Lipidomic and Metabolomic Analysis of Biological Response Mechanisms in Cancer Cells: A Multidisciplinary Approach

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

2016

Joanna L. Denbigh

School of Chemistry
# Table of Contents

Table of Contents .................................................................................................................. 2
List of Figures .......................................................................................................................... 10
List of Equations ...................................................................................................................... 20
Abbreviations .......................................................................................................................... 21
Publications .............................................................................................................................. 23
Abstract .................................................................................................................................. 24
Declaration ............................................................................................................................... 25
Copyright Statement ............................................................................................................... 26
Acknowledgements .................................................................................................................. 27

## Chapter 1  Introduction ......................................................................................................... 29

1.1 Thesis Roadmap .................................................................................................................. 29

1.2 Mammalian Cells and Cancer ........................................................................................... 30

1.2.1 Why Cell Analysis? ....................................................................................................... 30

1.2.2 The Cell Cycle ............................................................................................................... 31

1.2.3 Cancer Cell Metabolism ............................................................................................... 32

1.2.4 Lipogenesis and Cancer ............................................................................................... 33

1.2.5 Fatty Acid Biosynthesis .............................................................................................. 34

1.2.6 Increased Lipogenesis in Cancer .................................................................................. 36

1.3 Metabolomics and Lipidomics ......................................................................................... 38

1.3.1 Metabolomics ............................................................................................................... 38

1.3.2 Lipidomics .................................................................................................................... 39

1.3.3 Lipids ............................................................................................................................ 39

1.3.4 Analytical Techniques Employed in Metabolomics and Lipidomics ......................... 41

1.3.4.1 Expanding the Analytical Field in Metabolomics and Lipidomics.......................... 42

1.3.5 Traditional Metabolomic Workflow ............................................................................. 43
1.3.6 Traditional Lipidomic Workflow ................................................................. 45

1.4 Experimental Considerations in Metabolomic and Lipidomic Studies ....... 46

1.4.1 Experimental Design .................................................................................. 46

1.4.2 Sample Preparation .................................................................................... 47

1.4.2.1 Extraction and Staining .......................................................................... 47

1.4.2.2 Fixation .................................................................................................. 48

1.4.2.3 Aldehyde Fixation ................................................................................ 49

1.4.2.4 Organic Solvent Fixation ...................................................................... 51

1.4.3 Freeze Fracture .......................................................................................... 52

1.4.4 Data Interpretation ..................................................................................... 54

1.5 Applications of Analytical Techniques in Metabolomics and Lipidomics . 54

1.5.1 Mass Spectrometry in Metabolomics and Lipidomics ......................... 55

1.5.1.1 Metabolomic and Lipidomic Applications of ToF-SIMS in Profiling of Cancer ........................................................................................................ 56

1.5.1.2 Probing the Biochemistry of Other Diseases with ToF-SIMS ......... 58

1.5.1.3 Metabolomic and Lipidomic Applications of LC-MS ..................... 58

1.5.1.4 The Use of LC-MS in Understanding the Many Facets of Leukaemia ................................................................................................................................. 59

1.5.2 Spectroscopy in Metabolomics and Lipidomics ..................................... 60

1.5.2.1 Metabolomic and Lipidomic Applications of Raman Spectroscopy ......................................................................................................................... 61

1.5.2.2 Cellular Raman Imaging ..................................................................... 61

1.5.2.3 Raman Spectroscopy in the Study of Cancer ..................................... 62

1.5.2.4 Raman Spectroscopy Translation into the Clinic ................................ 65

1.5.2.5 Metabolomic and Lipidomic Applications of FTIR Spectroscopy ................................................................................................................................. 67

1.5.2.6 FTIR Applications for the Study of Leukaemia .............................. 69

1.5.2.7 Considerations in FTIR Studies .......................................................... 70

1.5.3 Applications of Multimodal Analytical Techniques to Metabolomics and Lipidomics ........................................................................................................ 71

1.6 Acute Myeloid Leukaemia ............................................................................ 74

1.6.1 Acute Myeloid Leukaemia Characterisation and Prognosis ............... 74
1.6.2 Treatment of AML ................................................................. 77
1.6.3 Drug Redeployment ............................................................. 79
1.6.4 Bezafibrate and Medroxyprogesterone Acetate (BaP)........... 79
1.7 Project Aims ........................................................................... 82

Chapter 2 Analytical Techniques and Instrumentation .................... 83

2.1 Secondary Ion Mass Spectrometry ........................................... 83
  2.1.1 Secondary Ion Formation and Sputtering ......................... 83
  2.1.2 Static SIMS ........................................................................ 85
  2.1.3 The Static Limit .................................................................. 85
  2.1.4 The SIMS Equation ........................................................... 86
  2.1.5 SIMS Instrumentation ....................................................... 87
    2.1.5.1 Time of Flight Mass Spectrometry (ToF-MS) ............... 89
    2.1.5.2 The Reflectron .............................................................. 92
    2.1.5.3 The Detector ................................................................. 93
  2.1.6 Methodology for ToF-SIMS Experiments ......................... 94

2.2 Liquid Chromatography-Mass Spectrometry .......................... 94
  2.2.1 Liquid Chromatography .................................................... 94
  2.2.2 Coupling Liquid Chromatography to Mass Spectrometry ...... 95
  2.2.3 Ion Trap Mass Spectrometry ............................................. 96
  2.2.4 The Orbitrap ................................................................. 96
  2.2.5 Methodology for UHPLC-MS Experiments ...................... 97

2.3 Infrared (IR) Spectroscopy ..................................................... 98
  2.3.1 Principles of Infrared Spectroscopy .................................. 98
  2.3.2 FTIR Instrumentation ...................................................... 103
  2.3.3 Synchrotron-FTIR ............................................................ 106
    2.3.3.1 Production of Synchrotron Radiation ......................... 106
  2.3.4 Synchrotron-FTIR Microspectroscopy ............................. 107
  2.3.5 Methodology for Synchrotron-FTIR Experiments ............. 109
2.3.6 ATR Spectroscopy ................................................................. 110
2.3.7 ATR-FTIR Methodology ......................................................... 111
2.4 Raman Spectroscopy ............................................................... 111
  2.4.1 Principles of Raman Spectroscopy ....................................... 111
  2.4.2 Raman Instrumentation ...................................................... 116
  2.4.3 Raman Microspectroscopy/Microscopy ................................ 117
  2.4.4 Methodology for Raman Experiments .................................. 118
2.5 Multivariate Analysis ............................................................. 118
  2.5.1 Unsupervised Analysis ....................................................... 118
  2.5.2 Principal Component Analysis (PCA) ................................... 118
  2.5.3 Data Pre-Processing ......................................................... 119
  2.5.4 Supervised Analysis .......................................................... 121
  2.5.5 Discriminant Analysis (DA) ............................................... 122
  2.5.6 Principal Component - Discriminant Function Analysis (PC-DFA)
  ................................................................. 122
  2.5.7 Partial Least Squares - Discriminant Analysis (PLS-DA) ............. 123
2.6 Summary ................................................................................. 124

Chapter 3 Experimental Principles and Methodology .................................. 126

3.1 Introduction ............................................................................. 126
3.2 Acute Myeloid Leukaemia In Vitro Cell Lines .............................. 126
  3.2.1 Cell Line Origin ................................................................ 126
  3.2.2 Cell Culture ...................................................................... 127
  3.2.3 Cell Counting .................................................................... 128
  3.2.4 Cell Viability Staining ....................................................... 129
  3.2.5 Cell Morphology and Characteristics ................................ 129
    3.2.5.1 Mycoplasma Testing of Cell Lines ................................. 130
3.3 Drug Compounds for Cell Inoculation ....................................... 131
  3.3.1 Drug Standards Preparation ............................................. 131
Chapter 3  Method Development for Single Cell Analysis

3.3.2 Raman of Drug Standards ......................................................... 131
3.3.3 Drug Treatment Protocol ......................................................... 133
3.3.4 Cell Cycle Considerations for Studying Drug-Cell Interactions ..... 133

3.4 Method Development for Single Cell Analysis ............................... 134

3.4.1 Deposition of Cells on Substrate for Analysis ............................ 134

3.4.1.1 Method Development to Adhere Cells to Substrate for Raman Microspectroscopy .......................................................... 134
3.4.1.2 Preparation of Whole Cells for Raman Microspectroscopy ...... 135

3.4.2 Raman Microspectroscopy Method Development ....................... 136

3.4.2.1 Raman Pre-Processing Considerations ................................. 137
3.4.2.2 Raman Excitation Lasers: 785 nm and 532 nm ..................... 139

3.4.3 Cell Fixation Method Development for Raman Microspectroscopy 140

3.4.3.1 Methanol Fixation Protocol ............................................... 140
3.4.3.2 Formalin Fixation Protocol .............................................. 141

3.5 Method Development for Cell Extract Analysis ................................ 144

3.5.1 Preparation of Cell Extracts for ToF-SIMS Analysis .................... 144
3.5.2 Results: ToF-SIMS Sample Preparation Study ........................... 145

3.5.3 Optimisation of Cell Extraction Procedure for Metabolomic and Lipidomic Studies ................................................................. 149

3.5.4 Experimental Procedure for Large Scale Metabolomics and Lipidomics Study Employing ToF-SIMS, UHPLC-MS and Raman Spectroscopy for the Analysis of Cell Extracts .............................................. 150

3.5.5 Preparation of Cell Extracts for Raman Analysis ........................ 151

3.6 Conclusions ............................................................................. 153

Chapter 4  A Spectroscopic Toolbox for Probing the Action of BaP Drug Therapy in AML Cells ......................................................... 154

4.1 Introduction ............................................................................. 154

4.2 Materials and Methods ............................................................. 155

4.2.1 Preparation of Cells and Measurement Details .......................... 155

4.2.2 Spectroscopic Data Processing .............................................. 157
4.2.2.1 FTIR Data Analysis ................................................................. 157
4.2.2.2 Raman Data Analysis ............................................................. 158
4.3 Results and Discussion .................................................................. 159
4.3.1 Mean Spectra Obtained from S-FTIR and ATR Analysis .............. 159
4.3.2 Statistical Analysis and Validation of S-FTIR and ATR Spectra ...... 161
4.3.3 Changes in Vibrations Arising from Lipid Species with BaP Treatment ................................................................. 164
4.3.4 Investigating the Effect of BaP on Cellular DNA with S-FTIR ....... 170
4.3.5 S-FTIR Analysis of Single Nuclei Extracted from Control and Drug Treated AML Cells ................................................................. 172
  4.3.5.1 Nuclei Extraction Methodology ............................................. 173
  4.3.5.2 Nuclei Extract Results ......................................................... 173
4.3.6 Raman Spectroscopy of Cell Extracts for Lipidomic Study ........... 176
4.3.7 Raman Microscopy to Map the Spatial Distribution of Lipids in Control and Drug Treated Cells ................................................................. 183
4.3.8 Biological Interpretation of the Effect of BaP on Cellular Lipids .... 188
4.3.9 Conclusion ............................................................................... 189

Chapter 5  A Metabolomics Investigation into the Anti-Leukaemic Action of BaP In Vitro by UHPLC-MS ................................................................. 190
5.1 Introduction ............................................................................... 190
5.2 Materials and Methods ............................................................... 190
  5.2.1 Cell Culture and Drug Treatment ........................................... 190
  5.2.2 Sample Preparation ............................................................... 191
  5.2.3 UHPLC-MS Methodology ..................................................... 191
  5.2.4 Data Quality ......................................................................... 191
  5.2.5 UHPLC-MS Data Processing .................................................. 191
  5.2.6 UHPLC-MS Metabolite Identification .................................... 192
5.3 Results and Discussion ............................................................... 192
Chapter 6 Exploring the Effect of BaP on the Lipidome of AML Cells Using Mass Spectrometry-Based Lipidomics

