters that described all groups simultaneously (Fig 6.8b). The spectra were not normalised and therefore K-means clustering classed the spectra into groups predominantly by sample thickness. This can be seen in Figure 6.8c where the class colours match thin (class 2; orange), medium (class 1; grey) and thick (class 3; white) areas of the sample.
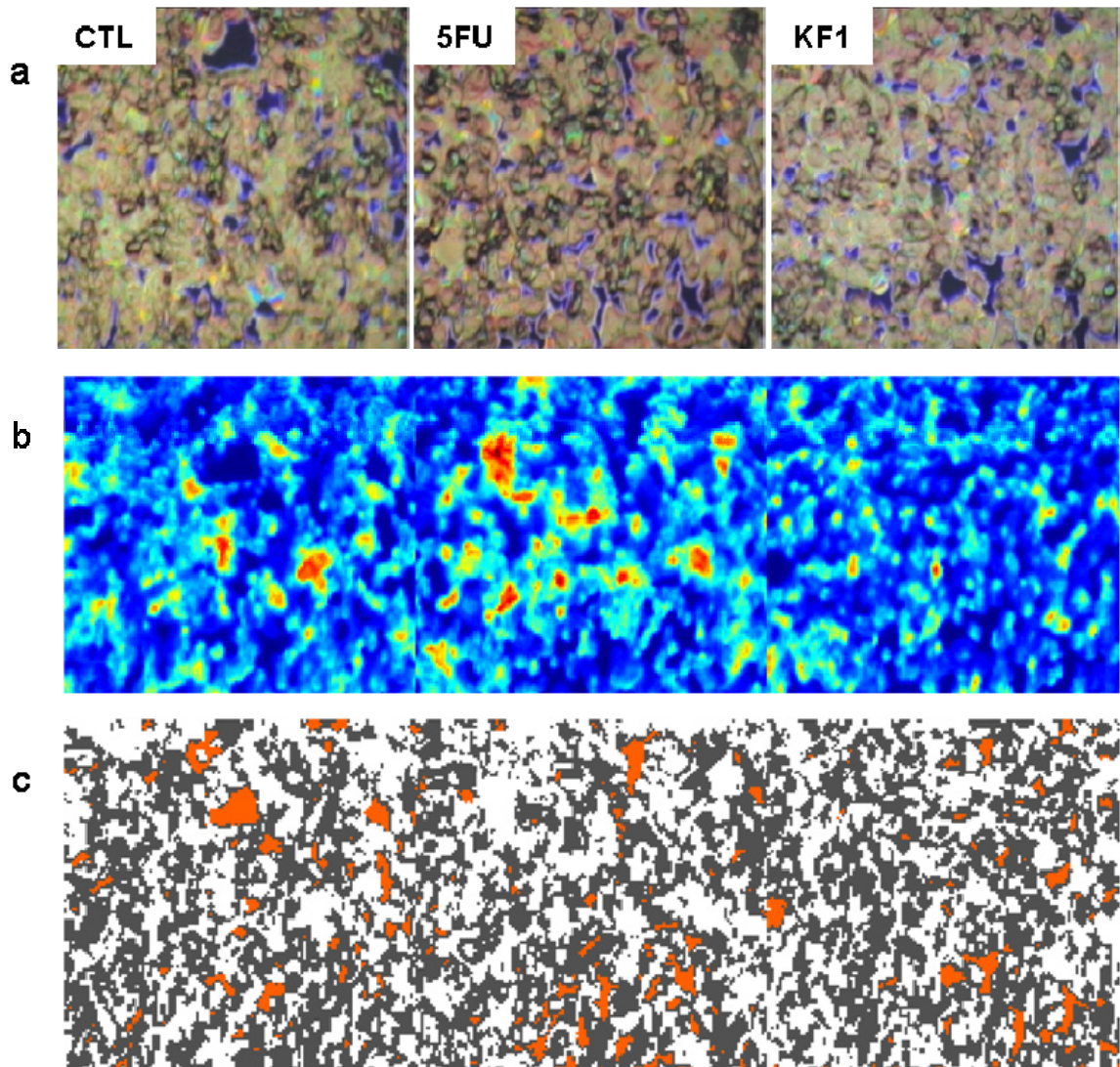
Figure 6.8 │ (a) optical images of the sample regions (b) IR hyperspectral images were acquired for each sample type and stitched together. Total intensity infrared images where intensity ranges from blue (minima) to red (maxima). (c) The resultant k-means clustering of the three images, split into class 1 (grey), class 2 (orange) and class 3 (white).

For robustness, the process was repeated a further two times in different areas of each slide. Figure 6.9 displays total intensity and corresponding k-means clustering for the replicate groups (a), (b) and (c).
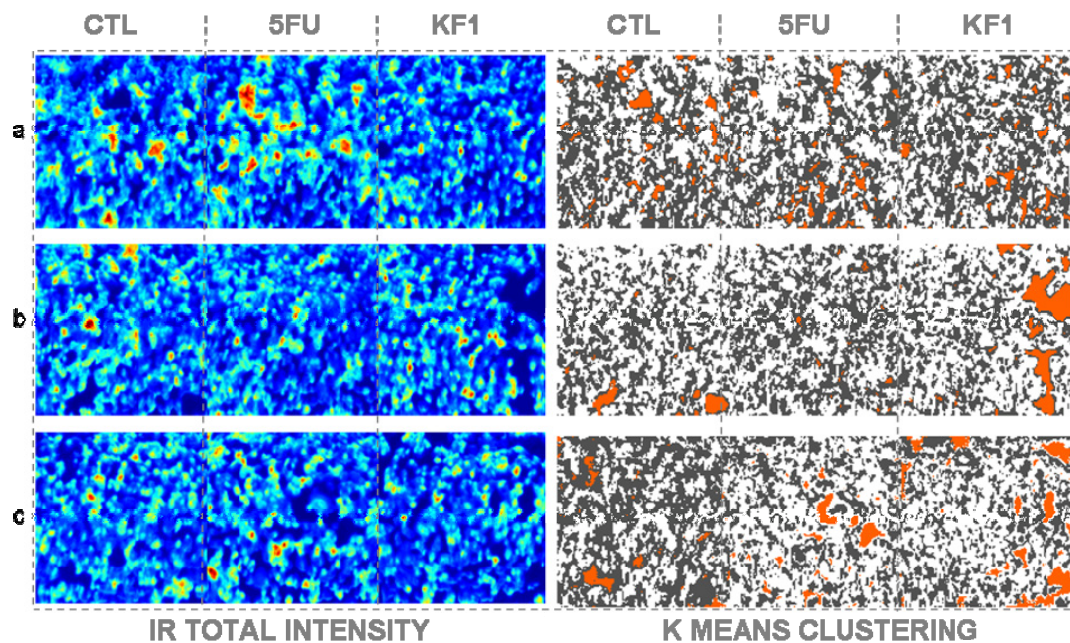
Figure 6.9 │ Stitched tiles of total intensity for replicates a, b and c (left) and the corresponding k-means clustering analysis (right).

The resulting mean spectra representing each replicate and for each group (CTL, 5FU and KF1) can be seen in Figure 6.10. Cancer cells have less contact inhibition than non-cancer cells and are more likely to grow on top of one another [29]. This is possibly why there are thickness differences on each slide. Alternatively, it may be a case of random attachment to the substrate when the cells are first introduced into the substrate-containing well. Class 3 spectra highlight the thickest sample areas displaying the highest absorbencies in the spectra.

Class 1 was most abundant (~8,000 spectra) and class 2 the least (~200 spectra). The class 2 spectra displayed characteristic fringing patterns and represented the thinnest sample areas of the slide. It was interesting to note that class 2 areas appeared vacant in the optical images, yet signals rich in protein, nucleic acid, lipid and carbohydrate are present. These spectral features may be evidence of cellular secretions.
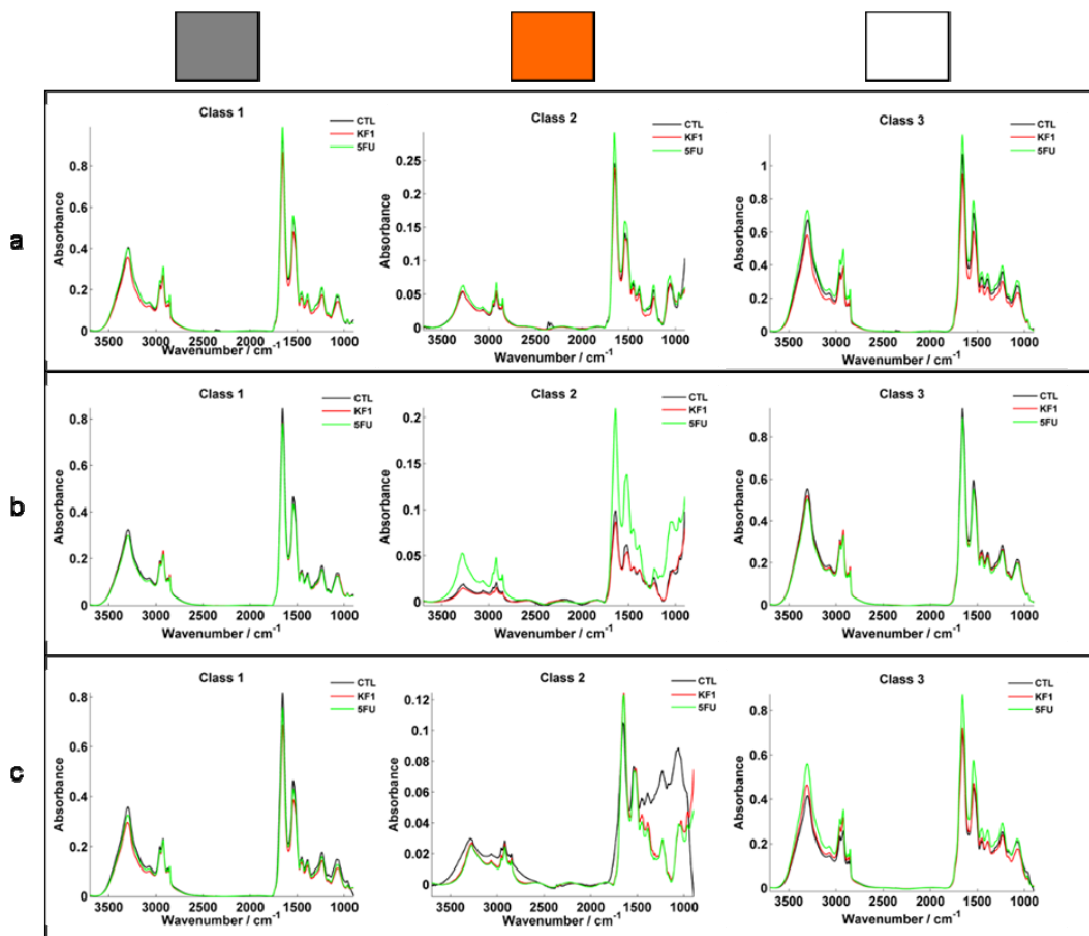
Figure 6.10 │ Mean spectra are represented for CTL, KF1 and 5FU for each class. (a), (b) and (c) are replicates equivalent to a,b and c in Fig. 6.9. Class 1 represent the grey shading in the K means cluster images, and are spectra of medium thickness. Class 2 represent the orange-shaded areas of the cluster images, with spectra of very low absorbencies where no cells appeared to be present. Class 3 represent the white-shaded cluster image areas and highlight regions of greater absorbencies, particularly in the fingerprint region.

The three classes of spectra were analysed separately by PC-LDA using 15 PCs (60% of variance explained). The LDA results in general (Fig 6.11) echo the general trend seen for the initial SR-FTIR dataset (Fig 6.7a) as cells treated with 5FU seem to project more on the DF2 axis, whereas the KF drugs appear further along the DF1 axis, respective to the control cells.
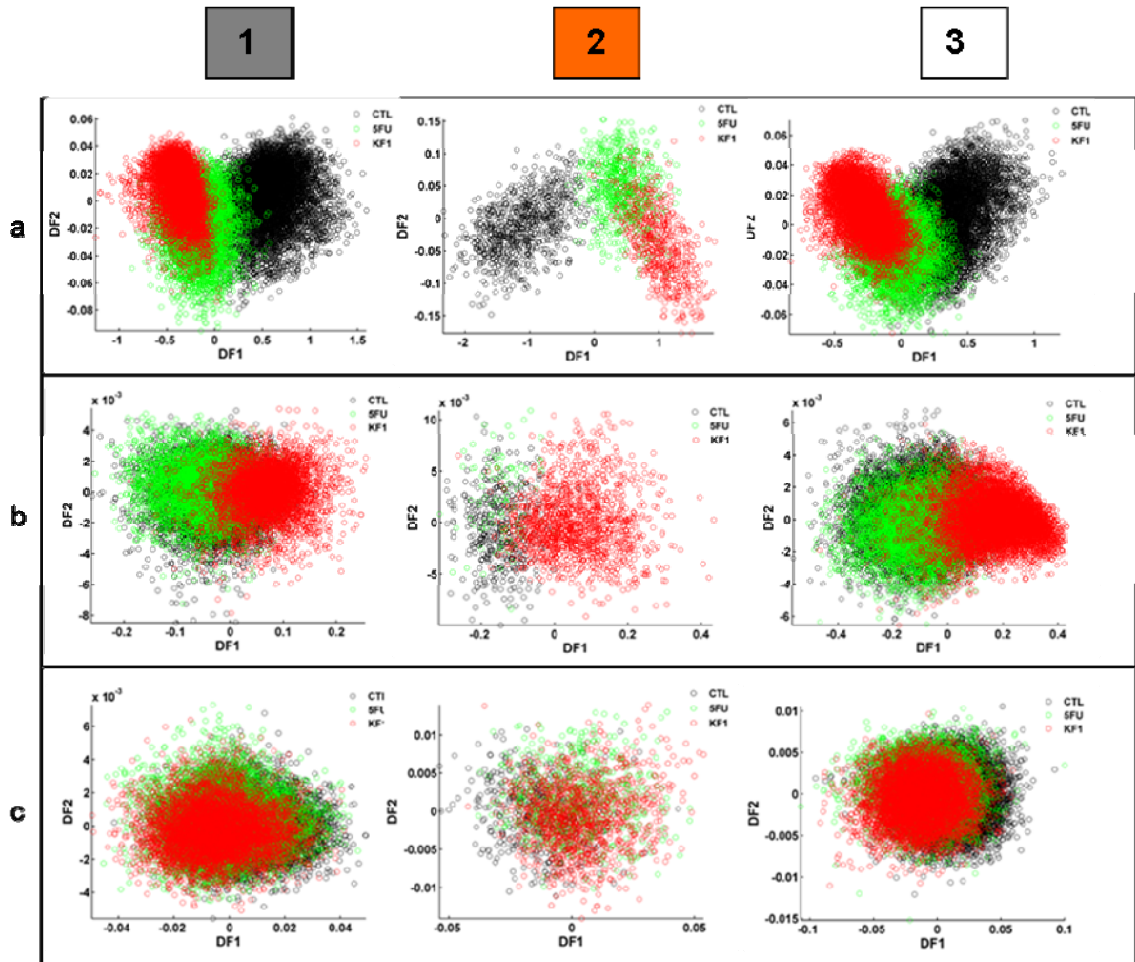
Figure 6.11 │ LDA scores of the spectral groups shown in Figure 6.10, using 15 PCs. Typically 15 PCs accounted for ~60% of variance. (a), (b) and (c) are replicates equivalent to a,b and c in Fig. 6.9 and in Fig. 6.10.