6.1 Introduction .................................................................................................................. 229

6.2 Materials and Methods ............................................................................................... 229

6.2.1 Cell Culture and Drug Treatment ............................................................................. 229

6.2.2 Preparation of Cell Extracts for Mass Spectrometric Analysis ......................... 229

6.2.3 UHPLC-MS Methodology ....................................................................................... 230

6.2.4 ToF-SIMS Methodology ......................................................................................... 230

6.2.5 Experimental Design for Obtaining Mass Spectrometric Data ....................... 230

6.2.6 UHPLC-MS Data Processing and Lipid Identification ....................................... 231

6.2.7 ToF-SIMS Data Processing ..................................................................................... 231

6.3 Results and Discussion ............................................................................................... 232

6.3.1 Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by UHPLC-MS .................................................................................................................. 233

6.3.2 PCA loadings analysis for UHPLC-MS non polar feature selection . 238

6.3.3 Unsupervised Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by ToF-SIMS .................................................................................................................. 246

5.3.1 Multivariate Analysis of UHPLC-MS Profiling Data from Drug Treated Cells ........................................................................................................................................ 193

5.3.2 PCA Loadings Analysis for UHPLC-MS Feature Selection ............................ 200

5.3.3 Putative Identification of Significant Features .................................................... 210

5.3.4 Pathway Analysis of Significant Metabolites Found in BaP Treatment Cell Extract Study for HL60 and K562 Cells ........................................................................................................................................ 212

5.3.4.1 Considering the Biological Significance of Pathway Topology Analysis Results ........................................................................................................................................ 221

5.3.5 Pathway Analysis of Significant Metabolites Found in Cell Extract Study for Comparison of HL60 and K562 Cells Response to BaP Treatment 222

5.3.6 Considerations for Metabolomic Data Interpretation ........................................ 227

5.4 Conclusion ................................................................................................................... 228
6.3.4 Supervised Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by ToF-SIMS .............................................................. 250
6.3.5 A Targeted Approach to Determine Lipid Saturation Status in Control and BaP Treated Cell Extracts Analysed by ToF-SIMS ....................... 255
  6.3.5.1 Probing BaP Induced Saturation Changes of Fatty Acid Acyl Chains .............................................................................. 256
  6.3.5.2 Probing Saturation Changes of Diacylglycerols in AML Cells Treated with BaP .................................................................. 260
6.4 Discussion and Conclusions .................................................................. 264

Chapter 7 Project Conclusions and Future Work ........................................ 265
7.1 A consideration of sample preparation and optimisation of instrument parameters facilitated the acquisition of high quality data covering a range of biomolecular compounds ................................................................. 265
7.2 A multi-platform spectroscopic investigation was employed to study the action of BaP therapy in AML and revealed biochemical perturbations on a single cell level .................................................................... 268
7.3 A metabolomics investigation into the action of BaP therapy in vitro by UHPLC-MS revealed a number of metabolic pathways that are disrupted by BaP therapy ..................................................................... 269
7.4 ToF-SIMS and UHPLC-MS proved to be powerful tools for probing the lipidome of AML cells and revealed saturation changes in cellular lipids post-BaP treatment ........................................................................ 271
7.5 Adopting a multi-disciplinary spectroscopic and mass spectrometric approach increased confidence in interpreting metabolomic and lipidomic changes with BaP treatment at the cellular level ........................................ 273
7.6 Future Prospects .................................................................................. 273

Appendix I ......................................................................................... 275
Appendix II ....................................................................................... 276
References .......................................................................................... 277

Final Word Count 75,557
List of Figures

Figure 1.1 Roadmap illustration summarising project and thesis chapters presented herein............................................................................................................................................... 29
Figure 1.2 Overview of the Cell Cycle. .................................................................................................................................................. 31
Figure 1.3 Overview of the fatty acid synthesis pathway showing the role of key enzymes ATP citrate lyase (ACL), Acetyl CoA carboxylase (ACC) and Fatty acid synthase (FAS)........................................................................................................................................ 35
Figure 1.4 Representative structures from each of the eight main lipid classes reproduced from reference [46]. .............................................................................. 40
Figure 1.5 Diagram to show typical untargeted and targeted Metabolomic workflows. ................................................................................................................................................................................. 43
Figure 1.6 Leukaemia incidence by age per year in the UK 2008 – 2010 [197]. ...... 75
Figure 1.7 Overview of the development of mature blood cells from haematopoietic stem cell. ........................................................................................................................................................................................................ 76
Figure 1.8 Structures of bezafibrate (A) and medroxyprogesterone acetate (B). ..... 80
Figure 2.1 The sputtering process in ToF-SIMS. The depth of penetration of the primary ions and (to a lesser extent) the information depth from which molecular ions are derived is dependent on the energy and incident angle of the primary ions, and the nature of the sample. Typically, in biological ToF-SIMS the information depth is <10 Å meaning that analysis is specific to the cell/tissue surface exposed to the analysis. .................................................................................................................................................. 84
Figure 2.2 Schematic detailing main components of BioTof SIMS Instrument. Sample is loaded into the SAC under vacuum where it is subjected to an incoming primary ion beam from either the Au\textsuperscript{+} or C\textsubscript{60}\textsuperscript{+} gun. Sputtered ions enter the ToF mass analyser, are reflected by the reflectron and subsequently detected .................................................................................. 88
Figure 2.3 Separation of ions in the drift region of a ToF-MS. Heavier ions take a longer time to drift the fixed distance and therefore arrive at the detector later than lighter ions.............................................................................................................. 90
Figure 2.4 Schematic of a reflectron positioned at the end of the drift region. ...... 92
Figure 2.5 A picture of an R-500 ToF mass analyser, reproduced from reference [240] ............................................................................................................................................................................................... 93
Figure 2.6 The Morse curve illustrating energy of a diatomic molecule undergoing anharmonic oscillation, with allowed vibrational energy levels illustrated......... 100
Figure 2.7 Typical biological FTIR spectrum and tentative peak assignments from a population of cells. Reproduced from reference [148].

Figure 2.8 Diagram of a basic FTIR spectrometer. Radiation from the IR source is reflected into the path of the interferometer where it is modulated by a moving mirror. This modulated radiation is then reflected from a further two mirrors through the sample, where it falls on a transducer. The data acquisition system records and stores the signal from the transducer.

Figure 2.9 Production of synchrotron radiation at the Australian Synchrotron facility.

Figure 2.10 FTIR spectra (offset) of a tissue section showing comparison of synchrotron radiation (upper) and globar-source (lower) acquired with a spectral resolution of 8 cm⁻¹, 1024 scans and from an area of 10 µm x 10 µm. Reproduced from reference [257].

Figure 2.11 Bruker Hyperion 3000 IR microscope at the Australian Synchrotron facility.

Figure 2.12 A schematic of the process of internal reflection spectroscopy. Adapted from reference [257].

Figure 2.13 Energetic processes occurring in Rayleigh (light scattered from a material is of the same frequency (energy) as incident light) and Raman scattering. In Stokes scattering, the molecule starts out in a lower vibrational energy state and ends up in a higher vibrational energy state after scattering. In anti-Stokes scattering, the opposite is observed. The frequency (energy) differences between the Raman lines and incident line are solely characteristic of the scattering molecules and are independent of the excitation frequency.

Figure 2.14 Typical cellular Raman spectrum with tentative peak assignments. Reproduced from reference [148].

Figure 2.15 Schematic of Renishaw inVia Raman microscope.

Figure 2.16 Representation of PCA for principal components 1 and 2. A K-dimensional variable space is set up showing three variable axes (A). Data points are plotted in the variable space (B). The vector of averages is calculated shown here as a red point (C). Mean-cenetering moves the average point to the origin (D). The first principal component, PC1 represents the maximum variance direction in the data (E) and the second principal component, PC2, reflecting the second largest variation in the data and orthogonal to PC1 (F). Figure adapted from reference [261].
Figure 3.1 Optical images of HL60 and K562 cells taken with a x 40 objective. A visual difference in morphology can be seen between the two cell lines. 

Figure 3.2 Raman spectra for bezafibrate (BEZ) and medroxyprogesterone acetate (MPA) drug standards (powder form) obtained at 532 nm (BEZ) and 785 nm (BEZ and MPA). Tentative peak assignments are numbered for each drug.

Figure 3.3 Representative noise filtered Raman spectra for control (black) and BaP treated (red) HL60 cells.

Figure 3.4 Raman whole cell extracted images (500–1800 cm\(^{-1}\)) for a typical HL60 control cell showing effects of data pre-processing. White light image is shown in (a). Heat maps portrayed in (b) and (c) depict regions of high intensity in red with lower intensity in blue. The spectrum (d) is an average spectrum which has undergone noise filtering and background subtraction depicting all biological material across the whole cell shown in (c).

Figure 3.5 White light microscopy images (x50 objective) of HL60 control cells with no chemical fixing and post methanol fixation.

Figure 3.6 White light microscopy images (x50 objective) of HL60 control cells with no chemical fixing and post formalin fixation.

Figure 3.7 White light microscopy image (x50 objective) of HL60 control cells post formalin fixation showing residual formalin crystals on surface.

Figure 3.8 Representative raw Raman spectra for HL60 BaP treated cells obtained at 532 nm excitation.

Figure 3.9 Plot to show m/z 184 peak area as a function of sample loading for HL60 cell extracts.

Figure 3.10 Plot to show m/z 184 average peak area (n=3) as a function of primary ion beam shots.

Figure 3.11 Positive ion ToF-SIMS raw spectra showing m/z 0 – 200 for HL60 BaP treated cell extract after washing with 0.15 M ammonium formate according to final wash procedure developed.

Figure 3.12 Plot to show ToF-SIMS peak signals for HL60 control cell extracts as a function of washing with 0.15 M ammonium formate solution.

Figure 3.13 Plot to show ToF-SIMS peak signals for HL60 BaP treated cell extracts as a function of washing with 0.15 M ammonium formate solution.

Figure 3.14 Workflow developed for AML cell extraction, adapted from the Folch and Bligh and Dyer methods for lipid extraction.
Figure 3.15 Preparation and treatment of cells in culture for large scale metabolomics experiment. ......................................................... 151
Figure 3.16 Images to show deposition of cell extract (outside and inside dried spots) for Raman analysis. ................................................................. 152
Figure 4.1 Microfluidics sample accessory device for use in the synchrotron FTIR Bruker Hyperion 2000 microscope (photograph courtesy of Robbin Vernooij). .... 155
Figure 4.2 Raman map data analysis work flow. .................................................. 159
Figure 4.3 Normalised, mean spectra from 9 replicates for ATR analysis of HL60 and K562 control and BaP treated cells and 3 replicates for each of 2 trials for S-FTIR analysis, incorporating a total of 441 individual cells (HL60) and 385 individual cells (K562). ................................................................. 160
Figure 4.4 Normalised, mean second derivate S-FTIR spectra from the high wavenumber region for control and BaP treated K562 and HL60 cells. ............... 161
Figure 4.5 Scores plots for the first oPLS-DA component for S-FTIR spectra of control (green) and BaP treated (red) cells. Each point represents a K562 single cell (a) and HL60 single cell (b). ................................................................. 162
Figure 4.6 Scores plots for the first oPLS-DA component for ATR spectra of control (green) and BaP treated (red) cells. Each point represents a replicate of cell pellet of K562 cells (a) and HL60 cells (b). ................................................................. 163
Figure 4.7 Classification model for PLS-DA of S-FTIR analysis of whole cell populations for K562 (a) and HL60 (b) cells showing classification rates for real (blue) and random (red), with a CCR of 1 being 100% correct classification. The embedded tables are the average confusion matrices with rows representing predicted classification and columns representing experimental. ..................................... 164
Figure 4.8 oPLS-DA loadings vectors overlaid for S-FTIR (red) and ATR-FTIR (blue) loadings in the high wavenumber lipid region, for K562 (a) and HL60 (b) cells.... 166
Figure 4.9 oPLS-DA loadings vectors overlaid for S-FTIR (red) and ATR-FTIR (blue) loadings in the fingerprint region, for K562 (a) and HL60 (b) cells. ...................... 166
Figure 4.10 Box and whisker plots for significant changes ($p < 0.05$, see Table 4.1) in peak area ratios of methylene to methyl stretching vibrations between control and 24 h drug treated cells analysed by S-FTIR. a. shows peak area ratios 2920:2955 cm$^{-1}$ for asymmetric C-H stretching and 2850:2870 cm$^{-1}$ for symmetric C-H stretching for K562 cells and b. shows the same ratios for HL60 cells. The median of each box plot can be compared to describe the changes in lipid peak area ratios with drug treatment;
with the maximum values, minimum values and outliers showing the spread in inter-cell variability within population sizes of 385 cells for K562 and 441 cells for HL60. Extreme outliers (n<10) were removed from the plot for visual appearance, however all cells analysed were included in the data analysis.