It is clear, however, that even though the mean spectra per replicate appear similar by class, the discriminant analysis of replicates in different areas of the sample illustrate that the trend is not completely representative of the entire cell population and there are fluctuations. This may be due to differences in cell chemistry as a result of cell cycle differences or cell density.

6.3.4 │ Single cell SR-FTIR analysis after 72 hours exposure to KF compounds

The clustering patterns from PC-LDA in the 24 hour datasets suggested that the Caki-2 cells may have been more sensitive to KF13 than KF1. This was most apparent in Figure 6.6 (III a) whereby KF13 cluster points were the furthest removed from the CTL data points on the DF1 axis.

As the calculated IC_{50} values were based on a three-day cytotoxicity assay it was thought necessary to see any differences in cell response at this same time point i.e. after 72 hours of drug exposure. At this time point, we could be sure that at least half of the cellular population should have been inhibited by the agent concentration and possibly larger spectral differences. Due to the washing steps involved before formalin fixation, dead cells and debris that would have lifted off the surface of the substrate would have been removed. In general, only viable cells that either did not receive a dose of drug or chemically differentiated cells that may show resistance would have remained.

Figure 6.12a displays the PC-LDA score plot of the total dataset (CTL, KF1 and KF13), compiled using 4 PCs and non-derivative spectra. Relative to KF1, KF13 data points again appeared to be more differentiated from the control. There was also a larger spread of data points for KF13. This could also be observed in the equivalent group mean spectra. KF13 appears to have greater variance in the protein amide I and II peaks (Fig 6.12b).
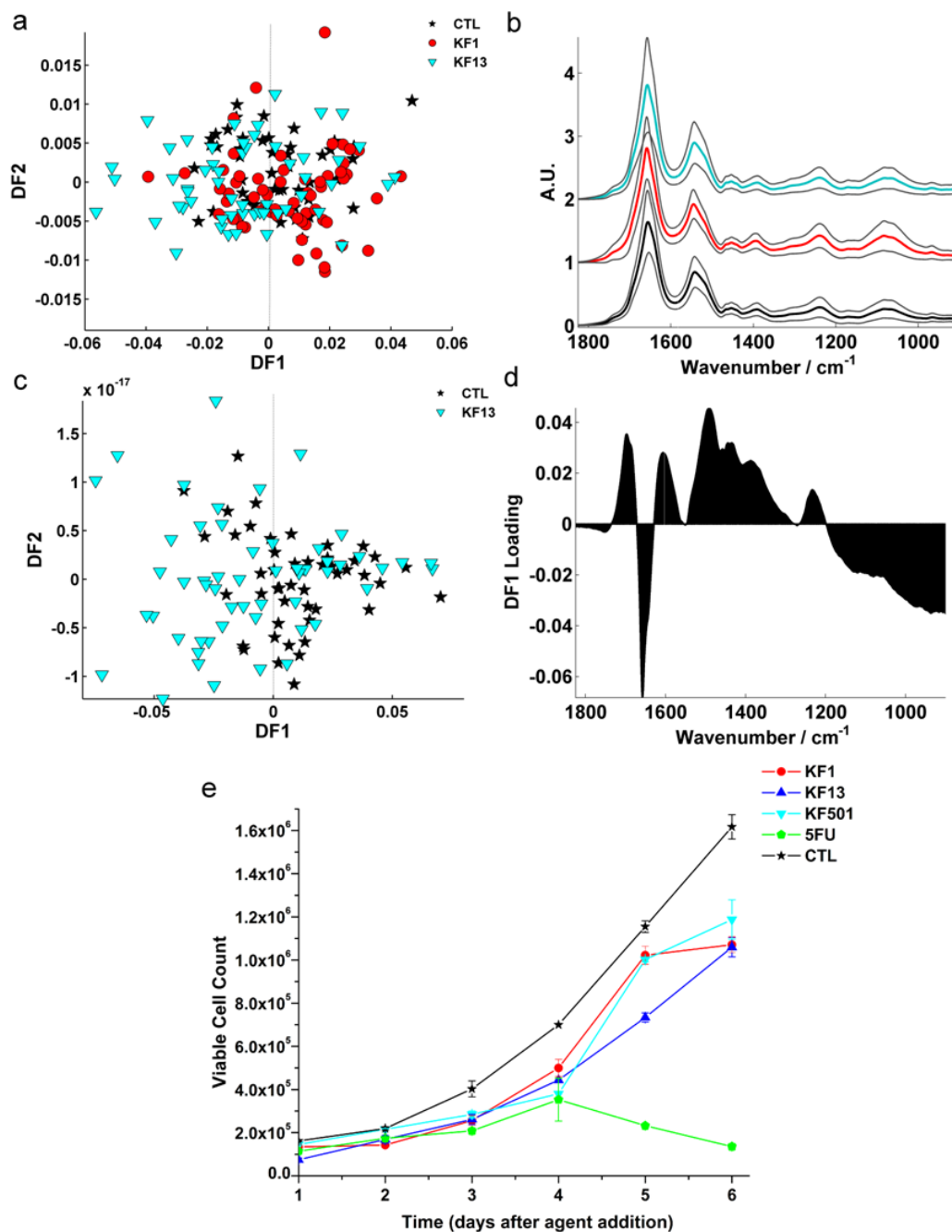
Figure 6.12 | LDA score plot of the total dataset for drug exposure over 3 days (a). Mean spectra ± standard deviation for each group. KF13 has maximum variance in the amide regions of the fingerprint (c). LDA score plot (d) and DF1 loading (e) for KF13 spectra against CTL spectra only. (f) displays viable cell proliferation counts acquired over a period of 6 days.

The KF13 dataset was analysed singularly against the control spectra (Fig 6.12c) in the same manner as the 24 hour equivalent set (Fig. 6.3). It appeared that the drug-treated cell spectra were again potentially distinguishable from the non-drug treated control as there is some separation in PC1. As well as changes in the amide I band, differences in the nucleic acid regions below 1200 cm^{-1} were anti-correlated with the mid-fingerprint range (1200-1500 cm^{-1}), rich in lipid and carbohydrate signals (Fig 6.12d).

There are many possible reasons as to why the drug-treated cell spectra are distinguishable against the control spectra due to a multitude of processes occurring:-

- The cells may be sensitive to the drugs and the cells are cells are being damaged
- The cells are responding to the drugs and are actively trying to repair and remove the foreign molecules
- Upon exposure to the KF agents, the cells have evolved and developed resistance to and that is why their infrared signature is separable to the sensitive control cell population.

A 6 day proliferation assay was performed in order to try and determine the eventuality of the cellular response (Fig 6.12e). In all cases, the cells treated with the KF series seemed to recover in viable cell number after 72 hours. In contrast, the viable cell count for 5FU treated cells remained more or less constant and relatively low, eliciting a profile indicative of a cytostatic drug effect.

6.3.5 │ Comparison between 3/6 days after initial drug exposure

The results of the proliferation assay in 6.3.4 prompted a final investigation into the cellular signatures after several days of drug exposure. In particular, it was questioned if the KF-treated cell population at 'day 6' were now indistinguishable to the control cells, as they were under proliferative recovery.

The single cell analysis in previous sub-chapters had highlighted issues affecting interpretation. There was a high level of data complexity, due to parallel processes in the cell cycle. The issue with choosing the number of PCs for LDA was also subjective. Upon consideration of the issues associated with single cell analysis it is perhaps more advantageous to acquire single-point spectra of cell monolayers and use PCA alone to ascertain a general trend. The single point technique has the advantage that a large aperture may be used to capture the average signal of several cells in one area of the substrate in a single acquisition. This approach also reduces the impact of cell cycle differences in chemistry.
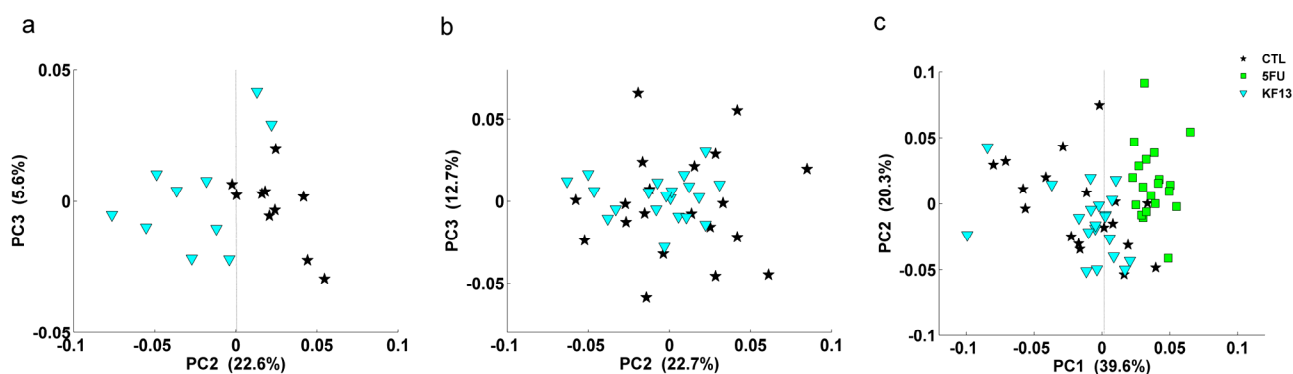
Figure 6.13 │ PCA of single-point monolayer spectra of (a) KF13 versus CTL after 3 days (b) KF13 versus CTL after 6 days (c) 5FU and KF13 versus CTL after 6 days of exposure

A new set of samples and a second proliferation assay were prepared for the final experiment. Figure 6.13 displays the PCA plot of the non-derivatised spectra. KF13 data points were distinguishable against the control only at the 72 hour time point (Fig. 6.13a). There was no distinction of theses two groups after 6 days of drug exposure (Fig 6.13b). Figure 6.13c displays the same KF13 data as figure (b) but with added 5FU data as an additional validation of this result. 5FU treated cells were very different from the control cells in terms of cell viability from days 3-6 in the previous proliferation assay.
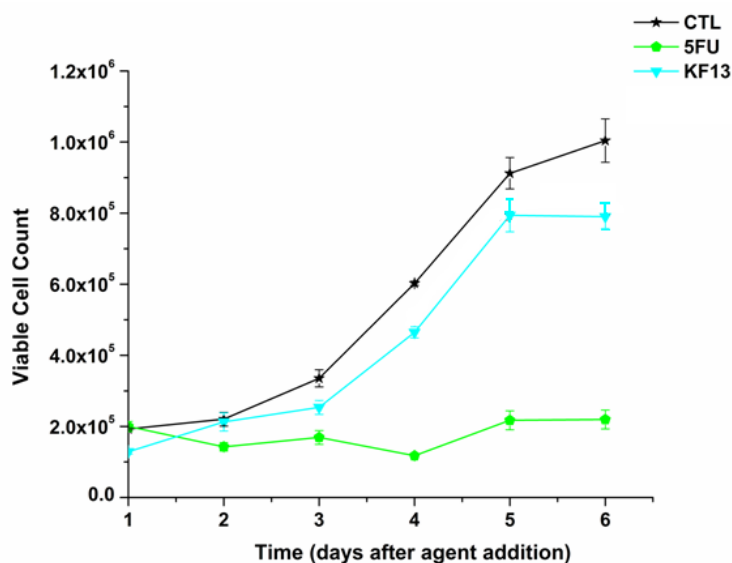
Figure 6.14│ The repeat of the cell viability proliferation assay for CTL, KF13 and 5FU treated cells.

The second proliferation assay displayed the same pattern of proliferation (Fig. 6.14). The difference between 5FU-treated cells and the CTL/KF13 cells were again apparent. A proportion of 5FU-exposed cells remained viable but no longer proliferated, whereas the KF-exposed cells recover from proliferation inhibition after three days. This result could support why the chemistry of 5FU-cell spectra separate from the chemistry of the KF-treated cell spectra in the PCA score plot (Fig. 6.13c).

Visual differences were also noted in the in-vitro photomicrographs taken for CTL, 5FU and KF13 (and KF1 in addition) in Figure 6.15 at the 6th day before cell counting protocol was implemented. There was a large display of un-healthy apoptotic cells in the 5FU image, with instances of cell blebbing. In contrast, there were smaller instances of blebbing however in general KF treated cells appeared similar in morphology to the control cells, suggesting a similar viability.
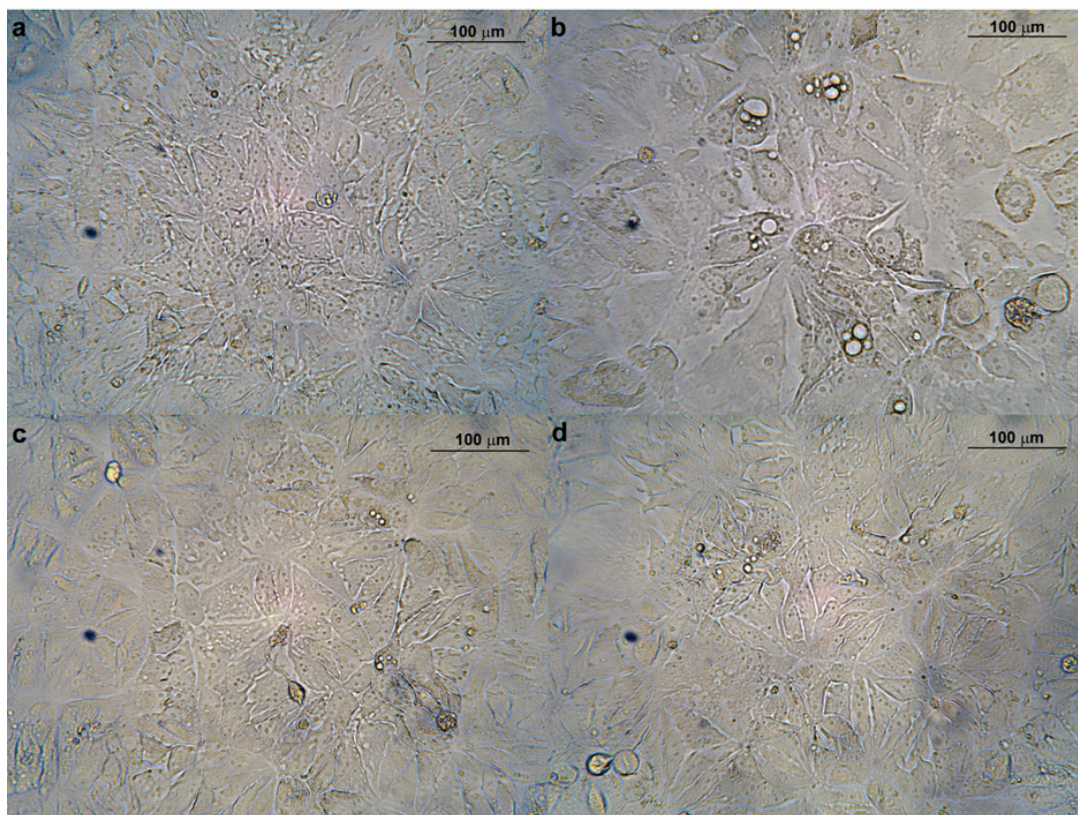
Figure 6.15 │ In-vitro photo-micrographs of the Caki-2 cells after 6 days of drug exposure: (a) CTL, (b) 5FU (c) KF1 (d) KF13

6.4 │Conclusions and Future Prospects

The analysis has shown heterogeneity in the cell populations treated with different chemotherapeutic agents. This was observed in the spread and overlap of data point for the single cell experiments and in the hyperspectral FTIR images. Single-point spectra of large cell monolayer areas, in combination with PCA was shown to be the best current method for detecting overall trends in the different cellular responses.