Figure 4.11  Box and whisker plots for significant changes ($p <0.05$, see Table 4.1) in peak area ratios of methylene to methyl stretching vibrations between control and 24 h drug treated cell pellets analysed by ATR. a. shows peak area ratios 2920:2955 cm$^{-1}$ for asymmetric C-H stretching and 2850:2870 cm$^{-1}$ for symmetric C-H stretching for K562 cells and b. shows the same ratios for HL60 cells. The trend observed is concurrent with synchrotron FTIR single cell data shown in Figure 4.10.

Figure 4.12  Box and whisker plots of peak height ratios for the asymmetric phosphate stretch at 1240:1220 cm$^{-1}$ calculated for the entire cell population of K562 cells (a.) and HL60 cells (b.) analysed by S-FTIR. The median line indicates no significant shift in asymmetric phosphate stretch peak position between control and drug treated cells for either cell class.

Figure 4.13  Normalised mean S-FTIR spectra of nuclei extracted from K562 control (light blue) and BaP (dark blue) treated cells.

Figure 4.14 PLS-DA scores plot for the first PLS-DA component showing discrimination between control (red) and BaP treated (green) cell nuclei analysed by S-FTIR.

Figure 4.15  PLS-DA loadings vector showing key bands in the spectra contributing to the discrimination between control and BaP treated cell nuclei.

Figure 4.16  Hierarchical cluster analysis for cell extract spectra selection from inside (a) and outside (b) of each dried sample spot. The x-axis is an arbitrary scale (0-1) of distance based on dissimilarity (closer to 0 equating to more similar).

Figure 4.17  Optical image from inside dried cell extract spot showing regions from where replicate Raman spectra shown in Figure 4.18 were obtained. Region 2 highlighted with white circle indicates sample heterogeneity.

Figure 4.18  Overlaid spectra from 3 different regions from the inside of dried extract spot of HL60 control cell extract.

Figure 4.19  PCA scores plots taken from inside (a) and outside (b) regions of dried sample for HL60 and K562 control and BaP treated cell extracts. Principal components 1 and 2 are shown, with cell type clearly separated along PC 1.
Figure 4.20  PCA scores plots taken from inside (a) and outside (b) regions of dried sample for HL60 and K562 control and BaP treated cell extracts. Principal components 1 and 3 are shown, with cell type clearly separated along PC 1.  

Figure 4.21 Loadings for PC 1 and PC 3 for inside (a) and outside (b) regions of dried cell extract.  

Figure 4.22  FCM images of a representative control and BaP treated K562 formalin fixed cell.  

Figure 4.23  VIP scores for PLS-DA of K562 control and drug treated cells showing three Raman lipid peaks.  

Figure 4.24  Average spectra from cluster 1 and cluster 2 regions, overlaid for control and BaP treated K562 cells. Black arrows in the enlarged spectra show the lipid regions that contribute to classification of control and drug treated cells.  

Figure 5.1  Representative example of UHPLC-MS data pre-processing showing selected features before and after log$_{10}$ transformation and Pareto scaling performed in MetaboAnalyst 3.0 prior to further statistical analysis.  

Figure 5.2  PCA scores plots obtained for HL60 and K562 control and BaP treated cells showing all replicates included in the polar UHPLC-MS metabolomics study. Coloured ellipses represent 95% confidence limits calculated from scores.  

Figure 5.3  PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of features directly associated with drugs administered. Coloured ellipses represent 95% confidence limits calculated from scores.  

Figure 5.4  3-dimensional visualisation of PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts.  

Figure 5.5  PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of features directly associated with drugs administered and including pooled QCs in the analysis. Coloured ellipses represent 95% confidence limits calculated from scores.  

Figure 5.6 Loadings plots obtained for HL60 and K562 polar cell extracts analysed by UHPLC-MS showing m/z features contributing to class separation between control and drug treated extracts along PC2 shown in Figure 5.3.  

Figure 5.7  Significant features identified by SAM (Significance Analysis of Microarray) in positive and negative ionisation. For features in green, the further the deviation from linearity, the greater their significance.
Figure 5.8 Pathway analysis for common features found in both HL60 and K562 cell extracts identified to within 2.5 ppm. Pyrimidine metabolism is identified as a key pathway targeted by BaP treatment with the highest significance (p-value) and pathway impact. Further pathways are also identified.

Figure 5.9 Pyrimidine metabolism pathway with significant features found in metabolomic analysis of control and BaP treated HL60 and K562 cell extracts shown in red.

Figure 5.10 Pathway topology analysis for features found in HL60 and K562 cell extracts respectively, identified to within 2.5 ppm. Amino sugar and nucleotide sugar metabolism is the most significant pathway affected by BaP treatment in K562 cells but is of much lower significance in HL60 cells.

Figure 5.11 Amino sugar and nucleotide sugar metabolism pathway with position of significant features found in metabolomic analysis of K562 control and BaP treated cell extracts shown in red.

Figure 6.1 PCA scores plots obtained for HL60 and K562 control and BaP treated cells and corresponding QCts analysed in positive and negative ESI modes. Data shown for all features found in 6 replicates included in the lipidomic UHPLC-MS study after the removal of features attributed specifically to drug administered. Coloured ellipses represent 95% confidence limits calculated from scores.

Figure 6.2 PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of K562 control outlier.

Figure 6.3 3-dimensional visualisation of PCA scores plots obtained for HL60 and K562 control and BaP treated non-polar cell extracts.

Figure 6.4 Loadings plots obtained for HL60 and K562 non-polar cell extracts analysed by UHPLC-MS, showing m/z features contributing to class separation between control and drug treated extracts along PC2 shown in Figure 6.2.

Figure 6.5 PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive and negative ionisation modes. Each point represents a replicate analysis from experimental replicates (n=4) and technical replicates (n=3).

Figure 6.6 Positive ion characteristic ToF-SIMS spectra (m/z 0-1000) and in-lay of enlarged spectra (m/z 300-900) showing an abundance of biological species for BaP treated K562 cell extract.
Figure 6.7 PCA scores and loading 3 plot obtained for K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive ionisation modes for lipid related features found in the region m/z 400-800. Each point represents a replicate analysis from experimental replicates (n=4) and technical replicates (n=3), extreme outliers removed.

Figure 6.8 PC-CVA scores and loading for CV 2 obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive ionisation mode, for features observed in the region m/z 300-900.

Figure 6.9 PC-CVA scores and loading for CV 2 obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in negative ionisation mode, for features observed in the region m/z 300-900.

Figure 6.10 Negative ion characteristic ToF-SIMS spectrum of BaP treated cell extract showing m/z range 100 – 500 with fatty acid acyl chains identified.

Figure 6.11 Plot to show average normalised peak areas for fatty acid acyl chains of different degrees of saturation, identified in HL60 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in negative ionisation mode. No data was determined to be significant, p-values > 0.05 obtained for all data in this series.

Figure 6.12 Plot to show average normalised peak areas for fatty acid acyl chains of different degrees of saturation, identified in K562 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in negative ionisation mode. Significant data indicated with * (p < 0.05).

Figure 6.13 Plot to show average normalised peak areas for diacylglycerols of different degrees of saturation, identified in HL60 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in positive ionisation mode. No data was determined to be significant, p-values > 0.05 obtained for all data in this series.

Figure 6.14 Plot to show average normalised peak areas for diacylglycerols of different degrees of saturation, identified in K562 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in positive ionisation mode. Significant data indicated with * (p < 0.05).

Figure 6.15 Plot to show saturated and unsaturated DAG normalised peak areas for 4 individual replicates of K562 control (blue) and BaP treated (red) cell extracts identified by positive ionisation ToF-SIMS. Both DAG(36:0) and DAG(36:2) data series were calculated to be statistically significant (p = 0.026 and 0.018 respectively).
List of Tables

Table 1.1  The eight primary lipid classes. ................................................................. 40
Table 2.1  Qualitative comparison summary of multimodal spectroscopic and mass spectrometric techniques employed in this project for the analysis of cells. ............ 125
Table 3.1  Typical cell populations counted using a haemocytometer during routine cell culture of AML cell lines. ............................................................................. 129
Table 4.1  Student t-test p-values to quantify the group separation in the PLS-DA scores of each cell line analysed by S-FTIR and ATR-FTIR. ............................................. 170
Table 4.2  PLS-DA K-fold-cross-validation model on 11 K562 cells. Rows represent predicted classification and columns represent experimental. .......................... 185
Table 5.1  Correlation between ANOVA and PC 2 loadings for significant features found in cell extracts. ................................................................................................. 202
Table 5.2  Numerical correlation between top 10 significant features according to ANOVA and PCA with SAM significance for cell extracts. The higher the SAM d-value, the more likely it is to be significant. ................................................................. 207
Table 5.3  Box plots for the 10 most significant features identified by univariate and multivariate analysis contributing to differences between control and BaP treated cell extracts. Red and navy blue box plots indicate relative amounts of each particular compound in HL60 and K562 BaP treated cell extracts respectively, with green and light blue box plots representing relative amounts of compound in HL60 and K562 control cell extracts respectively. ................................................................. 209
Table 5.4  Putative identifications of the top 10 significant features observed in positive and negative ESI LC-MS for drug treated cell extracts. ........................................ 211
Table 5.5  Top 10 pathways identified by pathway topology analysis of metabolite features affected by BaP treatment in HL60 and K562 cell extracts. Pyrimidine metabolism shows a significant change with highest pathway impact. See text for details. .................................................................................................................. 215
Table 5.6  Fold changes of metabolite features found in pyrimidine metabolism targeted by BaP treatment. Box plots show relative amounts of metabolite found in cell extracts for HL60 BaP treated (red), HL60 control (green), K562 BaP treated (navy blue) and K562 control (light blue). ................................................................. 220
Table 5.7  Top 10 pathways identified by pathway topology analysis of metabolite features affected by BaP treatment according to cell type. Amino sugar and nucleotide
sugar metabolism is the most significant pathway affected by BaP treatment in K562 cells but does not feature in the top 10 significant pathways for HL60 cells. 6 pathways out of the top 10 are identified as being common between the two cell types (illustrated here by matched colour shading).