One reason may be simply due to sampling numbers: ten to twenty spectra were taken for the single point data, encompassing numerous cells simultaneously. In comparison, the imaging investigation was performed with only 3 images per drug variable, due to the nature and size of the files. However, it could be said that the imaging technique provided the advantage of spatial resolution, where each

spectrum related to a pixel are of 5.5 μm^2. This meant that individual cells would have been probed every few pixels and cell cycle differences may have contributed to the differences in trends. It may also be that certain areas of the sample were denser than others. It has been shown that chemical differences in the infrared spectra of cells may differ depending on whether cells are in the exponential (proliferating) or plateau (non-proliferating) phase of growth [31].

As cells communicate with one another receiving chemical signals from nearby cells [1] (Ch. 1.1.2), this may also account for the different relationship patterns in the areas surrounding the cells (low-absorbance spectra, k means class 2, orange cluster areas), which most likely included traces of cellular chemical secretions. The mean spectra from these regions displayed relatively high intensities in the sub-900 cm^{-1} spectral region. (The area 1200-800 cm^{-1} included many nucleic acid signals [30]). In selected areas of the sample, these class 2 spectra were sometimes differentiated in terms of drug variable (CTL, 5FU or KF1) during discriminant analysis.

It is also possible that the spectral signatures in these sparse regions arose from cell secretions or signals due to stress. Normally in the human body, dead cells can attract phagocytes that engulf the dead cells before their cell membrane integrity is lost. This prevents the release of toxins into the extracellular matrix which would otherwise affect nearby cells [33]. In an *in-vitro* environment, the cell line cannot exhibit such a mechanism and it is feasible that there may be traces of cell debris or toxins.

The KF analogues appeared to recover in viable proliferative cell count, as seen by the proliferation assays after 3 days and the PCA plot after 6 days. One possible explanation is that the drugs were readily effluxed out of the cells, i.e. the drug effect was no longer observed as the drug was no longer present; P-gp and other membrane transporters such as MRP1, and ABCG2 recognize a wide spectrum of compounds, however, they share a common feature in that their transport substrate are typically hydrophobic, especially the case for P-gp and ABCG2. Therefore many drugs that exhibit similar hydrophobic properties could be substrates for these

drug efflux pumps and may readily cross lipid bilayer by passive diffusion [34]. For P-gp substrates, the now widely accepted 'hydrophobic vacuum cleaner' model, accounts for the lipophilic nature of P-gp substrates [34-36]. It is suggested that drugs that have entered into the phospholipid bilayer are pumped out of the membrane by the P-gp pump and are expelled into the extracellular aqueous phase.

In terms of the KF-series for renal cell carcinoma, this drug efflux proposal could be analysed further by means of blocking the efflux ability of the cells in a similar fashion as the use of Verapamil to sort the side population cells, enriched for stem-like cells (Ch. 5).

Further investigation for the methodology points towards a simplification step whereby perhaps differences in cell cycle chemistry could be separated from the different effects of particular anti-cancer drugs.

6.5 │ References

1. Lamberti A, Sanges C, Arcari P: **FT-IR spectromicroscopy of mammalian cell cultures during necrosis and apoptosis induced by drugs**. *Spectroscopy*, **24**(5):535-546.

2. Ahmed GAR, Khorshid FAR, Kumosani TA: **FT-IR spectroscopy as a tool for identification of apoptosis-induced structural changes in A549 cells treated with PM 701**. *International Journal of Nano and Biomaterials* 2009, **2**(1-5):396-408.

3. Mourant JR, Yamada YR, Carpenter S, Dominique LR, Freyer JP: **FTIR spectroscopy demonstrates biochemical differences in mammalian cell cultures at different growth stages.** *Biophysical Journal* 2003, **85**(3):1938-1947.

4. Zhao R, Quaroni L, Casson AG: **Fourier transform infrared (FTIR) spectromicroscopic characterization of stem-like cell populations in human esophageal normal and adenocarcinoma cell lines.** *Analyst* 2010, **135**(1):53-61.

5. Di Giambattista L, Pozzi D, Grimaldi P, Gaudenzi S, Morrone S, Congiu Castellano A: **New marker of tumor cell death revealed by ATR-FTIR spectroscopy.** *Analytical and Bioanalytical Chemistry* 2011:1-8.

6. Zelig U, Kapelushnik J, Moreh R, Mordechai S, Nathan I: **Diagnosis of cell death by means of infrared spectroscopy**. *Biophysical Journal* 2009, **97**(7):2107-2114.

7. Petibois C, Deleris G: **Evidence that erythrocytes are highly susceptible to exercise oxidative stress: FT-IR spectrometric studies at the molecular level.** *Cell Biology International* 2005, **29**(8):709-716.

8. Shoemaker RH: **The NCI60 human tumour cell line anticancer drug screen**. *Nature Reviews Cancer* 2006, **6**(10):813-823.

9. Juranka PF, Zastawny RL, Ling V: P-glycoprotein: **Multi-ddrug resistance and a superfamily of membrane-associated transport proteins**. *FASEB Journal* 1989, **3**(14):2583-2592.

10. Wu XX, Kakehi Y: **Enhancement of lexatumumab-induced apoptosis in human solid cancer cells by cisplatin in caspase-dependent manner.** *Clinical Cancer Research* 2009, **15**(6):2039-2047.

11. Ott I: **On the medicinal chemistry of gold complexes as anticancer drugs.** *Coordination Chemistry Reviews* 2009, **253**(11-12):1670-1681.

12. Flower K, Mcgown A: **Use of a compound having two gold (I) atoms each covalently binded to a carbon atom in a covalent link connecting the two gold (I) atoms for the treatment of cancer**. WO/2005/058421.

13. Derenne A, Gasper R, Goormaghtigh E: **The FTIR spectrum of prostate cancer cells allows the classification of anticancer drugs according to their mode of action**. *Analyst* 2011, **136**(6):1134-1141.

14. Flower KR, Khalifa I, Bassan P, Demoulin D, Jackson E, Lockyer NP, McGown AT, Miles P, Vaccari L, Gardner P: **Synchrotron FTIR analysis of drug treated ovarian A2780 cells: An ability to differentiate cell response to different drugs?** *Analyst* 2011, **136**(3):498-507.

15. Draux F, Jeannesson P, Gobinet C, Sulé-Suso J, Pijanka J, Sandt C, Dumas P, Manfait M, Sockalingum GD: **IR spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells**. *Analytical and Bioanalytical Chemistry* 2009, **395**(7):2293-2301.

16. Vichai V, Kirtikara, K.: **Sulforhodamine B colorimeric Assay for Cytotoxicity Screening**. *Nature Protocols*. vol. **1**; 2006.

17. Barreiro E, Casas JS, Couce MD, Sánchez A, Sánchez-Gonzalez Á, Sordo J, Varela JM, Vázquez López EM: **Synthesis, structure and cytotoxicity of triphenylphosphinegold(I) sulfanylpropenoates**. *Journal of Inorganic Biochemistry* 2008, **102**(2):184-192.

18. Lock EA, Reed CJ, Kinsey GR, Schnellmann RG: **Caspase-dependent and -independent induction of phosphatidylserine externalization during apoptosis in human renal carcinoma Cak1-1 and A-498 cells**. *Toxicology* 2007, **229**(1-2):79-90.

19. Griffin JP: The **Textbook of Pharmaceutical Medicine**. 6th Edn., *Wiley-Blackwell* 2009, p29, 174

20. Brown D: **Unfinished business: target-based drug discovery**. *Drug Discovery Today* 2007, **12**: 23-24

21. Defernez M, Kemsley, E.K.: **The use and misuse of chemometrics for treating classification problems**. *Trends in Analytical Chemistry* 1997, **16**:216-221.

22. Kaiser HF, Norman WT: **Coefficients alpha for components**. *Psychological Reports* 1991, **69**(1):111-114.

23. Kanyongo GY: **The influence of reliability on four rules for determining the number of components to retain**. *Journal of Modern Applied Statistical Methods* 2006, **5**(2):332-343.

24. Watkins MW: Monte **Carlo PCA for Parallel Analysis**: *State College, PA: Ed & Psych Associates*; 2000.

25. Ledesma RD, Valero-Mora P: **Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis.** In: *Practical Assessment, Research & Evaluation*. vol. **12**; 2007.

26. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

27. Grimbergen MCM, Van Swol CFP, Kendall C, Verdaasdonk RM, Stone N, Bosch JLHR: **Signal-to-noise contribution of principal component loads in reconstructed near-infrared Raman tissue spectra**. *Applied Spectroscopy* 2010, **64**(1):8-14.

28. Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, Diem M: **Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix.** *Biospectroscopy* 1998, **4**(1):47-53.

29. Elliott WH, Elliott, DC, **Biochemistry and Molecular Biology**, *Oxford University Press* 2001: 495-502.

30. Draux F, Jeannesson P, Gobinet C, Sule-Suso J, Pijanka J, Sandt C, Dumas P, Manfait M, Sockalingum GD: **IR spectroscopy reveals effect of non-cytotoxic doses of anti-tumour drug on cancer cells**. *Analytical and Bioanalytical Chemistry* 2009, **395**(7):2293-2301.

31. Mourant JR, Yamada YR, Carpenter S, Dominique LR, Freyer JP: **FTIR spectroscopy demonstrates biochemical differences in mammalian cell**

cultures at different growth stages. *Biophysical Journal* 2003, **85**(3):1938-1947.

32. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

33. Nagata S, Hanayama R, Kawane K: **Autoimmunity and the Clearance of Dead Cells.** *Cell* 2010, **14**0(5):619-630.

34. Eckford PDW, Sharom FJ: *ABC efflux pump-based resistance to chemotherapy drugs.* **Chemical Review**s 2009, **109**(7):2989-3011.

35. Higgins CF, Gottesman MM: **Is the multidrug transporter a flippase?** *Trends in Biochemical Sciences* 1992, **17**(1):18-21.

36. Del Moral RG, Olmo A, Aguilar M, O'Vaue F: **P glycoprotein: A new mechanism to control drug-induced nephrotoxicity**. *Experimental Nephrology* 1998, **6**(2):89-97.

Chapter 7 │ Discriminating Chemical Responses by Cell Cycle Phase

7.1 │ Introduction and Aims

The previous chapter outlined the difficulties in trying to separate cellular response from different agents. PCA score plots of the cell spectra, displayed widely dispersive intra-group clustering, as observed in Chapter 6 (Fig. 6.6) and by K. Flower et al. [1]. It is suggested that there may be small sub-clusters where spectra are grouping in PC1, predominantly because of cell cycle signatures, rather than slight chemical changes that may occur by cellular drug response. A collection of SR-FTIR spectra from a 'typical' population will yield a mixture of cell spectra displaying different quantities of biomolecular components, depending on the cell phase at the time of chemical fixation.

It is possible to sort and collect cells in different phases of the cell cycle through the use of a fluorescence-activated cell sorting (FACS) which analyses the DNA content of the cell, by use of a fluorescent dye that quantitatively binds to it (Chapter 2) [2].

Cells in the resting phase are G0/ G1, the pre-DNA synthesis phase. The cells in the S phase are actively undergoing DNA-synthesis and contain increasing quantities of DNA [3]. In clinical terms, patient tumours observed to have a high S phase fraction are typically said to be at high risk from tumour recurrence [4-5]. The post DNA-synthesis phase is referred to as G2/M where cells have twice the amount of DNA as the G0/G1 phase as they are entering into mitosis.

Propidium iodide (PI) stoichiometrically binds to double-stranded DNA by intercalation. As a ligand, the die fits between the base pairs of DNA, causing inevitable conformational changes in the DNA [2].

Using this dye for cellular spectral analysis and PCA discrimination will lead to an obvious cluster separation of the cells from different phases, as presumably the higher the PI concentration the greater the extent of DNA conformational change (Fig 7.4b).

Due to the stoichiometric nature of the staining reaction mechanism, however, it would still be possible to evaluate a single cell line treated with different agents with this approach, provided that the cell spectra were analysed from one phase at a time. This would be viable as the PI concentration per cell cycle phase would generally be a constant in one cell line and therefore can be omitted as a factor of multivariate discrimination.

The aim was to determine if cellular responses could be detected without having to consider cell cycle differences. It was predicted that interpretation and possible biomarker generation from the multivariate analysis loading plots would also be clearer and perhaps more accurate.

7.2 │ Experimental

7.2.1 │ SRB cytotoxicity assay

The cytotoxicity assay followed the same protocol as described in Chapter 2.4.4 to obtain IC_{50} values for the A-498 cell line.

7.2.2 │ FACS

The experimental procedures in Chapter 2.4.7 describe the experimental setup to enable cells to be stained with propidium iodide and sorted using flow cytometry.

7.2.3 │ Spectral sample preparation

Following cell sorting cells were spun and re-suspended in 150 μL of ddH_2O. The aliquots were deposited with a cell cytospinner (Ch. 2.4.2.2) onto CaF_2 substrates at 800 g for 5 minutes

7.2.4 │ Data acquisition

SR-FTIR transmission spectra were acquired on the SMIS beamline according to the protocol in Ch. 2.4.3.2. Spectra were recorded at 4 cm^{-1} resolution with 256 co-scans.

7.2.5 │ Data pre-processing and analysis

The day 1 control dataset contained 73 spectra in total and the day 3 dataset (control, 5FU treated or KF1 treated) contained 113 spectra. For the day 3 dataset a total of 40 spectra were recorded for each cell type. The raw spectra were corrected with a Matrigel reference spectrum and 150 iterations of the RMieS-EMSC algorithm. The corrected spectra were vector normalised and mean centred prior to PCA, using the spectral range 1110-1825 cm^{-1}.

7.3 │ Results and Discussion

7.3.1 │ Cell cytotoxicity
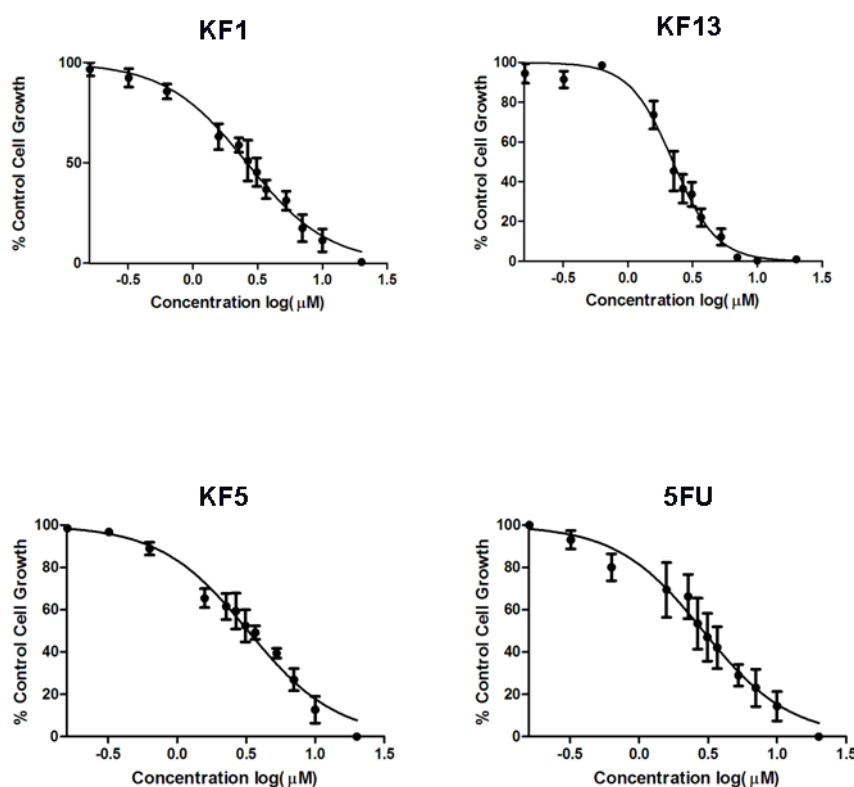
Figure 7.1 │ Dose response curves to determine IC$_{50}$ values for the A-498 cell line from the SRB assay data (repeated in triplicate). Graphs were created in Graphpad Prism using mean values ± standard error.

Both cell lines used throughout the course of the drug-response research were closely matched, defined as primary RCC with wildtype p53 expression [6]. As in Chapter 6 for the Caki-2 cell line, the IC_{50} values had to be obtained for the A-498 cell line. These were found to be: KF1 [2.61 μM]; KF13 [2.23 μM]; KF5 [3.26 μM]; 5FU [2.91 μM]. The IC_{50} values obtained from the replicate SRB assays revealed that the cell line A-498 was more sensitive to the gold-based agents than the Caki-2 cell line (Fig. 7.1), whereas 5FU did not vary distinctively. This may be due to P-gp expression differences as P-gp does not export small hydrophilic drugs such as Cisplatin, 5-Fluorouracil [7]. The A-498 cell line was used in this instance to possibly detect a greater response within the cells, with respect to the non-treated control cells.

7.3.2 │ FACS cell cycle sorting

Cells were collected after 24 and 72 hours of exposure to the agents and stained with propidium iodide for FACS sorting. There appeared no apparent discrepancies between control cells and KF1 treated cells during both time periods. The only unique histogram profile was obtained for cells 72 hours upon treatment with 5FU, highlighting cell cycle arrest in the S phase as expected (Fig. 7.2).

As described in further detail in Chapter 1, 5-fluorouracil belongs to the anti-metabolite category of chemotherapy agents. Anti-metabolites are very similar to normal substances within the cell affecting cells at very specific phases in the cycle. When the cells incorporate these substances into the cellular metabolism, they are unable to divide [8].

Interestingly in Figure 7.2, the effect of 5FU at the IC_{50} concentration was not observed until 72 hours of exposure to the drug. This time period coincides with the length of time the IC_{50} value was determined by the 72 hours cytotoxicity SRB assay. It may simply be that a higher proportion of cells are affected by the

194

concentration of 5FU at the 3 day time point and this is reflected by a notable change in the cell numbers at each phase.

5FU is said to inhibit cell cycle progression at S phase [8].; however we do not see a change in the FACS cell cycle profile until 72 hours after drug addition. The result may also suggest that 5FU did not have an immediate notable effect on the cell population 24 hours after drug addition. The cells were to establish onto the substrates in the tissue culture wells for 24 hours prior to drug addition, making a total of 48 hours for cells to complete their cycle. Although 48 hours seems a substantial time period for cell cycle competition, the cells may have remained in G0/1 phase longer when cultured onto a MirrIR substrate rather than the more preferable conditions of a tissue culture plastic alone. At the time point 72 hours after drug addition, the cells had been cultured for a total of 5 days in the substrate environment, a suitable duration for cells to settle onto the substrate and attempt to complete the cell cycle.
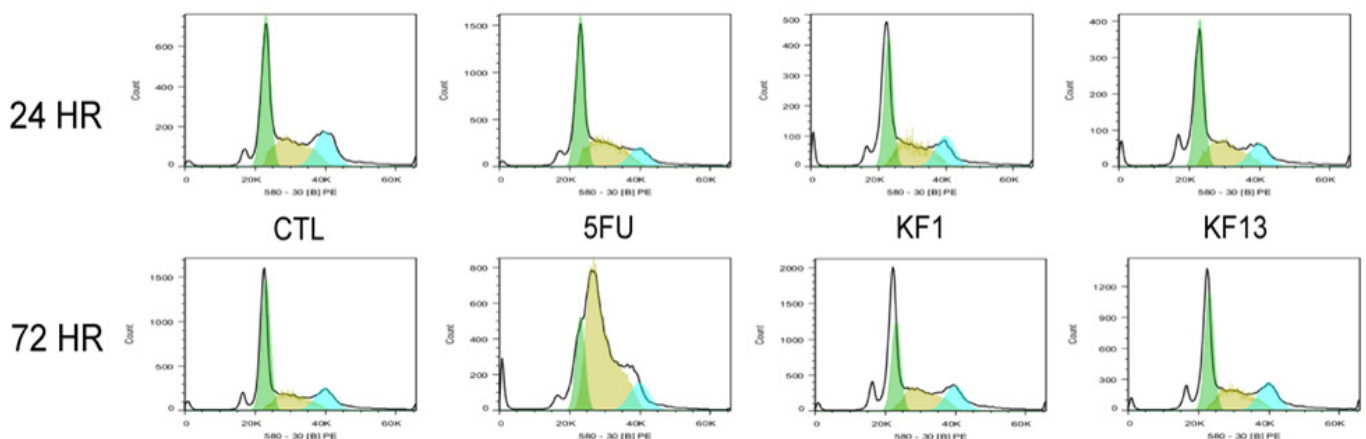
Figure 7.2 │ FACS histogram plots of G1 (green), S (gold) and G2/M (blue) cell counts. The A-498 cells were exposed to agents KF1, KF13 and 5FU at the IC_{50} for a period of 24 and 72 hours.

7.3.3 │ Spectral and sample quality

The control spectra for the 24 hour dataset were analysed in terms of G(1), S and G2/M phase. Figure 7.3 displays the mean spectra directly after RMieS-EMSC correction (without vector normalisation).
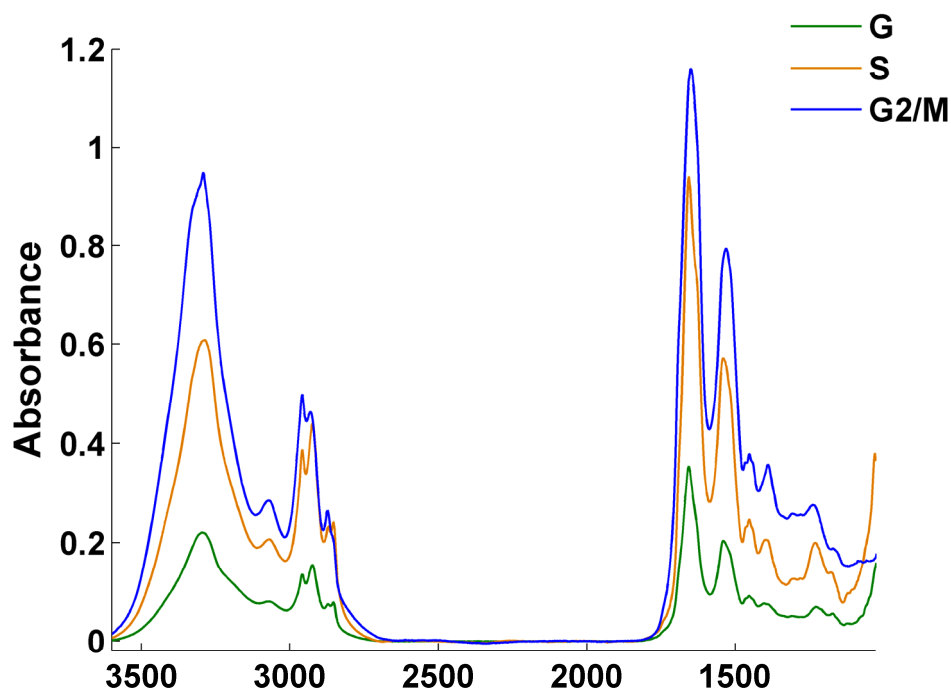
Figure 7.3 │ Mean RMies-EMSC corrected spectra for non-drug treated control cells after 24 hours of cell culture (not normalised).

Overall absorbance appeared to be greater in G2/M cells than in the other cell cycle spectra. This result favourably compared with the research of Holman et al. [9] (Fig. 7.4) who postulated this may due to either more material in the cell, or thickness differences in M phase. In the second case there could be a greater path length for the IR beam to traverse. The v_s(CH$_3$).peak at ~1392 cm^{-1} is noticeably larger in the G2/M phase, an observation again described by Holman et al. [9].

The positive correlations suggest that the PI dye is not seriously affecting the associated chemistry of the cell cycle within the spectral bands of molecules other than those associated with DNA.

Figure 2. IR spectra of individual cells in different stages of the cell cycle. Spectra were not normalized, but a linear baseline was subtracted over the range of 2000 to 650 cm^{-1}.
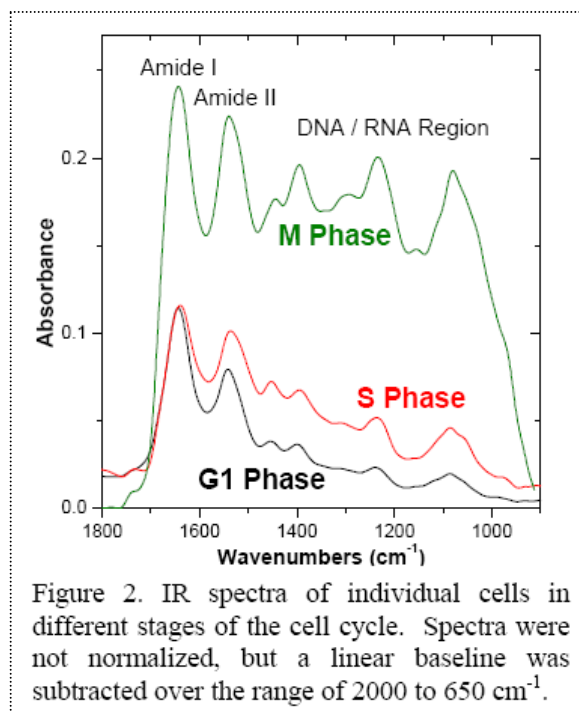
Figure 7.4 │A figure taken from the publication of Holman et al. [9] comparing the mean spectra of cell cycle phase in normal human foetal lung fibroblast cells IMR-90.

The spectra (Fig. 7.5) were typically of poorer quality in the lower wavenumber region (1200-1000 cm^{-1}), with low signal and increased noise, relative to other drug-treated (Chapter 6) or cytospun spectra (Chapter 5).
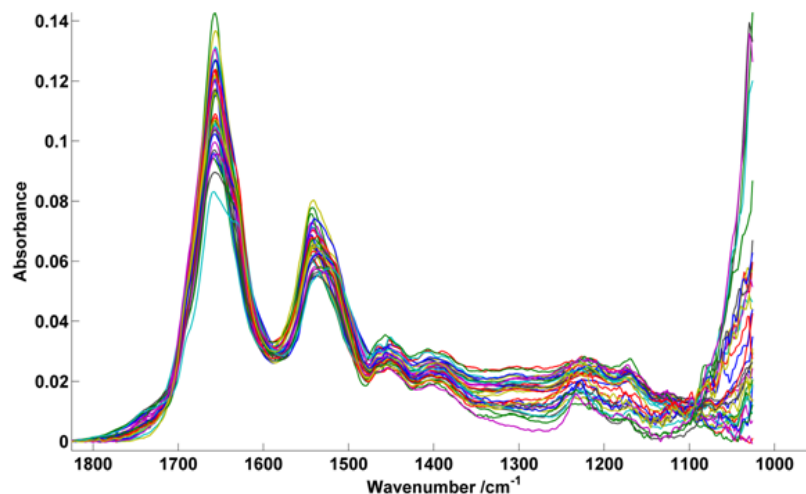
Figure 7.5 │ RMieS corrected dataset for 24 hour control. The spectral range used for analysis had to be reduced to 1110-1825 cm^{-1}.

After discussion with the SMIS beamline team at Soleil, the large absorbance below 1100 cm^{-1} probably arises from a misalignment of the condenser. It has been observed that the condenser not remaining aligned when the lower microscope illumination was left on. It is probable that this is due to the warming of the mirror and a change in the mirror curvature or alignment due to dilatation.

The fingerprint region was therefore decreased to 1110-1825 cm^{-1}, omitting the poor signal to noise areas that contained nucleic that may have otherwise been highly informative.

Figure 7.6a displays of RMieS-EMSC corrected, vector normalised mean spectra per cycle phase displaying their standard deviation. The amide I and II regions displayed notable variance, as well as the region 1200-1400 cm^{-1} spectral region in the G1 phase spectra. It may be that in the cells in G1 phase are more heterogeneous as DNA replication is occurring, perhaps faster in some cells than others, hence the larger variability in the nucleic acid region due to PI dye concentration.

The PCA score plot for the control dataset is shown in Figure 7.6b. G1 phase spectra are anti-correlated against S and G2/M phase spectra in PC1. Distinction

can also be made in PC2 where G2/M phase spectra anti-correlate with G1 and S data.

The cell spectra in different cell cycle phases could not be compared to one another due to the influence of PI. Increased DNA content in the later phases of the cycle, just before mitosis, means that the concentration of the PI dye will be higher in these cells, and therefore the extent of DNA conformational change.
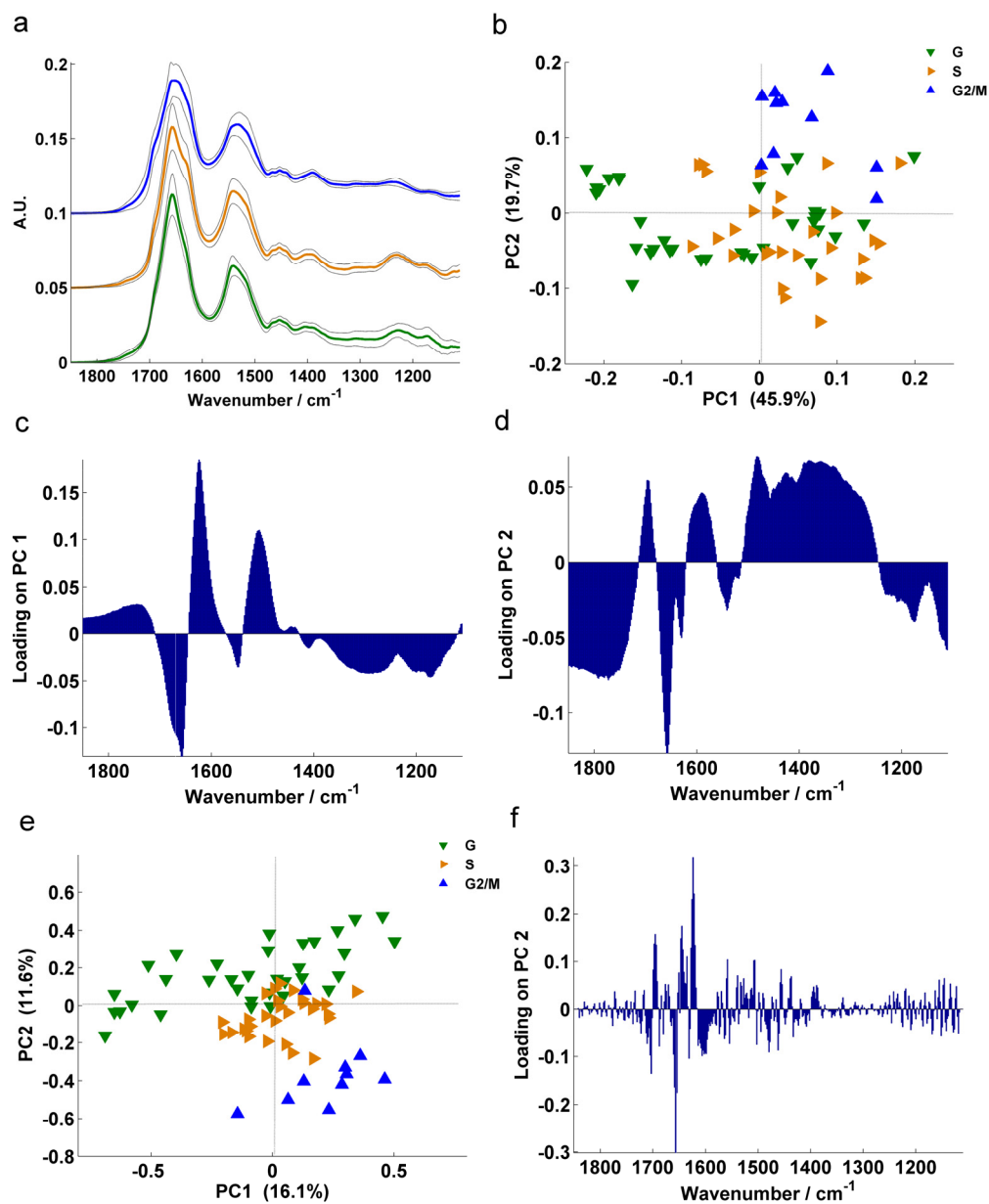
Figure 7.6 │ (a) mean (RMieS-EMSC corrected; vector normalised) fingerprint spectra for G1 (green), S (gold) and G2/M (blue) control cells describing the difference in standard deviation (not to scale). Variance in the spectra can be mainly found at the amide I and II absorbance bands (b) PCA of the dataset and subsequent loading plots (c, d for PC1, PC2). The equivalent 2nd derivative score plot (e) and PC2 loading (f).