Table 6.1 Putative identifications and relative fold changes of significant \((p < 0.05)\) lipid related species observed in positive and negative ESI UHPLC-MS identified by multivariate analysis as contributing to differences between control and BaP treated cell extracts. \(m/z\) values listed in red indicate a decrease with BaP treatment and those in green indicate an increase with BaP treatment. Red and green box plots indicate relative amounts of each feature in HL60 BaP treated and control cell extracts respectively, whilst blue and turquoise show the same for K562 cell extracts. All species identified are within 2.5 ppm mass error.
List of Equations

Equation 2.1 Fundamental SIMS equation [230]................................................................. 86
Equation 2.2 Kinetic energy of ions in relation to ion mass and velocity .................. 89
Equation 2.3 Relationship between kinetic energy of ions and potential applied .... 90
Equation 2.4 Relationship between time and distance travelled by ions .............. 91
Equation 2.5 Time of Flight equation ............................................................................... 91
Equation 2.6 Chromatography partition coefficient......................................................... 95
Equation 2.7 Frequency of oscillation of a spring in harmonic oscillation .......... 99
Equation 2.8 Reduced mass of a two-body system......................................................... 99
Equation 2.9 Absorbance of IR ......................................................................................... 102
Equation 2.10 The Beer-Lambert law ............................................................................. 103
Equation 2.11 Dipole induced by oscillating electric field .......................................... 112
Equation 2.12 Polarizability changed by vibrational displacement ...................... 112
Equation 2.13 Fluctuation with time of incident electromagnetic wave .............. 112
Equation 2.14 Time-dependent induced dipole moment .............................................. 113
Equation 2.15 Induced dipole moment in terms of polarizability change with vibrational displacement ........................................................................................................ 113
Equation 2.16 Raman equation describing three distinct frequencies of induced dipole moment......................................................................................................................... 113
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Acetyl CoA carboxylase</td>
</tr>
<tr>
<td>ACL</td>
<td>ATP-citrate lyase</td>
</tr>
<tr>
<td>A-DNA</td>
<td>A-Deoxyribonucleic acid</td>
</tr>
<tr>
<td>AML</td>
<td>Acute myeloid leukaemia</td>
</tr>
<tr>
<td>AMPK</td>
<td>AMP-activated protein kinase</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>ATR</td>
<td>Attenuated total reflection</td>
</tr>
<tr>
<td>BaP</td>
<td>Bezafibrate and Medroxyprogesterone acetate</td>
</tr>
<tr>
<td>BEZ</td>
<td>Bezafibrate</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast Cancer 1</td>
</tr>
<tr>
<td>CaF₂</td>
<td>Calcium Fluoride</td>
</tr>
<tr>
<td>CCR</td>
<td>Correct classification rate</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic myeloid leukaemia</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CON</td>
<td>Control</td>
</tr>
<tr>
<td>CV</td>
<td>Cross-validation</td>
</tr>
<tr>
<td>CVA</td>
<td>Canonical variates analysis</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DAG</td>
<td>Diacylglycerol</td>
</tr>
<tr>
<td>DESI</td>
<td>Desorption electrospray ionisation</td>
</tr>
<tr>
<td>DFA</td>
<td>Discriminant Function Analysis</td>
</tr>
<tr>
<td>DIMS</td>
<td>Direct infusion mass spectrometry</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EMSC</td>
<td>Extended multiplicative signal correction</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionisation</td>
</tr>
<tr>
<td>FAS</td>
<td>Fatty acid synthase</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>FCM</td>
<td>Fuzzy c-means</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared</td>
</tr>
<tr>
<td>G₁</td>
<td>Gap 1</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>HCA</td>
<td>Hierarchical cluster analysis</td>
</tr>
<tr>
<td>HIF</td>
<td>Hypoxia-inducible factor</td>
</tr>
<tr>
<td>HMDB</td>
<td>Human metabolome database</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LV</td>
<td>Latent variable</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix assisted laser desorption ionisation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MCR</td>
<td>Multivariate curve resolution</td>
</tr>
<tr>
<td>MCT</td>
<td>Mercury cadmium telluride</td>
</tr>
<tr>
<td>MoS</td>
<td>Morphological Scores</td>
</tr>
<tr>
<td>MPA</td>
<td>Medroxyprogesterone acetate</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MSI</td>
<td>Mass spectrometry imaging</td>
</tr>
<tr>
<td>MVA</td>
<td>Multivariate analysis</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OPLS-DA</td>
<td>Orthogonal partial least squares-discriminant analysis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>PC-PC</td>
<td>Principal component</td>
</tr>
<tr>
<td>PC-PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PC-CVA</td>
<td>Principal component-canonical variates analysis</td>
</tr>
<tr>
<td>PC-DFDA</td>
<td>Principal component-discriminant function analysis</td>
</tr>
<tr>
<td>PDMS</td>
<td>Polydimethylsiloxane</td>
</tr>
<tr>
<td>PE</td>
<td>Phosphatidylethanolamine</td>
</tr>
<tr>
<td>PGD_2</td>
<td>Prostaglandin D2</td>
</tr>
<tr>
<td>PI</td>
<td>Phosphatidylinositol</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial least squares</td>
</tr>
<tr>
<td>PLS-DA</td>
<td>Partial least squares-discriminant analysis</td>
</tr>
<tr>
<td>PQN</td>
<td>Probabilistic quotient normalisation</td>
</tr>
<tr>
<td>QC</td>
<td>Quality control</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>RMicS</td>
<td>Resonant Mie scattering</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RSD</td>
<td>Relative standard deviation</td>
</tr>
<tr>
<td>SAM</td>
<td>Significance analysis of microarrays</td>
</tr>
<tr>
<td>SCD1</td>
<td>Stearoyl-CoA desaturase</td>
</tr>
<tr>
<td>S-FTIR</td>
<td>Synchrotron-Fourier transform infrared</td>
</tr>
<tr>
<td>SIMS</td>
<td>Secondary ion mass spectrometry</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SNV</td>
<td>Standard normal variate</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid</td>
</tr>
<tr>
<td>TFA</td>
<td>Total fatty acids</td>
</tr>
<tr>
<td>ToF-MS</td>
<td>Time of flight-mass spectrometry</td>
</tr>
<tr>
<td>ToF-SIMS</td>
<td>Time of flight-secondary ion mass spectrometry</td>
</tr>
<tr>
<td>TUFU</td>
<td>Total unsaturated fatty acids</td>
</tr>
<tr>
<td>UHPLC</td>
<td>Ultra high performance liquid chromatography</td>
</tr>
<tr>
<td>UMP</td>
<td>Uridine cyclic monophosphate</td>
</tr>
</tbody>
</table>
Publications


Abstract

University: The University of Manchester
Candidate Name: Joanna Lynne Denbigh
Degree Title: Doctor of Philosophy
Thesis Title: Lipidomic and Metabolomic Analysis of Biological Response Mechanisms in Cancer Cells: A Multidisciplinary Approach
Date: March 2016

The 21st Century has seen a rise in incidence of complex diseases such as cancer and in the quest to develop essential new therapeutic options, the study of drug-cell interactions can yield powerful information. Acute myeloid leukaemia (AML) is an aggressive cancer that causes life-threatening deficits of functional blood cells in humans for which current treatment options are highly toxic and often poorly tolerated. A combination of two existing drugs, bezafibrate and medroxyprogesterone acetate in a drug redeployment situation has shown promise in vitro and in vivo and further investigations are crucial to elucidate the mode of action of this treatment.

This project investigated the mechanistic action of BaP at a cellular level. Orthogonal spectroscopic and mass spectrometric platforms were employed to probe the biochemical composition of two AML cell lines, HL60 and K562 in the presence and absence of this combined drug treatment. Analysis was performed on single living cells, dehydrated cells, fixed cells and cell extracts to give a large and detailed data set.

A consideration of the main spectral differences obtained by Synchrotron-FTIR and ATR-FTIR in conjunction with multivariate statistical analysis revealed a significant change to the cellular lipid composition with drug treatment; furthermore, this response was not caused by cell apoptosis. In particular, the ratio of CH₂:CH₃ was observed to increase with BaP treatment and this was determined to be a significant change in both cell lines (p < 0.05). An overall increase in lipid unsaturation suggests that BaP targets cellular lipid biosynthesis. Raman microspectroscopy added a further dimension to the spectroscopic study by providing spatial information of lipid distribution which suggested that BaP-induced saturation change is uniform across a single cell.

UHPLC-MS was employed for global metabolomics analysis of AML cell extracts and revealed a number of biochemical pathways that were indicated as targets of BaP therapy in both cell lines. Univariate and multivariate analysis determined statistically significant metabolites for which putative identifications were made. Pyrimidine metabolism was the most significant pathway identified for changes consistent in both HL60 and K562 cell lines.

The complementarity of ToF-SIMS and UHPLC-MS provided large coverage of the lipidome of AML cells through untargeted and targeted approaches. For data derived by both techniques, a general increase in polyunsaturated species for BaP treated cell extracts was observed which correlated well with findings from spectroscopic investigations.

Adopting a multi-disciplinary approach to cell analysis can afford a powerful insight into understanding drug mode of action at a cellular level and novel information regarding BaP mechanistic action in AML cell lines was revealed. This analytical approach could be extended to the future study of drug-cell interactions for other oncological systems.
Declaration

No portion of the work referred to in the 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.

Joanna L. Denbigh
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 Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and 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 (see http://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgements

I would sincerely like to thank my supervisors Dr Nick Lockyer and Professor Roy Goodacre for their scientific input, encouragement, patience and personal support during my project. Thank you to Nick for opening his office door to me so many times.

My collaborators at the University of Birmingham, in particular Dr Andrew Southam and Dr Farhat Khanim have been extremely helpful in advising on the biological aspects of the project. Through them I have found an area of research that really inspires me. Special thanks to Andy for frequent chats - some out of hours and way beyond the call of duty.

Belonging to two research groups has significantly broadened my scientific horizons and I would like to thank all members of the SARC and BioSpec groups, in particular Dr Kat Hollywood for cell culture and Raman support and Dr Alex Henderson and Dr Yun Xu for data analysis support. A big thanks to Irma for showing me the ropes when I first started and to Giles for making me laugh on difficult days.

Special thanks to Professor Peter Gardner who has been a fantastic mentor to me, always there for advice and support (despite not having to be), and thanks to members of the Gardner group with whom I have enjoyed many drinks and scientific discussions.

During the final year of my PhD, I was fortunate to spend 2 months at Monash University, Australia as a visitor of Associate Professor Bayden Wood. I am extremely grateful to all group members and synchrotron staff who supported me in my research there. Particular thanks to Bayden for inviting me and providing funds to part-finance my visit, to Dr David Perez-Guaita for assistance with data processing and to Robbin Vernooij – my synchrotron wing-man. I would also like to acknowledge two very special people in Melbourne who made my trip a truly memorable one - thank you Bev and Mark for fun, food and friendship.

I am very grateful to the BBSRC for funding this PhD project. I also gratefully acknowledge CLIRSPEC who awarded me a research grant to facilitate my trip to Monash and the RSC, BACR, BMSS and ABS for travel grants which allowed me to present my work at conferences in the UK, Germany, Poland, USA and Turkey.
I could not have given up work and embarked upon my PhD without the support of my wonderful friends and family and my sincere thanks to all far and wide for years of encouragement. Particular thanks to the fabulous tap girls for listening to me ramble on every week (the best friends any one could wish for) and to my Bro, Jonathan for providing the tunes for ‘making the moves’ when I needed it.

My heartfelt thanks to a very special man David Bradley, for so many things…but especially for always being by my side to share the highs and lows with me for the past four years and for giving me the determination to succeed.