Figure 7.7 displays optical images of the FACS-sorted control cells, cytospun onto calcium fluoride substrates. The S and G2/M cells appear darker, due to the extent of PI intercalation within the cell.

Interestingly, the cells in G1 phase appear larger in diameter than the other cells. During the cell cycle process, typically the cells become larger as these progresses through each stage. G2 stage cells should technically be the largest as cells continue to grow through the cell cycle, prior to mitosis [3].
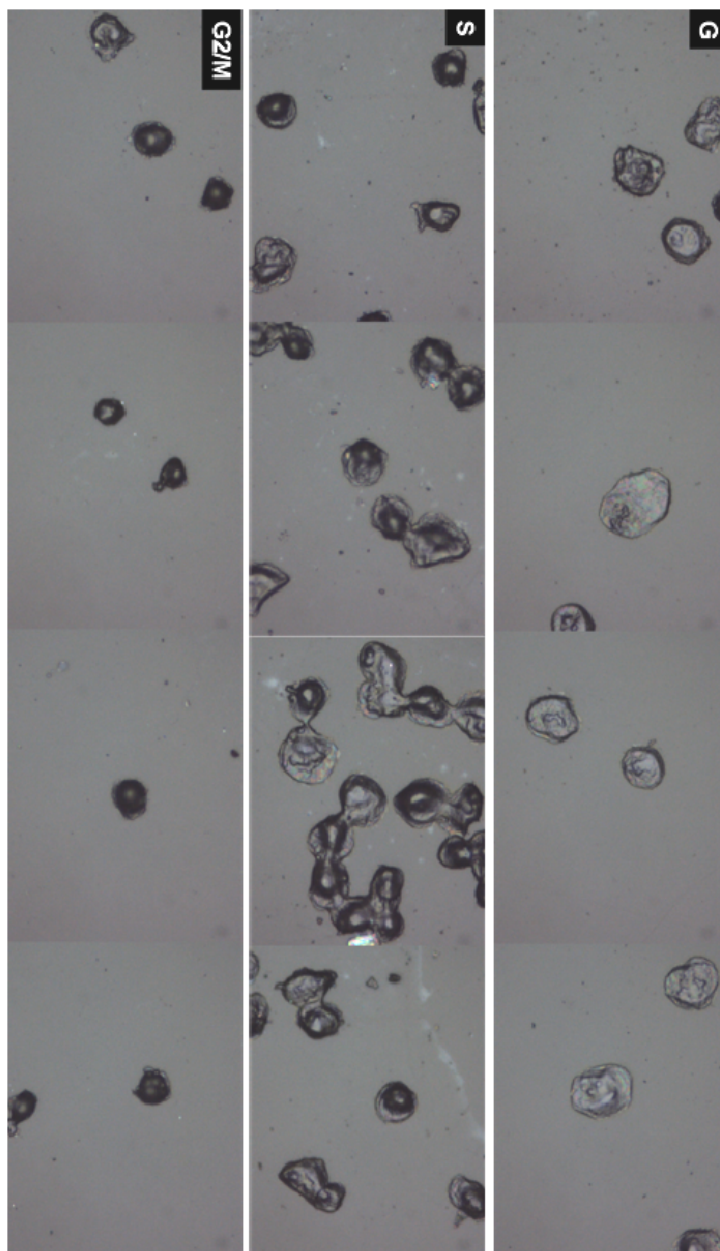
Figure 7.7 | **Optical images of a selection of control cells prepared for spectroscopy measurements, highlighting differences between G(1), S, and G2/M cells.**

The diameters of 50 cells were measured using a ruler tool in the Omnic software for each phase. The group means were 24 μm, 22 μm and 19 μm for cells in G(1), S and G2/M phase respectively and were found to be statistically different (P<0.001) using a one-way analysis of variance (ANOVA) test in Graphpad Prism (Fig. 7.8).
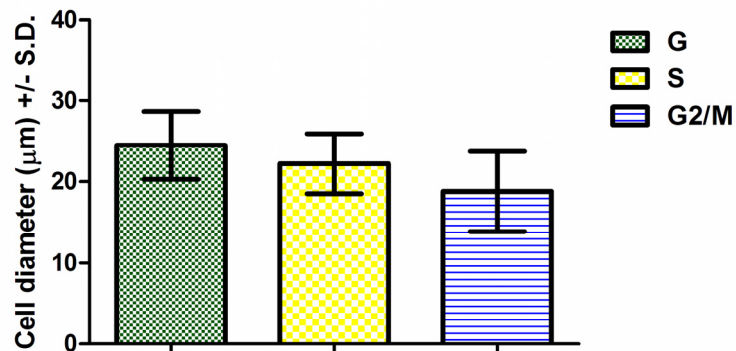
Figure 7.8 │ **Mean diameter of cytospun cells for each cell cycle phase ± standard deviation (S.D).**

It is possible that the cells collected in G2/M phase, which had undergone spectral acquisition, included a cell population richer in M phase cells than G2 cells. In M phase cell division would have occurred. The daughter cells would be smaller in size than cells in the rest of the cell cycle prior to mitosis [3].

It would be impossible, however, to make a definitive assumption about the natural size of the cells prior to processing for spectroscopy measurements. The cells were ethanol fixed prior to FACS sorting. The ethanol fixation may have an effect morphological distortion, such that dehydration may shrink the cells [10]. The difference in PI concentration between cell phases, however, is probably not a factor, as the dye is applied after ethanol fixation.

Research has shown that, for a cell line of different origin, G2/M cells were said to be more rigid than G1 cells [11]. This may be another plausible explanation as to why the G1 cells appeared larger after cyto-spinning. It was possible that the G2/M cells, displaying more rigid properties than the G1 or S cells, remained relatively

spherical when the centrifugal force was applied whereas the G1/S cells flattened slightly more on impact.

It was important to utilise the RMieS correction algorithm on this dataset, particularly because of the size difference. The cells from different phases may have all appeared similar in morphological shape; however, the difference in size would have an effect on the Mie scattering profile. Cells may have been discriminated against because of these different profiles, of which can be observed in the mean raw spectra (with standard deviation) in Figure 7.9.
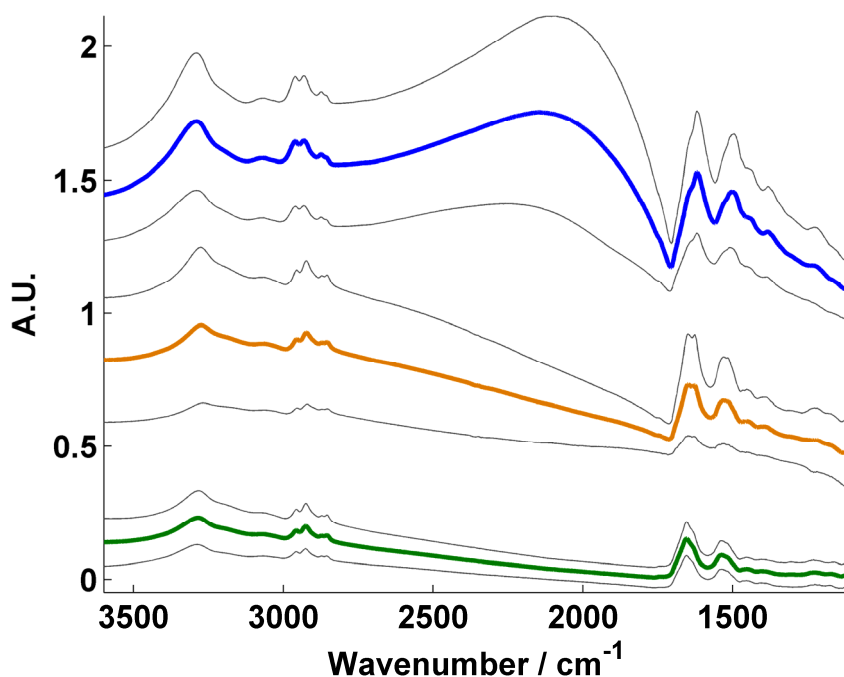
Figure 7.9 │ **The mean cell phase spectra for G1 (green), S (gold) and G2/M (blue), acquired from the non-drug treated control at the 24 hour time point.**

7.3.3 │ Cellular response after 24 hours of drug exposure

In Chapter 6, chemometric distinction between KF1 and 5FU treated Caki-2 cells after 24 hours was found. This distinction was only achieved, however, with the use of spectral derivatives and discriminant function analysis.

As previously discussed, it was not a viable option to compare drug-treated cell spectra from different cell phases due to the difference in DNA conformation from the propidium iodide concentrations.

As previously discussed, during FACS analysis whereby the cells were sorted in preparation for the SR-FTIR experiments, the cell count profiles showed no apparent difference in cell accumulation in each phase after 24 hours of agent addition, in all group cases (Fig 7.2)

Spectral data collection focused on the 72 hour time point samples, however, G1 phase cell spectra for the PI stained 24 hour drug exposure samples were also collected. PCA was performed on the stained, G1 phase SR-FTIR spectra, enabling a comparison of cells treated with different agents in same phase, irrespective of PI concentration.

Cluster separation between cells treated with different agents was achieved (Fig. 7.10). Due to the single-phase analysis where PI concentration and cell cycle chemistry remained a constant, the complexity of the dataset was reduced, such that the use of spectral derivatives and supervised LDA were not necessary to discriminate spectra. The SR-FTIR and chemometric approach found differences in cell chemistry after 24 hours of drug exposure, a result that was not detectable by corresponding FACS analysis. This demonstrates the sensitivity of the infrared technique.
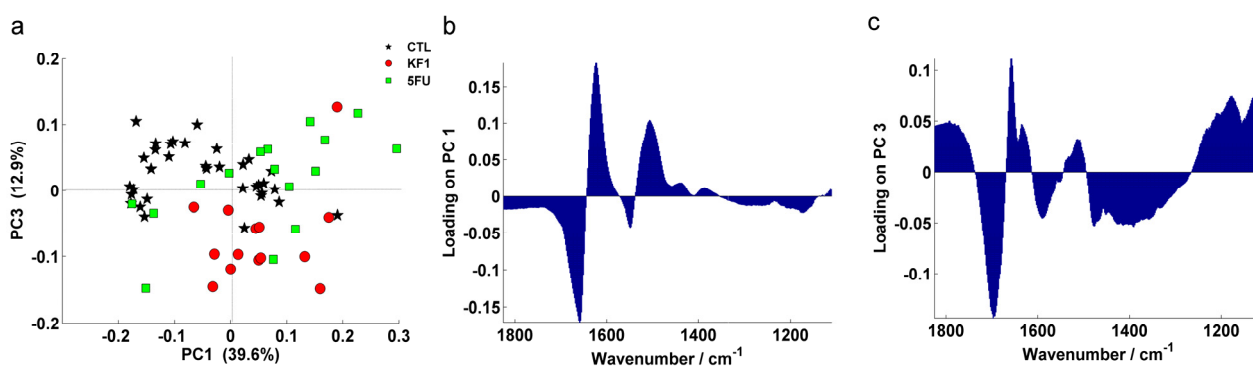
Figure 7.10 │ PCA of cell spectra in G1 phase after 24 hours of drug exposure (a) with PC1 and PC3 loadings (b, c)

Drug treated cell spectra could be discriminated from the control group in PC1. The loading plot in PC1 predominantly highlights protein differences. KF1 also appeared to discriminate from the CTL/5FU groups in PC2. The loadings in PC3 revealed a notable peak at ~1700 cm^{-1}. This peak resides near the protein region of the spectral fingerprint, although according to Dovbeshko et al. [12], it may be associated with nucleic acid damage and repair. Stretching C=O vibrations that are H-bonded (in guanine, thymine) may be connected with destruction of old H-bonds and creation of the new ones. A full assignment [13] of loadings peaks for PC1 and PC3 can be found in Table 7.1, 7.2 (Ch. 9, Appendix).

It may be postulated that the KF1 drug is interacting with the cellular DNA at this time point. Similarly when Caki-2 cells were treated with KF1 and KF13, they appeared to have a greater initial change in cell chemistry than the 5FU after 24 hours (Ch. 6). In some cases the loading plots from chapter 6 displayed similar peaks at ~1700 cm^{-1} however it was difficult to deduce cell cycle contributions from drug effect.

Proliferating cells are generally more susceptible than resting cells to toxic and mutagenic effects of DNA-damaging agents as DNA replication and chromosome segregation processes are easily disrupted by DNA damage [14]. To prevent cells with damaged DNA to continue to S-phase, signal transducers activate p53 and de-activate cyclin-dependent kinases to inhibit cell cycle progression. Cell cycle arrest can occur, which enables extra time for repair enzymes to work on the damaged DNA before cell cycle continuation [14-15].

As described (Ch. 1.2.2) activation and deactivation of signal transduction pathways are initiated by phosphorylation. In PC3, which separates the CTL and 5FU clusters from KF1, there is a notable contribution to the loading around the lower spectral fingerprint region, most notably ~1240 cm^{-1} which is primarily associated with phosphate $v_{as}(PO_2)$ stretching from phosphodiester groups of cellular nucleic acids [12].