Finally, I thank my parents, two inspirational people who have walked this path with me every step of the way, and without whose practical help and continual love and understanding this would not have been possible. My T.S. and P.G. - for a lifetime of believing in me and always shining the light – we did it!

I dedicate this thesis to the memory of Dr Angela Gernaey, who planted the seed many years ago.
Chapter 1 Introduction

1.1 Thesis Roadmap

This thesis presents a multi-dimensional study into the action of a novel combination therapy (BaP) at a cellular level in two acute myeloid leukaemia cell lines (HL60 and K562). Spectroscopic and mass spectrometric platforms have been used to probe drug-cell interactions to yield rich data sets which, upon interpretation, aid in understanding the mode of action of this therapy.

Figure 1.1 illustrates a roadmap summarising the contents of each chapter presented herein.

Figure 1.1 Roadmap illustration summarising project and thesis chapters presented herein.
1.2 Mammalian Cells and Cancer

1.2.1 Why Cell Analysis?

Eukaryotic organisms are made up entirely of cells, which may exist as single entities or come together to form more complex tissues, organs and systems. The field of cell analysis requires a dedicated multidisciplinary approach to provide insights into the complexity of mammalian life in health and disease. Cell analysis typically involves studying cells which have been extracted from their mammalian origin and cultured in vitro. Despite this being a fabricated environment, it provides scientists with the ability to study small biochemical changes both in single cells and within cell populations when subjected to stress or perturbations under defined conditions, which aids in understanding the mechanistic details of cellular processes. Single cell analysis is particularly desirable for cell populations showing significant heterogeneity in which the cells of interest are in the minority and could be masked by the behaviour of the majority of cells [1].

In studying cells at this level, one can derive a whole variety of qualitative and quantitative information regarding cellular characteristics such as morphology, genotype, metabolomic activity and physiological properties; the understanding of which are essential in the never ending drive to cure disease and maintain life. The techniques employed to explore the aforementioned properties are varied and span a range of disciplines. Traditionally, optical and fluorescence microscopy have been widely employed to image cells which are usually pre-stained with dyes to provide sub-cellular contrast [2]. Other well reported imaging techniques include electron microscopy, x-ray microscopy, confocal optical microscopy, scanning probe microscopy (such as atomic force microscopy) and more recently vibrational spectroscopy (Fourier transform Infrared (FTIR) and Raman) and mass spectrometry (matrix assisted laser desorption ionisation (MALDI) and time of flight-secondary ion mass spectrometry (ToF-SIMS) for example); the latter of which offer an extra dimension of chemical information previously unattainable with early microscopic techniques [3]. With this in mind, spectroscopic and mass spectrometric techniques are not confined to cell imaging, but are widely employed in probing the chemical nature of cells and cell extracts as well. Traditional molecular biological techniques which also aim to do this include polymerase chain reaction (PCR) and western
blotting for probing cells at the protein or gene level [4], but more recently the wider range of aforementioned analytical platforms are also being employed to offer complementary and often more detailed information to the traditional biological techniques. When such data obtained is interpreted in a multi-disciplinary setting combined with that from other techniques, it is possible to piece together biologically relevant information regarding cellular behaviour in both healthy and diseased states. As analytical technology further advances, there is no doubt that the toolbox for such applications will continue to expand.

1.2.2 The Cell Cycle

A typical mammalian cell replicates and divides in a process known as the cell cycle, which for a majority of mammalian cells spans approximately 24 hours, with some exceptions including faster growing embryonic cells and slower growing liver cells [5]. The cell cycle is traditionally described in terms of four successive phases, the M (mitosis), G₁ (gap), S (synthesis) and G₂ (gap) phases. The M phase is the most marked of these, in which mitosis (nuclear division) and cytokinesis (cytoplasmic division) pave the way for cell division. The other three phases make up Interphase, in which the cell grows in size (G₁ and G₂ phases) and most importantly the cell synthesises and replicates its DNA in the S phase. An overview of this cycle can be seen in Figure 1.2.

![Figure 1.2 Overview of the Cell Cycle.](image-url)
In a population of cells proliferating asynchronously in culture, cells will exist in a distribution across each of these four phases with, for example, approximately 30% typically being in S phase [5]. This heterogeneity adds to the complexity of studying a cellular system, for which labelling or fluorescent staining of cells can assist with phase of cell cycle determination and may be useful in bioanalytical studies reporting a small population of single cells. Cell cycle considerations are further addressed in section 3.3.4.

### 1.2.3 Cancer Cell Metabolism

Mammalian cells obtain energy via two main metabolic processes; lactic fermentation and aerobic respiration. Lactic fermentation converts glucose to pyruvate through glycolysis and sees pyruvate reduced to lactate, overall yielding two molecules of ATP per glucose molecule. Aerobic respiration is a more complex process in which glucose is fully oxidised to carbon dioxide and water via glycolysis, the Krebs cycle and oxidative phosphorylation. This relies on there being oxygen present and yields a much higher amount of energy than lactic fermentation [6].

‘Normal’ cells undergo the latter oxidative phosphorylation-ATP production in an oxygenated environment, however it has been known for almost a century that cancer cells exhibit an enhanced conversion of glucose to lactate even in the presence of oxygen [7]. This heavy dependence on glycolysis manifests in an increased uptake of glucose in contrast to normal cells and represents a shift in metabolism from respiration to fermentation, a phenomenon first reported by Otto Warburg in the 1920’s [8, 9]. ‘Aerobic glycolysis’ or the ‘Warburg effect’ has led to a multitude of studies into the biochemical pathways that redirect the metabolism in cancer cells. A number of key pathways have been indicated in increased biosynthesis, with cancer cells typically producing far greater amounts of macromolecules, nucleotides and lipids than observed in healthy cells and at the same time consuming at least 10 times more glucose [10]. Glycolysis, nucleotide synthesis and lipid synthesis are all key pathways upon which cancer cells depend and a deeper understanding of cancer metabolism can pave the way for the development of new anti-cancer therapies. This project focuses on the lipidomic and metabolomics changes in acute myeloid leukaemia cells when treated with a novel combination therapy.
1.2.4 Lipogenesis and Cancer

Fatty acids play an important role in many biochemical pathways and are essential for healthy cell biological processes. Synthesis of fatty acids endogenously is known as *de novo* lipogenesis and traditionally serves the purpose of converting excess carbohydrates into lipids for storage, lipid being more energy dense than carbohydrate. Lipids can also be obtained from the diet and can be stored or utilised according to the action of the liver. Because free fatty acids are toxic to cells, lipids which are employed as energy stores within the cell are usually stored in the esterified forms of mono-, di- and tri-acylglycerols which are often found deposited in lipid droplets [11]. Such reservoirs can then be utilised as an energy source or other cellular processes as required.

Lipids are crucial features of biological membranes, forming the lipid bi-layer responsible for structural integrity. Membrane embedded phospholipids and glycolipids are closely involved with membrane proteins involved in cell signalling and transporters, in fact the composition of lipid vesicles involved in membrane traffic has been discovered to be a crucial factor for protein targeting. Some lower abundance but highly important lipids fulfil roles as precursors of second messengers or have specialised functions as lung surfactants or in milk production [12]. Studies of lipid species are therefore of great interest and can yield specific information on physiological functions within tissues and cells.

Lipid biology has become a major research target since the birth of the systems biology approach to understanding disease. The plasma membrane in particular has been the focus of many phospholipid studies [13-16] and the widely reported strong link between enhanced lipogenesis and cancer has led to a vast amount of studies of the nature of lipids and enzymes, involved in the lipid biosynthesis pathway. The development of a cancerous tumour involves prolific cell growth and division which causes the cell to undergo significant morphological and biochemical changes, particularly to the lipid-rich cell membrane. Lipid droplet reservoirs are readily available for providing both energy and lipid requirements for the generation of new cell membranes in proliferating cells and such lipid bodies are reportedly to be observed in higher concentrations in cancer cells when compared with healthy cells [17].
As one would expect, de novo synthesis is highly active in foetal lungs to produce surfactant and in lactating adults to produce breast milk, but otherwise one would expect adult human de novo lipogenesis and expression of lipogenic enzymes to be low [18]. However, the biological processes by which fatty acids are synthesised in the mammalian cell have been a focal point of research for over half a century, primarily because of the large number of reports that link over-expression of several enzymes involved in de novo lipid biosynthesis with a variety of human tumours [12]. In fact, the list of types of cancer showing elevated lipogenic enzyme expression keeps growing and Fernandis et al suggest this as a basis for lipid biomarker discovery in cancer and other diseases [11].

1.2.5 Fatty Acid Biosynthesis

Fatty acids synthesis involves the condensation of two-carbon units, namely acetyl CoA, to form long chain hydrocarbons in a series of biochemical reactions. Acetyl CoA is produced from pyruvate, the end product of glycolosis, which is transported into mitochondria and decarboxylated by the pyruvate dehydrogenase enzymatic complex to give acetyl CoA. Transportation of the acetyl CoA out of the mitochondria is required as fatty acid synthesis takes place in the cytosol, however the inner mitochondrial membrane is impermeable to acetyl CoA, so it becomes combined with oxaloacetate to form citrate, which readily crosses the membrane. Once in the cytosol the citrate is cleaved by ATP-citrate lyase (ACL) to regenerate acetyl CoA. The remaining oxaloacetate is returned to the mitochondrial matrix through conversion to malate and then pyruvate, a step which generates NADPH which can be used in fatty acid synthesis.

Acetyl CoA, now in the cytosol, is available for the first committed step of fatty acid synthesis; the carboxylation to malonyl CoA by acetyl CoA carboxylase (ACC), a biotin containing enzyme found in the endoplasmic reticulum. Following this, malonyl CoA and acetyl CoA undergo a sophisticated series of elongation steps to form long chain fatty acids, primarily palmitate. This elongation cycle of fatty acid synthesis consists of four sequential stages of condensation, reduction, dehydration and further reduction, all catalysed by the enzyme fatty acid synthase (FAS) and employing nicotinamide adenine dinucleotide phosphate (NADPH) as a reducing agent (acquired
from the conversion of pyruvate to malate or via the pentose phosphate pathway). Each cycle adds two-carbon units to the lengthening acyl chain and after 7 cycles (14 carbons added to original 2-carbon unit), the C16 fatty acid palmitate is formed. The elongation of fatty acids beyond C16, to C18 (stearate) and further, is carried out by elongases and desaturases, enzymes located in the endoplasmic reticulum. These newly synthesised fatty acids are used for subsequent synthesis of different cellular lipids such as sphingolipids, triglycerides, phospholipids and cholesterol esters. A summary of the pathways and biochemical reactions involved in fatty acid biosynthesis can be seen in Figure 1.3.

![Figure 1.3](image-url)  
**Figure 1.3** Overview of the fatty acid synthesis pathway showing the role of key enzymes ATP citrate lyase (ACL), Acetyl CoA carboxylase (ACC) and Fatty acid synthase (FAS).
1.2.6 Increased Lipogenesis in Cancer

In the early 1990’s it was observed that human breast cancer cells exhibit high levels of fatty acid synthase (FAS) expression and activity [19]. Since then a multitude of clinical and laboratory studies have deduced that human cancers are able to bypass the regulatory signals that down-regulate fatty acid synthesis, and provide themselves with their own fatty acid supply. The phenomenon of enhanced lipogenesis has recently been described as one of the most important hallmarks of cancer cells [20]. Kuhjda et al have compiled a review table of cancers that were reported in scientific literature to show presence of FAS at unusually high levels [21], which at the time included cancers of the breast, prostate, colon, ovary, thyroid and endometrium. As one would expect, this increased lipogenesis corresponds to significantly elevated activity and high-level expression of lipogenic enzymes such as fatty acid synthase and ATP citrate lyase [18]. As well as these examples of elevated FAS levels, the key role of lipogenesis in cancer has also been demonstrated through studies which have inhibited lipogenesis in cancer cells by RNA interference or pharmaceutical agents. The outcomes were suppressed DNA replication, cell cycle arrest and apoptosis of the cancer cells [22], [21].