7.3.4 │ Cellular response after 72 hours of drug exposure

The analysis of cells after 24 hours of exposure will sample cells that are most likely within the process of cellular damage and repair. Here, cell spectra were acquired for cells treated with KF1 and 5FU (and non-treated control cells) after 72 hours after initial exposure to the drugs. At 72 hours, however, cells likely to be affected by the chemotherapeutic agents would no longer be viable and would not have been part of the cell population. Dead cells and debris would have been removed during the sample processing. During the flow-cytometry forward angle light-scatter gating would have remove small debris. Dead cells prior would also have been gated out due to a reduced DNA fluorescence signal because of DNA breakdown in dead cells prior to chemical fixation [16-17]. Therefore the cell populations studied in the 72 hour dataset included largely viable cells unaffected by the initial drug dose. Cells in the process of repair or damage would also be present however, though perhaps at a reduced frequency than the 24 hour dataset.

The dataset, including all G1, S and G2/M phases was analysed by PCA as a whole dataset (Fig. 7.11). The resulting spread of data points appeared similar to previous experiments in chapter 6, where cell cycle chemistry appeared to inhibit or mask the effects of the different drugs. Similar plots were also seen by Flower et al. in drug-treated ovarian cancer cell lines [1].

However the spectra corresponding to the individual stage were known and were highlighted accordingly in the sub-plots below the main figure. Without the use of derivatives, or further discriminant analysis, it was clear that most of the variation in the main dataset cluster stemmed from the differences in cell chemistry in S phase.
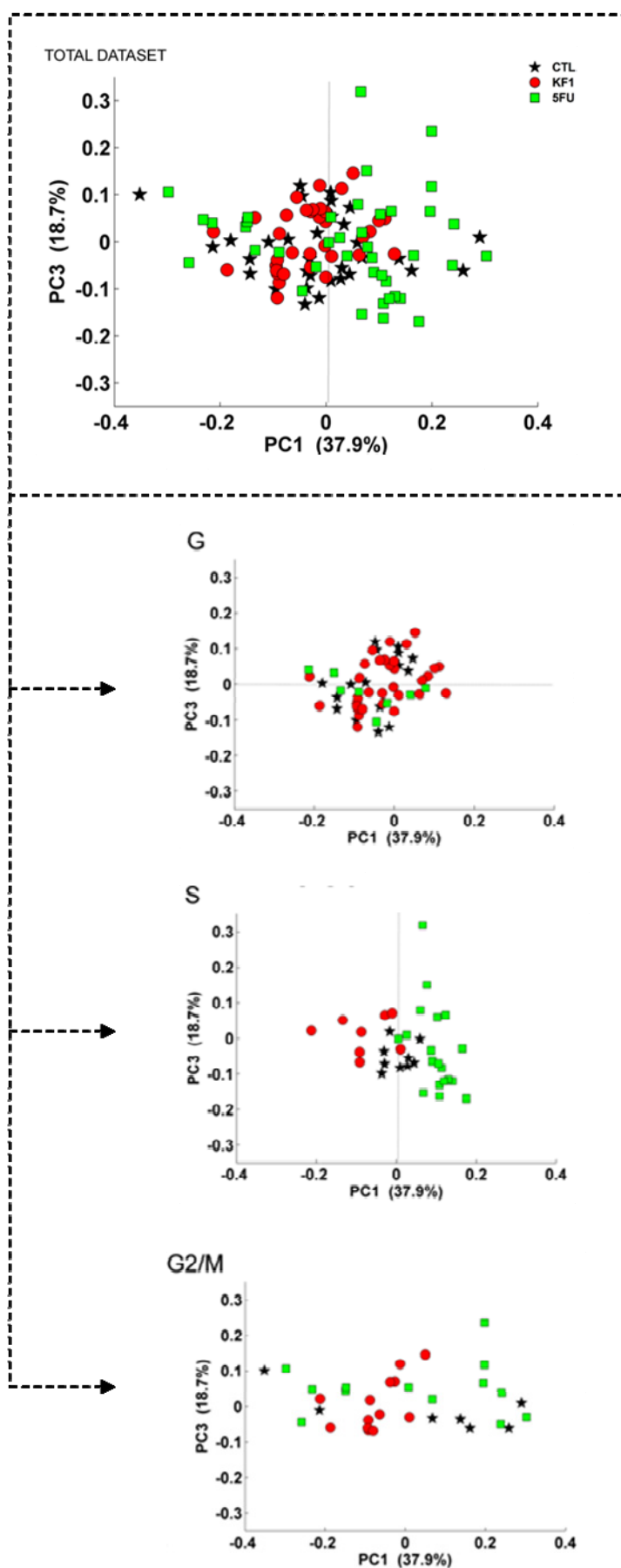
Figure 7.11 │ PCA of total dataset (top) and the respective spectra from each phase that made up the total dataset are shown below where it is apparent that the greatest variance between the drug types occurs in S phase.

Analysis of the complete dataset loadings would not be viable due to the variable extent of DNA conformational change from PI concentration from phase to phase. To validate this result, PCA was repeated using S phase group spectra alone. The same relationship was observed (Fig. 7.12a), suggesting that the cell chemistry in KF1-treated cell spectra is anti-correlated with the 5FU-treated cell spectra. This is again in agreement with the FACS analysis which illustrated a significant change in the cell population profile in S phase after 72 hours of drug exposure. The use of 5FU validates the FACS methodology and therefore it can be said that KF1 does not appear to cause significant cell cycle arrest in the A-498 RCC cell line.
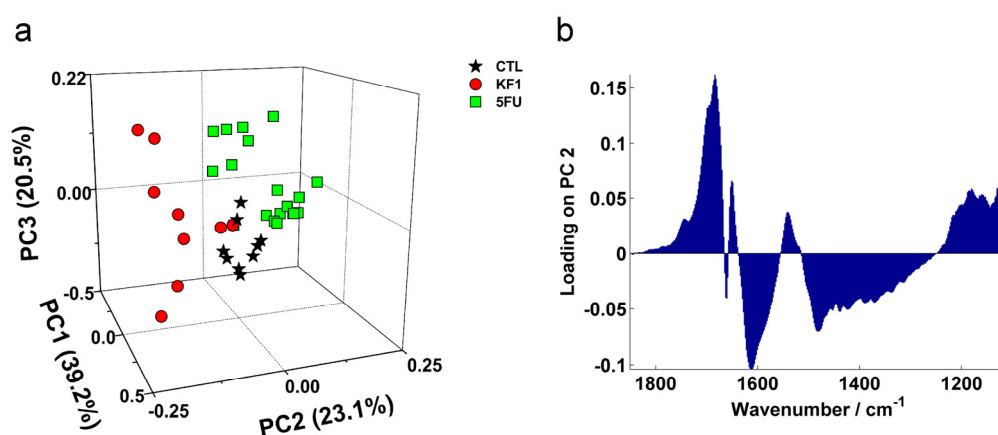
Figure 7.12│ PCA using S phase data only (a) and the PC2 loading plot that highlights the discriminatory chemical differences between the two different drug-treated cell populations

The peak at 1700 cm^{-1} found in the PC loading when G1 phase (Fig. 7.10b) was compared is not present in the S phase PC2 loading plot above (Fig. 7.12b).

A compete assignment [13] is given in Table 7.3, (Ch.9, Appendix). Notable loading peaks concern protein bands, and the $v_s(PO_2)$ symmetric phosphodiester stretching band of RNA at ~1120 cm^{-1}. The protocol for PI staining requires RNAse treatment [6] so that both RNA and DNA are not stained, thus evidence is given of a chemical difference detected that is wholly unaffected by PI staining, therefore validating the use of intra-phase analysis.

As described previously (Ch. 1), 5FU is an analogue of uracil with a fluorine atom at the C-5 position in place of hydrogen. Upon cellular entry it is converted to

several active metabolites: fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP) and fluorouridine triphosphate (FUTP). These active metabolites disrupt RNA synthesis and the action of thymidine synthase (TS) [15].

There is also an intra-S-phase checkpoint that can block replication. It is activated by damage encountered during the S phase, or previously unrepaired damage that has escaped the G1/S checkpoint. The damage sensors for the intra-S checkpoint include a large set of checkpoint and repair proteins [15].

As a final exercise, PC-LDA was performed (Fig 7.13) using 5 PCs and second derivative spectra, a combination used in Chapter 6. In this case, however, there are 9 different groups (G1, S and G2/M for CTL, KF1 and 5FU) and the size of the dataset was too limited for any results to provide 100% confidence at this stage.

It does however enhance the separation of cell-treatment groups in S phase. To a certain extent, it could be argued that there are cellular chemical differences in G2/M phase also. The G2/M checkpoint prevents cells from undergoing mitosis in the presence of DNA damage. It could be postulated, as the cells selected for spectroscopy are viable ones at this 72 hour stage, there may be an indication of repair mechanisms in motion.
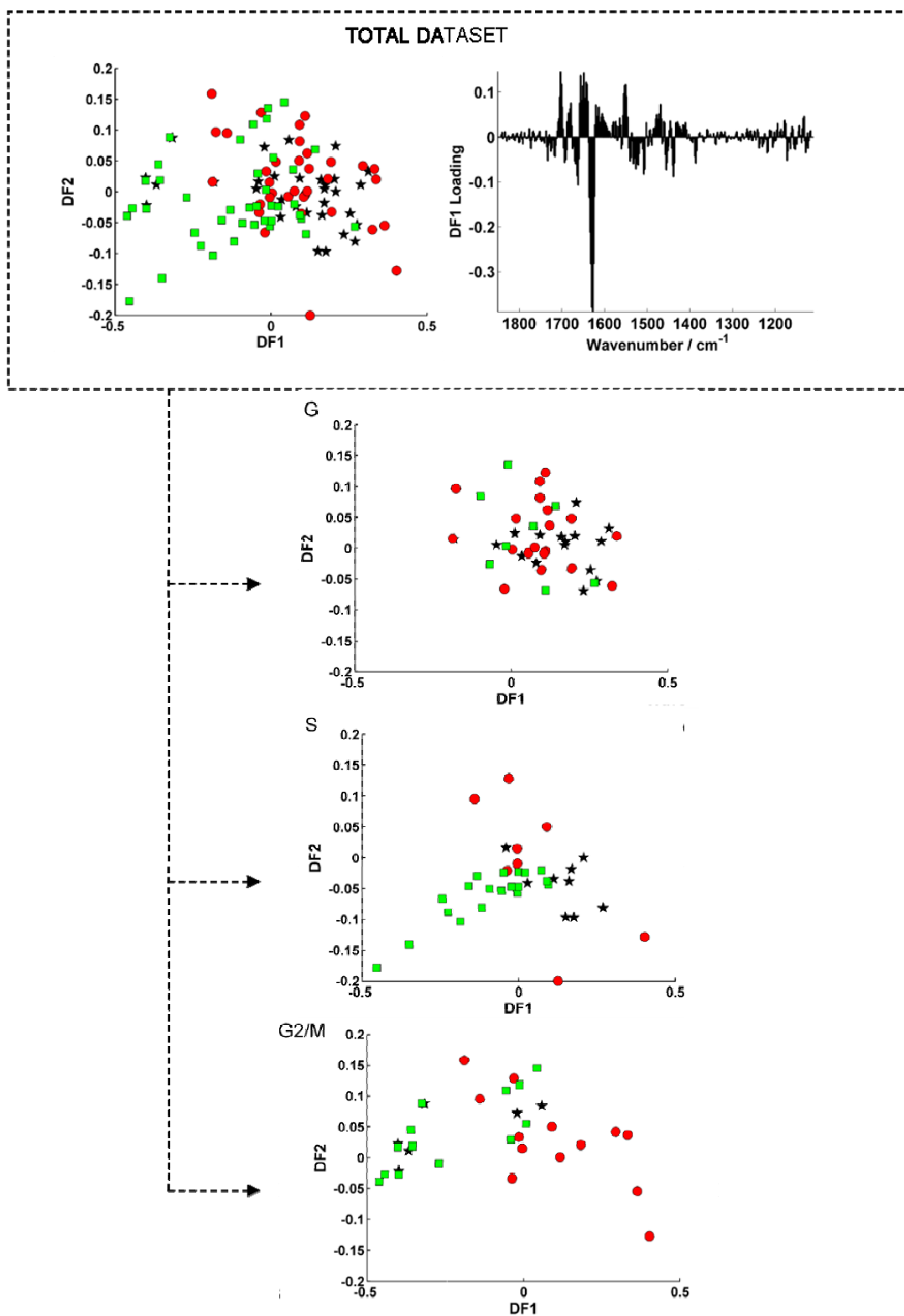
Figure 7.13 │ LDA of the total 2nd derivative dataset using 5PCs. The respective spectra from each phase that made up the total dataset are shown below where it is apparent that the greatest variance between the drug types occurs in S phase and possibly some difference in G2/M phase.

7.4 │ Conclusions and Future Prospects

The use of FACS sorting the cells enabled differences due to drug effect to be analysed alone, which simplified interpretation of the multivariate analysis results. In chapter 6 it appeared that generally, the KF compound series caused a greater change cellular chemistry earlier.

This chemical change may pertain to the fact that the KF drugs are causing DNA cross-linking at this stage, as suggested. It is known that 5FU acts in a very different mode and disrupts progression in the S phase of the cell cycle.

This evidence may help determine the timing of future experiments, depending upon the question to be asked. Is it an initial response that is to be observed, or is it the greatest extent of the response and chemical change, which will be determined by the length of the corresponding inhibitory concentration assay, 50% of cells will not be inhibited. It is these viable cells that are analysed (Ch. 7.3.4), as well as perhaps partially-viable cells.

As seen in chapter 6 with the Caki-2 cell line, a cellular response to the anticancer compounds could be detected after 24 hours of exposure, unlike the other techniques used. This highlights an advantage of the technique, and shows promise that an indication of drug efficacy may be detectable in a short time frame, in the context of high-throughput screening of a multidrug panel. Result at this time point has also raised the question as to what earliest time point an initial response may be detected by FTIR spectroscopy. An ideal way to monitor this would be to use a live cell experiment. Flow cytometry is a progressive field, particularly with the use of synchrotron radiation sources for single cell experiments [18-20]. Water contributions, however, would dominate the signal unless a correction could account for the contributions, or a flow cytometry cell is well designed to limit the effects of water in the spectrum. Alternatively Raman spectroscopy has an advantage over infrared for liquid based experiments, however, there is an added risk of damaging the sample due to the heat generated, which would affect cell viability. Regardless of this, recent research has shown promise [21-23].

Through the use of FACS analysis, sorting single cells by their progression phase in the cell cycle, it was possible to analyse differences largely due to drug effect by analysing one phase at a time. Where cluster patterns and chemical signatures could be interpreted by assessing same-phase spectra, a number of chemical differences agreed with the FACS analysis and the known mode of action of 5FU. The results including tentative band assignments would warrant further confirmation with alternative methods in order to be validated.

The necessary use of propidium iodide staining, however, was not the most ideal situation. As chemical fixation was used prior to staining, the staining should not have had too much of an effect on the cellular chemistry overall. It did, however, affect the detection of changes in nucleic acids; a highly important feature in monitoring cellular response to drugs. Cells analysed across different phases would most likely involve difference in peaks associated with nucleic acids, due to the DNA conformational changes from the intercalation of different concentrations of PI dye.

It is still desirable to separate chemistry based on cell cycle differences from chemistry changes from cellular drug response. Removal or isolation of cell cycle chemistry as a variable makes for more simple and accurate analysis of drug response effects. The next logical step in this line of research, therefore, would be to establish a method to define spectra to respective cell cycle phase without having to stain the cells first, thereby introducing chemical changes in a highly important region of spectrum.