The mechanistic detail explaining the overexpression of FAS in cancer has been the subject of many scientific studies and is complex and still poorly understood [12]. However, since blocking the activity of FAS inhibits the growth and proliferation of cancer cells, much research is focussed on identifying compounds with FAS inhibitory properties for the potential development of new anti-tumour drugs. Compounds including some polyphenols (as found in green tea) [23], [24] and steroid-based molecules (such as ursolic acid) [25], [26] have been reported.

Two further enzymes, acetyl CoA carboxylase (ACC) and ATP citrate lyase (ACL), which also have strong indications for increased lipogenesis in cancer cells are the topic of inhibition investigations, but only a limited number of compounds with selectivity for these enzymes have currently been identified. ACC is the rate limiting enzyme in fatty acid synthesis and yet has historically attracted less attention than FAS. Elevated ACC expression has been reported early in the development of breast cancer [27] but it wasn’t until the breakthrough in the discovery of the BRCA1 gene mutations in breast cancer that the subsequent importance of ACC became apparent, when BRCA1 gene mutations were observed to prevent interaction with ACC [28]. In a re-
deployment scenario, the drug Metformin (developed for hyperglycemia and type 2 diabetes) has been found to inhibit the proliferation of a wide range of cancer cells through the activation of AMP-activated protein kinase and subsequent phosphorylation of ACC which downregulates lipogenesis [29].

Rysman et al reported a significant finding regarding the saturation state of lipids found in tumours [30]. Intact phospholipids from prostate cancer cells that had been treated with a highly potent inhibitor of ACC (soraphen A), were analysed by gas chromatography-mass spectrometry (GC-MS) and electrospray ionisation tandem mass spectrometry (ESI-MS/MS). Mammalian cells have a limited availability to synthesise polyunsaturated fatty acids \textit{de novo} as they lack the \( \Delta 12 \) desaturase required, Rysman et al hypothesised that \textit{de novo} lipogenesis would enrich cancer cells mainly with saturated and/or mono-unsaturated acyl chains. Indeed, their study provides direct evidence that cancer cells promote saturation of their membranes and modulate this by activating \textit{de novo} lipogenesis. It is known that mono-unsaturated or saturated lipids are less susceptible to lipid peroxidation than the more reactive polyunsaturated acyl chains and results of this study also suggested that enhanced lipogenesis of cancer cells renders them less susceptible to lipid peroxidation by limiting the degree of phospholipid unsaturation [30]. This is significant as results also supported the growing evidence that oxidised phospholipids and their degradation products play a key role in immune response, inflammation and cellular apoptosis [31, 32]. Data shown supports the idea that the change in the level of saturation of cellular membranes in cancer cells protects them from oxidative stress-induced cell death.

In 2011 it was stated that there remains much room for improvement in our understanding of the mechanisms that couple the processes of lipogenesis and tumour development [12] and many researchers hypothesise that the design of new anticancer drugs that will inhibit lipid production could provide breakthroughs in the field of cancer therapy.
1.3 Metabolomics and Lipidomics

It has become widely accepted that it is not possible to fully understand biological systems, from microbial or plant species to higher organisms, simply by studying a small part of this system. Instead, the emerging ‘systems biology’ approach is being widely adopted to complete the missing gaps. Recent efforts are focussed on merging all branches of ‘-omics’, such as metabolomics, lipidomics and proteomics and combining this with key understanding of the biological pathways affected to make valuable inroads into this vast field of scientific research.

1.3.1 Metabolomics

Metabolomics refers to the comprehensive measurement of the small molecule component of a biological sample to give a picture of what the physiological status of the organism or sample was at the time of collection [33]. It emerged as a complementary discipline to genomics and proteomics in the last couple of decades, offering a completely different insight into biological behaviour by viewing biological systems in relation to the metabolite content of anything from a single cell to whole organism. The close proximity of metabolites to the phenotype of the biological system means that small and often rapid biochemical changes in the cell can be observed [34]. Metabolites span a diverse range of low molecular weight molecules including lipids, amino acids, nucleic acids, vitamins and carbohydrates and as such no single analytical platform is able to characterise the global metabolome.

The study of such biochemical species is employed to gain powerful insights into a variety of applications such as clinical [35], pharmaceutical [36], nutritional [37] and environmental [38]. Both untargeted and targeted approaches are employed to detect and analyse metabolic species and examples of these are discussed in section 1.3.5.

Metabolites are small molecules of organic origin that are responsible for regulating many biological processes, and there are reportedly in excess of 16,000 named metabolites on record [39]. They can be regarded as the ultimate output of cellular machinery [40] and encompass a wide variety of roles including those of by-products, substrates, cell signallers and cellular building blocks. Thus the presence or absence of a certain metabolite could suggest disruptions to a particular cellular pathway and
allow key biomarkers to be discovered and/or investigated. The implication for this is huge, taking cancer as an example, which is responsible for millions of deaths worldwide every year. The development of reliable methodologies for the screening of individuals with the disease for early diagnosis and even prediction of treatment could potentially save the lives of many people.

### 1.3.2 Lipidomics

Lipidomics can be defined as the systems-level analysis and characterisation of lipids and their interacting moieties and today has been described as very much an emerging field of research [41]. Because lipids and associated species are crucial components of cells and tissue, many recent studies have demonstrated how they play an important role in health and disease. Ivanova et al in 2009 discuss how the many facets of lipidomics reflect both the diversity of lipid species and the wide roles of functions mediated by lipids [42]. They reference studies that see lipids implicated in HIV replication [43], embryo implantation [44] and Parkinson’s disease [45].

Ten years ago, lipidomics was focussed on mass spectrometry based profiling to detect and identify lipid species, however today a far more comprehensive, systems-biology approach is adopted in which the interaction of lipids with other cellular molecules and processes to gain insight into their functions is adopted. This has largely been driven by developments and advances in analytical instrumentation.

### 1.3.3 Lipids

Lipid classification is vast, reflecting the diversity of lipid species found in nature. Lipid structures are generally much more complex than genes and proteins due to the number of different biochemical transformations which occur during their biosynthesis. In 2005, a comprehensive report was published to facilitate international communication about lipids for emerging lipid communities to discuss the vast amounts of data that were, and still are, being generated [46]. This was the work of the International Lipid Classification and Nomenclature Committee on the initiative of the LIPID MAPS Consortium [47] and attempted to integrate all lipid information from throughout the world into one common language. Initially, eight primary categories
were proposed, based on well-defined biochemical principles. These categories can be seen in Table 1.1. and representative structures of lipids within each class can be seen in Figure 1.4 [46, 48].

<table>
<thead>
<tr>
<th>Lipid Category</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acyls</td>
<td>FA</td>
</tr>
<tr>
<td>Glycerolipids</td>
<td>GL</td>
</tr>
<tr>
<td>Glycerophospholipids</td>
<td>GP</td>
</tr>
<tr>
<td>Sphingolipids</td>
<td>SP</td>
</tr>
<tr>
<td>Sterol lipids</td>
<td>ST</td>
</tr>
<tr>
<td>Prenol lipids</td>
<td>PR</td>
</tr>
<tr>
<td>Saccharolipids</td>
<td>SL</td>
</tr>
<tr>
<td>Polyketides</td>
<td>PK</td>
</tr>
</tbody>
</table>

Table 1.1 The eight primary lipid classes.

Figure 1.4 Representative structures from each of the eight main lipid classes reproduced from reference [46].
Each main lipid category is further divided into classes, subclasses and in the case of some prenols, fourth level subclasses. This classification system is based on the concept of two fundamental ‘building blocks’, the ketoacyl groups and isoprene groups and has since been widely adopted as the main classification system in lipidomic studies. Under the LIPID MAPS system, each individual lipid is assigned a unique 12 or 14 character identifier which assists with lipid identification when mining vast amounts of data [49]. Fatty acid tail groups which are encountered in the majority of lipid classes have standard nomenclature to describe their composition. This states the number of carbon atoms present in the fatty acid chain and the number of double bonds, thereby providing information on the saturation of the chain. Examples are C16:0 which consists of a 16 carbon chain with no double bonds (a fully saturated species), and C18:1 with an 18 carbon chain and one double bond (a monounsaturated species).

1.3.4 Analytical Techniques Employed in Metabolomics and Lipidomics

Due to the chemical diversity of small molecules (typically of molecular weights less than 1500 Da), it is not possible to study the whole lipidome or metabolome using a single analytical technology. Instead, a multidimensional approach is typically employed in which data from a number of different orthogonal analytical techniques is mined.

The two approaches that historically dominate metabolomic studies are nuclear magnetic resonance (NMR) and mass spectrometry (MS) which is often coupled to liquid chromatography (LC) or gas chromatography (GC) at the front end. However, there are a number of alternative and equally significant techniques that are employed. Spectroscopic methods offer rapid spectral acquisition and are thus frequently run as a highly useful initial screening step before more time consuming mass spectral analysis. FTIR and Raman spectroscopies offer high throughput technology in which thousands of samples per day can be screened at a relatively low cost as well as affording the possibility of a more detailed sub-cellular investigation when these spectroscopic techniques are coupled to a microscope.
Capillary electrophoresis (CE) is a less common technique but has recently been applied in a study by Canuto et al. in which CE was interfaced with electrospray ionisation mass spectrometry (CE-ESI-ToF-MS) for use as a complementary technique to LC for separation of metabolites of Leishmania [50].

### 1.3.4.1 Expanding the Analytical Field in Metabolomics and Lipidomics

Raman microspectroscopy and FTIR microspectroscopy offer the capability to probe a single cell with sub-cellular resolution and thus derive substantial information about the chemical composition and biological pathways of a cell in its more native state. In recent years Raman and FTIR have been widely used in the field of oncology, for example in non-invasive diagnosis of skin cancer [51, 52], bladder cancer diagnosis [53, 54] and to characterise metabolism of breast cancer cells with different degrees of malignancy [55]. The collection of spectra can be *in vitro, ex vivo or in vivo* without disrupting the chemical environment, providing localised information about a cell’s nucleic acid, protein and lipid contents [56, 57]. In fact, the FTIR absorbance spectrum has been described as one of the most information-rich and concise ways to represent the whole ‘omics’ of a cell, thereby fulfilling all the requirements for the development of clinically relevant biomarkers [58].

Mass spectrometry imaging techniques (MSI) have become a rapid growth area in recent years, particularly in the field of metabolomics and lipidomic applications. MSI can also provide spatial distributions of various biomarkers adding a further dimension to that of conventional MS. Two methods that have been the main focus of development in recent years are that of matrix assisted laser desorption ionisation mass spectrometry (MALDI-MS) and time of flight-secondary ion mass spectrometry (ToF-SIMS).

Examples in the literature of metabolomic and lipidomic studies, including drug-cell interactions and cellular imaging utilising four key techniques employed in this project (ToF-SIMS, LC-MS, Raman spectroscopy and FTIR spectroscopy) are explored in sections 1.5.1 (MS) and 1.5.2 (spectroscopy).
1.3.5 Traditional Metabolomic Workflow

Approaches are typically either ‘untargeted; or ‘targeted’. This refers to whether the analyst is profiling ‘unknowns’ in the case of untargeted analysis; or studying ‘known’ species in the case of targeted analysis which usually aims to quantify the metabolite or lipid species. The workflows can be seen in Figure 1.5.