To do this a supervised training algorithm would need to be implemented; using a technique similar to Boydston-White et al. [24]. Cell spectra could be acquired and subsequently stained with 4',6-diamidino-2-phenylindole (DAPI) which is a fluorescent stain that binds strongly to DNA in a stoichiometric reaction. It can be used to quantify cellular DNA content by measuring the integrated intensity of the dye in the nucleus and is used extensively in fluorescence microscopy and for the analysis of both live and fixed cells [25-26]. It is more advantageous to PI which

requires RNAse treatment so that both RNA and DNA are not stained. DAPI does bind to RNA but is not as strongly fluorescent and the use of RNAse is not required. Cell cycle could be deduced by applying the appropriate fluorescence intensity thresholds where intensity = G1 < S < G2/M.

In order to locate individual cells before and after spectral acquisition and staining the use of an England finder graticule would be advantageous, held underneath a reflective slide such as MirrIR. Once a cell was selected for acquisition, the focus could be shifted to the graticule and the position co-ordinate recorded along with the image.

An example of this is shown (Fig. 7.14) using Omnic software at the Soleil synchrotron. Another advantage of this approach would be that cells could be grown directly onto the infrared substrate, removing the need for cyto-spinning and preserving true cellular morphology (as true as can be for cells grown onto a flat, 2D surface).
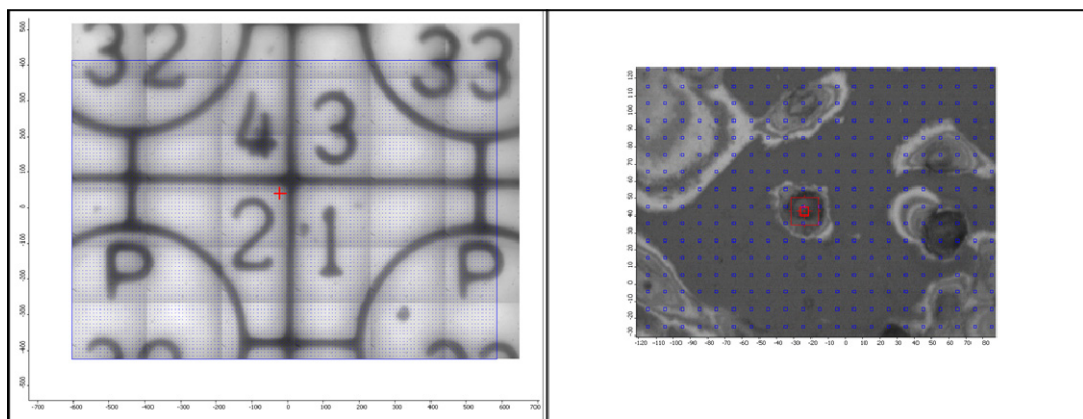
Figure 7.14│ Optical images of the England finder graticule (left) and corresponding region on the sample surface (right). The cross-hair marks the co-ordinate of the cell in both images.

The cultured cells would then be stained for DNA content and defined in cell cycle phase using the fluorescent DAPI dye as illustrated in Figure 7.15. Once a supervised training algorithm was in place, cell spectra may be characterised without the use of staining in the future, which would be advantageous not only to the analysis of cellular response to drugs, but for many other applications.
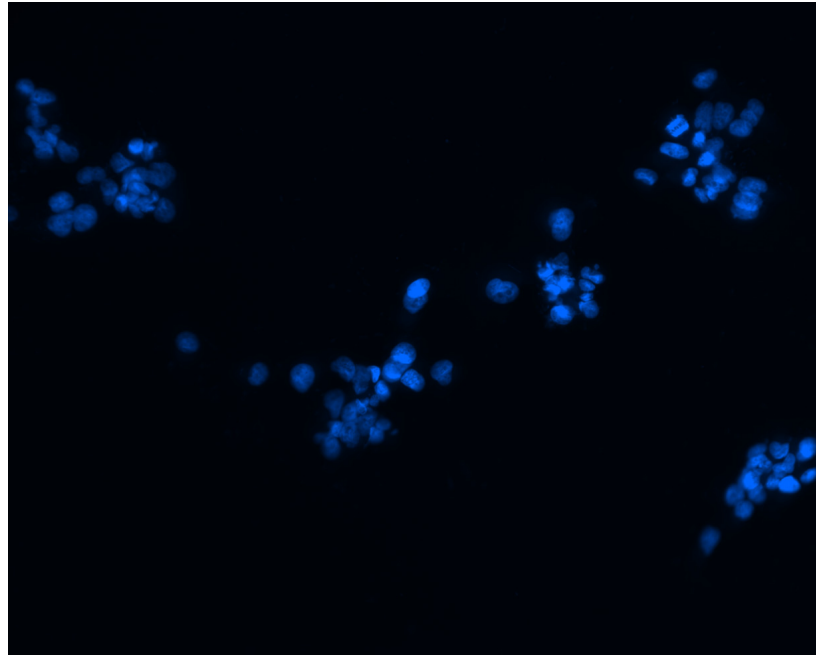
Figure 7.15│ Fluorescence microscopy: example of A-498 cells with their nuclei stained with DAPI (in mountant) taken with an Olympus microscope equipped with a 20× lens for an exposure time of 300 ms.

7.5 │ References

1. Flower KR, Khalifa I, Bassan P, Demoulin D, Jackson E, Lockyer NP, McGown AT, Miles P, Vaccari L, Gardner P: **Synchrotron FTIR analysis of drug treated ovarian A2780 cells: An ability to differentiate cell response to different drugs?** *Analyst* 2011, **136**(3):498-507.

2. Krishan A: **Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining**. *Journal of Cell Biology* 1975, **66**(1):188-193.

3. Elliott WH, Elliott, DC: **Biochemistry and Molecular Biology**, *Oxford University Press* 2001: 495-502.

4. Xu J, You C, Zhou L, Li Q, Zhou P, Chen N: **The cell-cycle kinetics of craniopharyngioma and its clinical significance**. *Journal of Neuro-Oncology*, **98**(1):71-76.

5. Raghavan R, Dickey Jr WT, Margraf LR, White Iii CL, Coimbra C, Hynan LS, Rushing EJ: **Proliferative activity in craniopharyngiomas: Clinicopathological correlations in adults and children**. *Surgical Neurology* 2000, **54**(3):241-248.

6. Warburton HE, Brady M, Vlatkovia N, Linehan WM, Parsons K, Boyd MT: **p53 regulation and function in renal cell carcinoma**. *Cancer Research* 2005, **65**(15):6498-6503.

7. Smith CD, Zilfou JT: **Circumvention of P-glycoprotein-mediated multiple drug resistance by phosphorylation modulators is independent of protein kinases.** *Journal of Biological Chemistry* 1995, 270(47)24,28145–28152

8. Longley DB, Harkin DP, Johnston PG: **5-Fluorouracil: Mechanisms of action and clinical strategies.** *Nature Reviews Cancer* 2003, **3**(5):330-338.

9. Holman HYN, Martin MC, Blakely EA, Bjornstad K, McKinney WR: **IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based Fourier transform IR spectromicroscopy**. *Biopolymers-Biospectroscopy Section* 2000, **57**(6):329-335

10. Poulin NM, Matthews JB, Skov KA, Palcic B: **Effects of fixation method on image cytometric measurement of DNA content and distribution in cells stained for fluorescence with propidium iodide**. *Journal of Histochemistry and Cytochemistry* 1994, **42**(8):1149-1156.

11. Tsai MA, Waugh RE, Keng PC: **Cell cycle-dependence of HL-60 cell deformability.** *Biophysical Journal* 1996, **70**(4):2023-2029.

12. Dovbeshko GI, Gridina NY, Kruglova EB, Pashchuk OP: **FTIR spectroscopy studies of nucleic acid damage.** *Talanta* 2000, **53**(1):233-246.

13. Movasaghi Z, Rehman S, Rehman IU: **Fourier transform infrared (FTIR) spectroscopy of biological tissues**. *Applied Spectroscopy Reviews* 2008, **43**(2):134-179.

14. Ljungman M: **The DNA damage response-Repair or despair?** *Environmental and Molecular Mutagenesis* 2010, **51**(8-9):879-889.

15. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S: **Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.** In: *Annual Review of Biochemistry*. vol. **73**; 2004: 39-85.

16. Kressel M, Groscurth P: **Distinction of apoptotic and necrotic cell death by in situ labelling of fragmented DNA**. *Cell and Tissue Research* 1994, **278**(3):549-556.

17. Muirhead KA, Kloszewski ED, Antell LA, Griswold DE: **Identification of live cells for flow cytometric analysis of lymphoid subset proliferation in low viability populations**. *Journal of Immunological Methods* 1985, **77**(1):77-86.

18. Nasse MJ, Ratti S, Giordano M, Hirschmugl CJ: **Demountable liquid/flow cell for in vivo infrared microspectroscopy of biological specimens.** *Applied Spectroscopy* 2009, **63**(10):1181-1186.

19. Tobin MJ, Puskar L, Barber RL, Harvey EC, Heraud P, Wood BR, Bambery KR, Dillon CT, Munro KL: **FTIR spectroscopy of single live cells in aqueous media by synchrotron IR microscopy using microfabricated sample holders.** *Vibrational Spectroscopy*, **53**(1):34-38.

20. Oshima Y, Takenaka T, Sato H, Furihata C: **In situ cell cycle phase determination using Raman spectroscopy.** In: *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*; San Francisco, CA.

21. Snook RD, Harvey TJ, Correia Faria E, Gardner P: **Raman tweezers and their application to the study of singly trapped eukaryotic cells.** *Integrative Biology* 2009, **1**(1):43-52.

22. Harvey TJ, Hughes C, Ward AD, Faria EC, Henderson A, Clarke NW, Brown MD, Snook RD, Gardner P: **Classification of fixed urological cells using Raman tweezers.** *Journal of Biophotonics* 2009, **2**(1-2):47-69.

23. Harvey TJ, Faria EC, Henderson A, Gazi E, Ward AD, Clarke NW, Brown MD, Snook RD, Gardner P: **Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers**. *Journal of Biomedical Optics* 2008, **13**(6).

24. Boydston-White S, Romeo M, Chernenko T, Regina A, Miljkovic M, Diem M: **Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: Principal component and artificial neural network analysis.** *Biochimica et Biophysica Acta - Biomembranes* 2006, **1758**(7):908-914.

25. Lin M, Alfi O: **Differential fluorescent staining of sister chromatids and the study of sister chromatid exchanges with 4,6 diamidino 2 phenylindole.** *Fifth International Congress of Human Genetics* 1976, No. **397**.

26. Kapuscinski J: **DAPI: A DNA-specific fluorescent probe**. *Biotechnic and Histochemistry* 1995, **70**(5):220-233.

Chapter 8 │ Project Conclusions and Future Potential

To lay the foundation work in order to investigate the feasibility of FTIR spectroscopy as a tool for cellular response to drugs, three levels of progression were involved.

8.1│Validation of the RMieS-EMSC algorithm for spectral interpretation using cell line characterisation of renal carcinoma cell lines

The novel correction RMieS-EMSC algorithm to correct FTIR spectra of biological material is a welcomed improvement in spectral pre-processing. The algorithm has been shown to correct simulated spectra with greatly improved performance over older algorithms, such as EMSC alone. While using simulated spectra can result in a definite evaluation of the algorithm's performance, it is more difficult to comment on the performance with real biological spectra, as the 'true' corrected spectrum answer cannot easily be known.

If the spectroscopy community is to rely heavily on such an algorithm in order to confidently interpret biological spectra, it was important to validate the RMieS-EMSC corrected spectra with other scientific techniques that did not suffer from such optical distortions.

In addition, validation with more clinically familiar -omic techniques was also explored in order to perhaps reach a wider audience, such as members of the clinical community that are unfamiliar with infrared spectroscopy.

Interpretation from the multivariate analysis of the corrected FTIR spectra highlighted key differences in the protein composition of the RCC cell line 2247R, relative to the 2245R and 2246R cell line. For validation, label-free LCMS analysis of the cell lysates revealed that the 2247R protein chemistry was more anomalous to that of the other two cells lines. In particular, alpha-rich sub-units of Histone H2A were revealed to be highly expressed in cell line 2247R, which also appeared to be

rich in alpha helical protein secondary structure in the FTIR analysis. The over-expression of Histone H2A was in turn validated by Western blot for confirmation.

8.2│Characterisation of renal carcinoma cell lines and cells displaying 'stem-cell like' properties

In the eventual analysis of the changes in cellular chemistry to different drugs, spectroscopic and chemometric discrimination would involve tracking discrete changes in one cell line, not simply discriminating cell lines of different origin.

The low proportion of cells displaying stem-cell like characteristics as a sub-population provided yet another model of cell characterisation with increased difficulty. A project running parallel to the investigation looking at targeted side population cancer therapy meant that these cells were readily available for spectroscopic analysis. They were typically small and required the use of an infrared source of high brilliance (synchrotron radiation) and use of the RMieS-EMSC correction algorithm in order to fully characterise.

Not only could the general side population cells be characterised by a different spectral infrared signature to the general, non-side population but the two sub-side populations were also discriminated. This was an encouraging result and allows potential to follow up the parallel, targeted therapy experiment with FTIR and chemometric analysis to determine drug-side population response in the future.

8.3│Exploring methods of detecting cellular response to chemotherapeutics

In the model, it was recognised that 5FU is not very efficient in treating RCC in clinical practice; it provided a good basis as a model agents as it is an established drug with known mode of action, including cell cycle arrest.

For the gold-based KF series of novel drugs, it was not known whether they would perform. At the end of the initial investigation, however, it was deemed that while the KF drugs did have an initial effect, where signs of possible DNA damage and repair were observed, subsequent analysis after 3 days showed disappointing results. Viability assays showed that proliferation had recovered.

In single point monolayer analysis using PCA, the chemical profiles of the infrared spectra of the cells seemed no different to control, suggesting that the cells had not developed resistance or perhaps the drugs were simply no longer present due to cell efflux, or breakdown.

This also highlights the question as to which day of exposure to drugs should cell spectra be analysed? It may be hard to predict response from day 1 as some agents may take more time to act, particularly if they cause cell cycle arrest. The results have also sparked another question concerning dosage: as it is difficult to translate drug dosages in vitro to in vivo, it is better to look at a single dose at the IC_{50}, or multiple doses at multiple concentrations?

PCA works well enough for cell monolayers looking at general trends of a population of cells and is preferable over supervised chemometric techniques that can include levels of subjectivity.

The issue of sample heterogeneity due to cell cycle differences was amplified when analysing single cell spectra. Efforts must be made to separate the infrared information of drug response from differences simply due to inherent cell cycle chemistry if the technique is to be further developed. It was shown in chapter 7 that advanced techniques of discrimination such as LDA may not be required if cell cycle can be omitted factor of discrimination.

In the context of the experiments (Ch.6 and Ch.7.) the FTIR methodology has been successful in indicating a cellular response to the novel gold compounds, which suggests that they are certainly active and able to infiltrate into the cells, particularly in the first 24 hours. While the overall efficacy of the compounds against RCC is

poor, the result is powerful enough to suggest that the compounds may be more effective in other types of cancer cell. Using multivariate analysis it was also possible to distinguish between the gold compounds from 5FU, which has a distinct mode of action that affects cells in the S phase in particular.

In conclusion, use of FTIR spectroscopy appears a feasible technique to monitor drug effects in cells. While the full extent of this method is not yet validated and there are many unanswered questions to account for, the conclusions gained from the spectral analyses were in line with complimentary methods such as the FACS and viability assays. Once further development of this approach to early assessment of drug efficacy can be achieved, it may attract pharmaceutical interest.

Chapter 9 │ Appendix

Table 4.1 │ Calculating the alpha/beta ratio of protein secondary structure

Peak Type	Cell Line	Centre (cm-1)	Area	Height (H)	Structural Assignment	% a (H) % b (H)	a/b Ratio (H)
Voigt	2245R	1640.354	50.739	1.662	Beta	41.7 58.3	0.715
		1658.164	53.193	1.797	Alpha		
		1675.821	21.525	0.852	Beta		
	2246R	1636.517	31.9823	1.3782	Alpha	44.5 55.5	0.803
		1652.082	51.675	2.0597	Beta		
		1669.44	37.3563	1.1873	Alpha		
	2247R	1636.858	39.2453	1.5085	Beta	56.4 43.6	1.293
		1655.329	104.9906	3.5259	Alpha		
		1675.862	33.3777	1.2182	Beta		

Table 4.2 │ MS peptide feature abundance and best peptide match

#	Mass to Charge (M/z)	2245R Mean Abundance (Normalized)	2246R Mean Abundance (Normalized)	2247R Mean Abundance (Normalized	Peptide Match	Score	Associated Protein
1	421.7615063	7.9171E6	9.5045E6	1.58777E7	IPAMTIAK	30.57	CH60_HUMAN
2	425.7688569	1.89476E6	1.95899E6	6.41181E6	HLQLAIR	26.44	B2R5B3_HUMAN
3	472.7726884	4.22063E6	3.82139E6	1.17995E7	AGLQFPVGR	67.93	B2R5B3_HUMAN
4	477.3090516	3.20094E6	2.70028E6	8.64672E6	LLLPGELAK	52.21	B2R5L4_HUMAN
5	491.0108808	4216.56958	640.82403	89.79852	No ID.	No ID.	n/a
6	495.2964495	1.67021E6	1.97688E6	6.42973E6	VFLENVIR	59.15	B2R4R0_HUMAN
7	516.8015562	1.26253E6	1.44931E6	6.75539E6	YRPGTVALR	23.24	B2R4P9_HUMAN
8	581.3171229	8.80499E6	8.53604E6	3.51485E6	EITALAPSTMK	49.37	Q96HG5_HUMAN
9	590.8183468	2.50839E6	2.95895E6	9.15641E6	ISGLIYEETR	64.75	B2R4R0_HUMAN
10	599.8602551	8.67073E6	8.5968E6	4.35356E6	AVFPSIVGRPR	67.83	Q96HG5_HUMAN
11	644.3991758	3.81624E6	3.82042E6	1.18767E7	VTIAQGGVLPNIQAVLLPK	72.66	B2R5B3_HUMAN
12	652.0317232	2.45875E7	2.57645E7	1.12324E7	VAPEEHPVLLTQAPLNPK	43.9	Q562M3_HUMAN
13	654.3133903	1.10724E7	1.17317E7	4.77769E6	YPIEHGIITNWDDMEK	29.39	ACTS_HUMAN
14	663.8881318	1.74117E6	1.90685E6	6.63861E6	KASGPPVSELITK	52.96	A5K4I2_HUMAN
15	706.4040474	1.65411E7	1.28018E7	4.82296E6	GALQNIIPASTGAAK	58.8	G3P_HUMAN
16	730.9198794	9.25962E6	1.31778E7	5.21833E6	No ID.	No ID.	n/a
17	760.3892537	8.58686E6	9.0466E6	1.9198E6	EKQPVAGSEGAQYR	17.31	TES_HUMAN
18	783.3668362	1.2907E7	1.12343E7	5.11403E6	No ID.	No ID.	n/a
19	820.4762747	4.77423E6	8.12601E6	2.97849E6	KVPQVSTPTLVEVSR	96	A8NBZ8_HUMAN
20	822.4382957	2.64375E6	6.21726E6	2.03178E6	TATESFASDPILYRPVAVALDTK	41.41	KPYM_HUMAN
21	839.1373486	4.68637E6	4.5216E6	366061.23553	VETGVLKPGMVVTFAPVNVTTEVK	56.16	EF1A1_HUMAN
22	882.4110866	7.11337E6	4.73826E6	1.45213E6	LISWYDNEFGYSNR	76.53	G3P_HUMAN
23	895.9556043	2.25458E7	2.34813E7	1.0416E7	SYELPDGQVITIGNER	95.56	Q96HG5_HUMAN
24	909.0303409	28407.66929	22483.37969	76445.58341	VAEQTPLSALYLASLIK	47.28	AL1B1_HUMAN
25	966.0942524	1.39806E7	1.44152E7	4.59418E7	VTIAQGGVLPNIQAVLLPK	87.07	B2R5B3_HUMAN
26	977.5430526	1.38088E7	1.50737E7	5.86319E6	VAPEEHPVLLTEAPLNPK	67.82	Q96HG5_HUMAN
27	980.9668273	6.51371E6	7.17552E6	2.63485E6	YPIEHGIITNWDDMEK	47.97	ACTS_HUMAN

Table 4.3 │ MSMS protein identification and secondary structure estimation

#	Associated Protein (MS Data)	Protein Expression Up (≥X2 fold)	Max Fold	Anova-P	Description	% a	% b	% other	Structure Search Type
1	CH60_HUMAN	47R	2.66145681	1.46695E-09	60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2	48.8751155	17.45361883	33.6712558 7	Primary
2	B2R5B3_HUMAN	47R	3.182153103	2.10227E-10	cDNA, FLJ92406, highly similar to Homo sapiens histone 1, H2ag (HIST1H2AG), mRNA OS=Homo sapiens PE=2 SV=1	61.21390586	5.317960047	33.4681140 9	Tertiary
3	B2R5B3_HUMAN	47R	3.182153103	2.10227E-10	cDNA, FLJ92406, highly similar to Homo sapiens histone 1, H2ag (HIST1H2AG), mRNA OS=Homo sapiens PE=2 SV=1	61.21390586	5.317960047	33.4681140 9	Tertiary
4	B2R5L4_HUMAN	47R	3.087437393	1.44511E-11	cDNA, FLJ92519, Homo sapiens histone 1, H2bn (HIST1H2BN), mRNA (Histone 1, H2bn, isoform CRA_a) OS=Homo sapiens GN=HIST1H2BN PE=2 SV=1	49.55238095	5.432539683	45.0285714 3	Quaternary
5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
6	B2R4R0_HUMAN	47R	3.853757517	2.35123E-12	Putative uncharacterized protein OS=Homo sapiens GN=HIST1H4A PE=4 SV=1	51.6184466	8.609708738	39.7834951 5	Quaternary
7	B2R4P9_HUMAN	47R	4.317624575	1.16227E-08	cDNA, FLJ92171, Homo sapiens H3 histone, family 3A (H3F3A), mRNA (H3 histone, family 3A, isoform CRA_a) (H3 histone, family 3B (H3.3B), mRNA) OS=Homo sapiens PE=2 SV=1	42.29336235	6.024264706	51.6933823 5	Quaternary
8	Q96HG5_HUMAN	45/46R	2.387189693	6.27374E-11	Actin, beta (Fragment) OS=Homo sapiens PE=2 SV=1	34.23848878	24.84061393	40.9208972 8	Secondary
9	B2R4R0_HUMAN	47R	3.853757517	2.35123E-12	Putative uncharacterized protein OS=Homo sapiens GN=HIST1H4A PE=4 SV=1	51.6184466	8.609708738	39.7834951 5	Quaternary
10	Q96HG5_HUMAN	45/46R	2.387189693	6.27374E-11	Actin, beta (Fragment) OS=Homo sapiens PE=2 SV=1	34.23848878	24.84061393	40.9208972 8	Secondary
11	B2R5B3_HUMAN	47R	3.182153103	2.10227E-10	cDNA, FLJ92406, highly similar to Homo sapiens histone 1, H2ag (HIST1H2AG), mRNA OS=Homo sapiens PE=2 SV=1	61.21390586	5.317960047	33.4681140 9	Tertiary
12	Q562M3_HUMAN	45/46R	2.613167058	5.05921E-06	Actin-like protein (Fragment) OS=Homo sapiens GN=ACT PE=2 SV=1	29.83396056	26.98543689	43.1912621 4	Quaternary
13	ACTS_HUMAN	45/46R	2.34749351	2.62568E-13	Actin, alpha skeletal muscle OS=Homo sapiens GN=ACTA1 PE=1 SV=1	40.29662839	13.90222932	45.8011423	Primary
14	A8K4I2_HUMAN	45/46R	2.613167058	5.05921E-06	cDNA FLJ76039, highly similar to Homo sapiens histone 1, H1c (HIST1H1C), mRNA (Histone 1, H1c) OS=Homo sapiens GN=HIST1H1C PE=2 SV=1	22.48169014	6.928169014	70.5760563 4	Quaternary
15	G3P_HUMAN	45/46R	3.792046395	9.82547E-14	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3	29.95715726	26.67842742	43.3644163 2	Primary
16	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
17	TES_HUMAN	45/46R	2.554313906	5.52873E-09	Testin OS=Homo sapiens GN=TES PE=1 SV=1	13.62359551	39.18539326	47.1910112 4	Primary
18	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
19	A8NBZ3_HUMAN	45/46R	2.845971277	6.55724E-09	Uncharacterized protein ALB OS=Homo sapiens GN=ALB PE=4 SV=2	71.52217483	0	28.4778251 7	Tertiary
20	KPYM_HUMAN	45/46R	2.624504219	6.42875E-12	Pyruvate kinase isozymes M1/M2 OS=Homo sapiens GN=PKM2 PE=1 SV=4	39.27658738	18.79054453	41.9326681	Primary
21	EF1A1_HUMAN	45/46R	1.817973533	5.29076E-08	Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1	25.09691775	31.49532035	43.4112554 1	Quaternary
22	G3P_HUMAN	45/46R	3.792046395	9.82547E-14	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3	29.95715726	26.67842742	43.3644163 2	Primary
23	Q96HG5_HUMAN	45/46R	2.387189693	6.27374E-11	Actin, beta (Fragment) OS=Homo sapiens PE=2 SV=1	34.23848878	24.84061393	40.9208972 8	Secondary
24	AL1B1_HUMAN	47R	3.270953016	6.39062E-07	Aldehyde dehydrogenase X, mitochondrial OS=Homo sapiens GN=ALDH1B1 PE=2 SV=2	37.83694391	17.06806511	45.1137330 8	Quaternary
25	B2R5B3_HUMAN	47R	3.182153103	2.10227E-10	cDNA, FLJ92406, highly similar to Homo sapiens histone 1, H2ag (HIST1H2AG), mRNA OS=Homo sapiens PE=2 SV=1	61.21390586	5.317960047	33.4681140 9	Tertiary
26	Q96HG5_HUMAN	45/46R	2.387189693	6.27374E-11	Actin, beta (Fragment) OS=Homo sapiens PE=2 SV=1	34.23848878	24.84061393	40.9208972 8	Secondary
27	ACTS_HUMAN	45/46R	2.34749351	2.62568E-13	Actin, alpha skeletal muscle OS=Homo sapiens GN=ACTA1 PE=1 SV=1	40.29662839	13.90222932	45.8011423	Primary

Table 4.4 │ MSMS protein identification and secondary structure estimation

Cell Line	x							N	Mean	Std. Dev.	α/β Ratio
2245R α↑	34.23848878	29.95715726	13.62359551					3	25.93974718	10.87879332	0.858
2245R β↑	24.84061393	26.67842742	39.18539326					3	30.23481154	7.805707814	
2245R α↓	61.21390586	48.8751155						2	55.04451068	8.724842335	4.834
2245R β↓	5.317980047	17.45361863						2	11.38679934	8.581192336	
2246R α↑	34.23848878	39.27658738	13.62359551	29.83398058	71.52217483			5	37.69896542	21.21293654	1.717
2246R β↑	24.84061393	18.79054453	39.18539326	26.98543689	0			5	21.96039772	14.34010262	
2246R α↓	61.21390586							1	61.21390586	0	11.511
2246R β↓	5.317980047							1	5.317980047	0	
2247R α→	61.21390586	51.6184466	42.29338235	49.55238095	48.8751155			5	50.71064625	6.831977149	5.919
2247R β→	5.317980047	8.609708738	6.024264706	5.432539683	17.45361863			5	8.56762236	5.143449418	
2247R α↓	40.29662839	37.83694391	34.23848878	29.95715726	25.09891775	29.83398058	13.62359551	7	30.12653031	8.921206878	1.171
2247R β↓	39.18539326	26.98543689	24.84061393	26.67842742	13.90222932	13.90222932	13.62359551	7	25.73664375	8.490088561	

Table 7.1 │ PC1 Loading plot analysis of G1 phase cell spectra after 24 hours of exposure

PC1 separates CTL against KF1, 5FU

(+) PC1 Wavenumber	Association	(-) PC1 Wavenumber	Association
1620	Nucleic acids (base carbonyl stretching and ring breathing mode)	1660	Amide I band of proteins, in-plane stretching of the C=O bond
1510	In-plane C-H bending from phenyl rings	1550	Amide II of proteins N-H bending and C-N stretching
1430	Lipid $\delta_s(CH)_2$	1240:1350	Phosphate $\nu_{as}PO_2$ stretching bands from phosphodiester groups of cellular nucleic acids. Amide III, C-N stretching and N-H in-plane bending.
1390	Lipid $\delta_s(CH)_3$	1170	Phosphate $\nu_{as}PO_2$ stretching bands from phosphodiester groups of cellular nucleic acids. Amide III, C-N stretching and N-H in-plane bending.
1110	Polysaccharides, $\nu(CO)$, $\nu(CC)$ ring (carbohydrates)		

Table 7.2 │ PC3 Loading plot analysis of G1 phase cell spectra after 24 hours exposure

PC3 separates KF1 against CTL, 5FU

(+) PC3 Wavenumber	Association	(-) PC3 Wavenumber	Association
1750	Ester C=O stretching vibration (phospholipids)	1700	Guanine, thymine stretching C=O vibrations that are H-bonded (changes in the C=O stretching vibrations could be connected with destruction of old H-bonds and creation of the new ones)
1660	Amide I band of proteins in plane stretching of the C=O bond	1590	Ring C-C stretch of phenyl, C=N, NH2 adenine
1510	Peak of nucleic acids due to the base carbonyl stretching and ring breathing mode	1480:1270	Extremely weak peaks of DNA and RNA. Mainly due to the C-H vibrational modes of methyl and methylene groups of proteins and lipids and amide groups
1390	In-plane C-H bending from phenyl rings		
1269:1110	Phosphate $v_{as}PO_2$ stretching bands from phosphodiester groups of cellular nucleic acids Amide III, C-N stretching and N-H in-plane bending		
1170	v_{as} CO-O-C, C-O bands from glyco-materials (carbohydrates) and proteins		
1110	Polysaccharides, v(CO), v(CC) ring (carbohydrates)		

Table 7.3 │ PC1 Loading plot analysis of S phase cell spectra after 72 hours exposure

PC1 separates KF1 against 5FU

(+) PC2 Wavenumber	Association	(-) PC1 Wavenumber	Association
1750	ν(C=C) lipids, fatty acids	1660	Amide I band ν(C=C) cis, lipids, fatty acids
1680	Unordered random coils and turns of amide I	1610	Adenine vibration in DNA
1650	Amide I absorption (predominantly the C=O stretching vibration of the amide C=O)	1490	C=C, In-plane CH bending vibration, deformation in C-H
1540	Protein amide II absorption-predominately b-sheet of amide II		
1180	Amide III band region		
1120	Symmetric phosphodiester stretching band of RNA		