Untargeted metabolomics provides information on the simultaneous presence of many metabolites which can typically be hundreds at a time, in the form of features such as peaks (NMR/chromatographic) or ion traces. Metabolic ‘fingerprints’ are generated which aim to detect all of the metabolite or lipid species present without regard for their identification [59] and are subsequently processed as a comparison between control and test material to look for observable differences between groups.

Figure 1.5 Diagram to show typical untargeted and targeted Metabolomic workflows.
The identification of unknown compounds is often defined as the main challenge in untargeted metabolomics [60], and thus combining data from a number of techniques that give information such as accurate mass, fragmentation data, isotopic pattern and structural location of protons, can assist in contributing to the basis for compound identification. Quantitation however is very limited in this approach since differences in abundance of metabolites is observed rather than absolute concentrations.

Mass spectrometry today is able to detect thousands of lipids and metabolites with high accuracy and precision. Fourier Transform Ion Cyclotron Resonance MS (FT-ICR-MS) and new generation Time of flight MS (ToF-MS) can provide accurate masses to within 0.1 ppm or 2 ppm respectively [61]. This high mass accuracy, together with isotopic abundance information, allows prediction of chemical formulae for unknown species. Formulae can then be searched against comprehensive databases for identification of species but whilst this is a crucial step in metabolomics, it does not stand alone; as the identified species need to then be placed into context of a biological system - the ‘systems biology’ approach to ‘–omics’ studies previously referred to.

Targeted metabolites encountered can have a wide range of physio-chemical properties and occur in a wide concentration range making them challenging to quantify, but targeted metabolomics works with a defined set of metabolites. Janeckova et al focussed on the diagnosis of inherited metabolic disorders in plasma samples using a targeted metabolomic approach [62]. In this study the samples were analysed with flow injection analysis-tandem mass spectrometry with electrospray ionisation. Multiple reaction monitoring (MRM) was used for quantitation of 163 endogenous metabolites from different metabolite classes.

Often the two approaches are combined whereby after initial screening and identification of key biomarkers, some attempt at quantitation of these is made. Using a combination of metabolomic profiling by LC/GC-MS followed by targeted GC-MS, Sreekumar et al established high throughput methodology to profile more than 1,126 metabolites across 262 clinical samples related to prostate cancer (tissue, plasma and urine) [63]. The unbiased metabolomic profiles obtained were able to distinguish between benign, clinically localised prostate cancer and metastatic disease.
1.3.6 Traditional Lipidomic Workflow

There are two predominant methods employed in lipidomics, ‘shotgun’ lipidomics and tandem mass spectrometry, both of which have been used to identify and quantify phosphohlipids [42].

Shotgun lipidomics refers to the direct infusion of lipid species into the mass spectrometer with no prior chromatographic separation. There are a number of techniques such as direct infusion mass spectrometry (DIMS) with electrospray ionisation of the species of interest, yielding a single mass spectrum representative of the whole metabolite composition in the sample. Precursor ion scans and neutral loss scans are employed on tandem quadrupole instruments to identify key lipid fragments, such as the polar head groups. A recent study by Stahlman et al demonstrates a shotgun lipidomic approach by hybrid quadrupole time of flight mass spectrometry for the absolute quantitation of hundreds of intact molecular lipid species such as sphingolipids, glycerolipids and glycerophospholipids [64]. Here sample infusion was automated by nanospray, and precursor ion scan analysis was used for quantitation of lipid species.

Alternatively, flow injection mass spectrometry (FIMS) has been employed as previously demonstrated by Vaidyanathan et al in 2002 [65] and more recently Jenackova et al in 2012 [62]. However, according to Allwood et al, the aforementioned DIMS is more likely to produce information-rich spectra as the extract is not diluted in the flow of solvent as in FIMS [66]. Indeed, mass spectrometric direct analysis techniques have been the subject of significant instrument developments in recent years and a large number of applications across a wide range of direct analysis techniques are now employed. Desorption electrospray ionisation (DESI) has been used in imaging mode to differentiate between renal cell carcinoma tissue and adjacent normal tissue by lipid profiling the spatial distribution of glycerophospholipids and free fatty acids [67].

Tandem mass spectrometry (MS/MS) is a well proven technique for the identification and quantitation of lipids because there are many isobaric lipid species which are not separated by direct infusion MS. Within the 1000+ phospholipid species present in mammalian cells, there are even intra-class isobaric compounds such as 38:4 PI which can be composed of 18:0/20:4, 18:1/20:3 or 16:0/22:4 fatty acid combinations [42] and
it is often not possible to separate these chromatographically, posing a further problem to the analyst. However, for a vast number of lipid species, there are many examples of the successful application of tandem mass spectrometry [68-70].

1.4 Experimental Considerations in Metabolomic and Lipidomic Studies

There is one pertinent question that arises in the ‘-omics’ field, specifically with regard to metabolomics and lipidomics: ‘is what you see the true story’?

With recent advances in state of the art technology allowing detection and quantification of metabolites and lipids to extremely low levels, which are often employed in a high throughput environment, it is easy to find a marker that appears to distinguish between a ‘biological’ state and ‘control’ state. However, without rigorous control over experimental design, including appropriate and sufficient quality controls in place, one can easily be misguided by an apparent ‘significant’ find, which is in fact an artefact of the sample, instrument or environment.

1.4.1 Experimental Design

An organised approach to experimental design is crucial to the success of metabolomics experiments. A set of clear objectives should be chosen, from which the experimental design is devised. This is particularly important if multivariate analyses (MVA) are to be carried out on the data, as all too often scientists will turn to this in an attempt to solve a problem when it is too late [71]. The collection of data sets needs to be carefully considered according to which MVA techniques are likely to be employed; so that potential issues with sample groups or populations can be identified early. This becomes even more crucial when sample supply is limited, which is a common occurrence in biological and cellular analyses.

The need for an optimised extraction technique is also a key part to this experimental design. The quest for the most comprehensive and information-rich data set does not stop at the choice of analytical technique(s) employed. There have been many studies
that compare and contrast the methodology by which samples are prepared for analysis.

1.4.2 Sample Preparation

A great challenge in metabolomics and lipidomics, in particular for single-cell metabolomic studies, is sample preparation. Because the chemistry and concentrations of intracellular metabolites can change rapidly, cell preparation techniques to capture the most accurate picture of the metabolome at a particular point in time are essential. Whilst there have been a number of studies in sample preparation, there is still scope for further research to be carried out in this field as a key focus in metabolomics and lipidomics studies to date has been on analytical instrumentation and techniques [72].

1.4.2.1 Extraction and Staining

There are a number of published methods for the extraction of lipids and metabolites from cells for subsequent analysis. The Bligh and Dyer extraction method is widely employed as a gold standard for metabolomic assays [73]. It is one of a number of solvent-extraction based methods, after which, chromatographic separation and analysis of lipid species can be carried out by mass spectrometry or NMR. This approach (of cell extraction followed by LC-MS, GC-MS or NMR) provides useful qualitative and occasionally quantitative data of the extracted lipids and metabolites but is labour intensive and time consuming, which in itself is not as big a disadvantage as the fact that by taking these biomolecules out of their native environments and homogenising them, any functionality linking them to specific biological pathways occurring within the cell is lost [74]. It therefore follows that a consideration of cell extract data in conjunction with single cell analysis can provide further information on cellular lipidomic and metabolic biochemistry which the aforementioned methodology alone cannot.

Tissue and cell staining has been a popular option for studying in vivo single cell lipidomics, as fluorescent probes, such as Nile red which is a stain for neutral lipids, are readily taken up by lipid membranes into the cell. Fluorescent microscopy enables the visualisation of sub-cellular structures and Gbelcova et al in their recent study on
the chemoprotective effects of the drug class statins, used Nile red to detect lipid droplets by fluorescent spectroscopy in human embryonic kidney cells and human pancreatic cancer cells [75]. The disadvantage of such staining however, is that fluorescent probes can significantly interfere with the biological activity of the cell and many organic fluorophores can be toxic to cells at high concentrations [76]. The popularity of this technique is still widespread, particularly with recent advances in optical microscopy providing opportunities for very high resolution images, but it is now becoming commonly used as a confirmatory or complementary technique alongside other, less invasive cellular analytical techniques.

But, the question of sample preparation for single cell analysis still arises. While a more ‘native’ chemical state can be analysed without homogenisation or fluorescent tagging, one still needs to consider how a cell behaves when removed from the environment in which it has grown.

Cell behaviour and chemical composition can change considerably when removed from a temperature regulated, pH controlled, growth medium. Cells dehydrate and internal structures can collapse leading to delocalisation of biomolecules; osmotic pressure is disrupted and can result in cell shrinkage or even swelling which can rupture the cell membrane and as well as these physical changes, autolysis can begin. Autolysis is the destruction of a cell by intracellular enzymes which initiate the denaturing of proteins and dephosphorylation of mononucleotides, phospholipids and proteins and is undesirable since the biochemical composition of the cell will be significantly altered from its healthy natural state [77].

1.4.2.2 Fixation

A number of cell fixation options are available and there currently appears to be little agreement in the literature as to which method is preferable, indeed the effects of fixation on the organ, tissue or cell being prepared have been intensively studied for more than a century [78]. As well as choice of fixative, there are many other parameters to consider such as the concentration of fixative, duration of fixing, temperature and pH of fixative and time-span from removal of cells from their natural environment to point of fixing. Zeng et al have recently pointed out that much of our current
understanding of fixation comes from historical investigations that focussed mostly on organs and tissues; and still effects of fixing on individual cells continues to be relatively understudied, which possibly explains why there is so much diversity currently reported [78].

The objective of fixation is to preserve the structural and biochemical constituents of cells in as close to in vivo conditions as possible [77]. However, fixation is known to change the conformation of the DNA from the B-form to the A-form, which can be easily identified by FTIR analysis. In the more disordered A-form, the DNA FTIR bands appear weak and broad compared to B-form and it is therefore difficult to discern them from other macromolecules like protein, RNA and carbohydrates [79]. This is pertinent in drug-cell interaction studies when trying to assess whether or not there has been intercalation of the drug with the DNA resulting in either conformational change or denaturing of the molecule. It is clear, therefore, that there can be significant advantages of live cell studies over that of fixed cells, primarily because fixation involves the use of chemicals which can interfere with the inherent biochemical signature of the cell, thus potentially altering the data derived from the cell.

Common methods involve chemical fixation or flash freezing for subsequent freeze drying, and the choice not only depends on the chemical nature one wants to preserve best but also on the substrate that will be used in the analysis. Chemical fixation methods generally fall into 2 classes, organic solvents and cross-linking reagents.

1.4.2.3 Aldehyde Fixation

Aldehyde fixation, with glutaraldehyde, paraformaldehyde/formaldehyde or formalin, is by far one of the most commonly employed cross-linking methods. It fixes cells by forming cross-linkages between proteins without harming the protein structure [80]. It is also reported that formalin can preserve lipids by the reaction of hydrated formalin with double bonds present in unsaturated hydrocarbon chains [81].

A recent study investigated the minimum concentration of fixative required to completely fix individual cells and reported that the effect of glutaraldehyde fixation on cell structures was more than ten-fold stronger than that of formaldehyde or
paraformaldehyde, according to fixative concentration [78]. Various cellular structures were investigated and it appeared that the concentration of fixative required was correlated with the size of these structures. A limiting parameter for cell fixation was deduced not to be the rate of penetration of the fixative into the cell, but more the speed of reaction between the fixative and cellular proteins; because the diffusion law of fixatives suggests that all fixatives will penetrate deeply into cells within seconds. But still this study, as with many, only focuses on one or two aspects of fixing. Here the formation of cell blebs as imaged by confocal microscopy was considered in the main, which might give a good indication of the rigidity of the cell after cross-linking of proteins, but tells us nothing about the actual chemical composition of the cell.

Gazi et al fixed prostate cancer cells in 4% formalin in phosphate buffered saline solution (PBS) and subsequently washed the fixed cells in deionised water to remove the PBS. Upon checking the integrity of the cell membranes with trypan blue staining, some loss of plasma membrane lipids was observed (positive stain result) [77]. This did not seem to have a significant effect upon the intracellular localisation of biochemical species but certainly implies that it is not one hundred percent effective for the study of lipids. A further example of the delocalisation of certain chemical species by aldehyde fixation has been reported. Glutaraldehyde fixation was compared to cryofixation by plunge freezing and both scanning electron microscopy and ToF-SIMS revealed cracks on the cell membranes showing poor membrane integrity. Moreover, potassium diffusion from the cell interior was indicated, a direct result of the reaction of glutaraldehyde with the membrane proteins, terminating their activity in the sodium/potassium pump within the cell membrane. ToF-SIMS spectra also showed relatively lower yields for the phosphocholine ion at $m/z$ 184 than those previously suggested indicating that the glutaraldehyde fixed cells had less well preserved membrane bi-layer structure than for the cryofixation method [82].

Meade et al employed Raman microspectroscopy to compare the effects of three fixatives: neutral-buffered formalin, Carnoy’s fixative and a methanol-glacial acetic acid mix for the analysis of three different mammalian cell lines [83]. This study is of particular interest because data obtained for fixed cells were directly compared to the same cells analysed in their live state. Despite spectral adjustments observed for bands assigned to lipids, which was suggestive of formalin affecting the vibrational freedom
of these biomolecules, formalin fixed cells displayed the closest spectral resemblance to that of live cells.

In a recent FTIR study investigating tumour cell sensitivity to an apoptosis inducing drug, it was acknowledged that there were limitations of evaluating changes in lipids as a marker because cells had been fixed in formalin which introduces CH$_2$ groups into the samples studied, thereby contributing to the IR spectra at the same vibration frequency as CH$_2$ groups from fatty acid chains in lipids [84]. This is surprising given the number of widely reported studies which employ formalin as a fixative for cell analysis and make no mention of this, however it could be taken into account when interpreting resulting data.

1.4.2.4 Organic Solvent Fixation

Alcohols are classified as precipitating or coagulant fixatives which act by reducing the solubility of proteins, aggregating them on the cellular architecture. The widespread use of organic solvent fixation, in particular methanol fixation, could be a concern given the numerous studies to date that report the dehydration effects that alcohols inflict upon cells as well as their ability to dissolve and remove cellular lipids. Vekemans et al reported an interesting study which investigated the localisation of two membrane bound molecules that play a key role in apoptosis. Effects of fixation of liver endothelial cells with methanol, paraformaldehyde and a mixture of paraformaldehyde and glutaraldehyde, were investigated. DNA, as observed by confocal laser scanning microscopy, was seen to have escaped the nuclear compartment indicating that methanol had permeabilised the nuclear cell membrane, leaching biological molecules [85]. In a study of silicon used as an alternative matrix free substrate for MALDI-MS, human embryonic kidney cells were prepared for imaging by fixing in 70% methanol. Results were compared and verified with staining methodology to confirm the spatial distribution of observed biomolecules at the cellular level by this new 2-D technique, with no mention of the effects of the methanol on the cells at all. An optical image was shown with the corresponding mass spectrometry ion map of the phosphocholine (PC) lipid headgroup at m/z 184.1, making the assumption that the PC molecules remained in situ with the methanol
fixing. A closing statement alluded to the need to focus further on sample preparation for further work, but more with the goal of improving image performance rather than spatial accuracy of cellular molecules [86].

Lipid droplets are ubiquitous components of most types of cells and play an important role in cell trafficking and lipid storage in mammalian cells. A number of proteins are known to associate specifically with lipid droplets and it is therefore important to employ appropriate analytical methodology when studying these species. DiDonato et al were concerned with the lack of standardisation of microscopy methods in use and carried out a study to compare the most commonly used methods of fixation and looked at the effects of fixatives on cellular lipid content and lipid droplet structure. Cold methanol fixation, which has been classically used to study cytoskeletal structure and subsequently commonly employed as a cell fixative, was compared with acetone fixation and paraformaldehyde fixation, both of which have also been previously widely employed. Lipid droplets and their associated proteins in whole cells were imaged by fluorescence microscopy and effects on lipid distribution and chemical composition (as determined by protein assay and lipid extraction followed by thin layer chromatography (TLC) analysis) were assessed. Larger lipid droplets were observed in the methanol fixed cells when compared to paraformaldehyde fixation, suggesting fusion of lipid droplets, most likely due to removal of surface phospholipids which otherwise act as a barrier to fusion. Furthermore, quantitative TLC revealed a depletion of more than 80% of phospholipids in methanol fixed cells when compared to unfixed control cells.

1.4.3 Freeze Fracture

For imaging techniques such as SIMS which operates under a high vacuum, cells need to be dry and/or frozen prior to entry into the sample analysis chamber, which can lead to loss of cell viability [87]. As well as considering preservation of morphological and chemical features, one also needs to consider the removal of media such as undesired salts which could contaminate or suppress the ion signal. Chandra et al first demonstrated the successful use of cryogenic freeze fracture preparation of cells for ion microscopy and SIMS analysis in 1986 [88]. The preparation involves initial flash freezing of cells whilst sandwiched between two substrates to render the cells in either
a frozen hydrated form if maintained at low temperatures; or they can be freeze dried to remove any excess water content. This flash freezing in a cryogenic liquid ensures that the formation of large ice clusters is prevented, thereby preserving the native distribution of cellular species. The substrates are then pulled apart under vacuum to expose various internal cell surfaces for analysis [87]. The technique has since been employed by many groups, in particular for SIMS analysis of frozen-hydrated cells because the frozen hydrated environment has been shown to increase ion yields for lipids and other cellular species [89],[90]. Roddy et al report molecule-specific images for PC12 neuronal cells, used to identify which regions of the cells were exposed during the freeze fracture process. Images were generated for signals such as sodium and potassium which have differing intracellular and extracellular concentrations at the time of flash freezing and these images were used to determine whether the cell was cross-fractured exposing the cytoplasm and if either the outer leaflet of the cell membrane or the inner leaflet of the cell membrane was exposed to the surface for analysis [91].

Sophisticated advancements in instrumentation have been required to successfully maintain a cell in its frozen hydrated form and to ensure the success of the freeze-fracture analysis. A freeze-fracture device for frozen hydrated samples that is compatible with high resolution imaging was reported by Lanekoff et al. PC12 samples were freeze fractured in situ and analysed with a Bi cluster primary ion beam to generate high resolution images of typically low abundant lipids such as phosphatidylcholine (PC) and phosphatidylethanolamine (PE). The J105 SIMS instrument at the University of Manchester is equipped with a novel in vacuo freeze fracture device which uses a modified sample holder referred to as the ‘mousetrap’ [92]. The process of sample preparation and transfer has been described in detail and a study was carried out to compare data from HeLa cells which had been either fixed and freeze dried, or analysed in a non-fixed, frozen hydrated state, concluding that the latter shows improved localisation of nuclear material and increased signal levels for characteristic biomolecular species [92].

More recently, research has focussed on developing procedures for the analysis of frozen hydrated cells without the need for a freeze-fracture device [93]. This presents a simplified approach to sample preparation with a reported advantage being the large
numbers of intact biomolecules such as membrane lipids which can be observed in the higher mass range.

1.4.4 Data Interpretation

For the majority of metabolomic and lipidomic studies, a vast array of raw data is generated, for which robust chemometrics must be employed to analyse this data and turn it into useful information from which biochemical understanding can be gained. The complexity and volume of data acquired from a single SIMS experiment alone is so enormous that patterns or correlations in data can be extremely difficult to find. Data processing and analysis with sophisticated software routines is often crucial for the subsequent interpretation of ToF-SIMS experiments [71, 94-96].

Data interpretation has also become increasingly challenging in recent years as mass spectrometric and spectroscopic instrumental improvements have given rise to increased resolution and sensitivity. It is therefore no coincidence that the majority of biological mass spectrometry or spectroscopy studies of disease that are discussed herein, employ multivariate analysis to deconvolute the large spectral data sets.

Statistical analysis for metabolomics studies is widely reported and further information on a range of chemometric methodologies and considerations such as data scaling, cross validation and avoiding false discoveries in metabolomics studies is readily available [97-103]. Two of the most commonly applied multivariate analysis techniques, principal component analysis and principal component-discriminant function analysis are described in the next section.

1.5 Applications of Analytical Techniques in Metabolomics and Lipidomics

Traditionally, LC-MS and NMR based approaches as discussed in section 1.3.4 have been employed in metabolomics and lipidomics for many years and are reportedly applied in greater than 80% of all published metabolomics studies [101]. Despite this, in the past decade analytical techniques for single cell studies have emerged and very much expanded the scope of the field of metabolomics and lipidomics, offering
different information to the aforementioned methodologies [57, 72]. Specific applications of some of these techniques are explored further in this section.

1.5.1 Mass Spectrometry in Metabolomics and Lipidomics

Mass spectrometry coupled to liquid chromatography or gas chromatography at the front end are gold standard methodologies in metabolomics and lipidomics, having been extensively employed for decades. The nature of these hyphenated techniques requires cells or tissue to be extracted prior to analysis. The efficiency of extraction protocols for isolating metabolites and lipids is therefore of paramount importance to the success of subsequent analysis and must be considered as an integral part of the methodology.

Mass spectrometry imaging (MSI) is a rapidly developing analytical tool for the study of biomolecule distribution in cells and tissues in situ. This expanding field includes a variety of ionisation techniques such as SIMS, MALDI and DESI. There have recently been a number of comprehensive reviews of MSI which cover a range of techniques, including biological applications of MALDI, which cover a wide variety of applications across different diseases [104-108]. ToF-SIMS is an emerging technique for probing the chemistry of cells and other biological material and brings a number of desirable features to the analytical tool box. In contrast to MALDI, no addition of matrix to sample is required and the ability to focus the primary ion beam to small spot sizes (down to a few hundred nanometres) potentially allows for higher spatial resolution. Cluster and polyatomic primary ion beams such as \( \text{Au}_n^+ \), \( \text{Bi}_n^+ \) and \( \text{C}_{60}^+ \), which have been developed in the last decade, are capable of generating increased ion signal from molecular materials than previously observed with atomic ions such as \( \text{Ga}^+ \) and \( \text{Cs}^+ \) [109, 110]. This has significantly increased the sensitivity of the technique making ToF-SIMS a novel tool for probing the surface of biological materials. The combination of improved sensitivity and high spatial resolution mean that single cell analysis can be performed in which the distribution of molecular species on the surfaces of single cells is mapped; which can provide a deeper understanding of the role of these biomolecules in disease. ToF-SIMS analysis has been applied to a wide range of biological tissues and cells, and lipids found within these materials.
particularly lend themselves to characterisation by ToF-SIMS due to their relative ease of ionisation in the SIMS process.

1.5.1.1 Metabolomic and Lipidomic Applications of ToF-SIMS in Profiling of Cancer

The composition and spatial distribution of lipids in cell membranes and tissue has been of interest for many years as a factor to understanding the role of lipids in disease. Brunelle and Laprevote [111] and Passarelli and Winograd [112] extensively reviewed developments in ToF-SIMS lipid analysis, reporting the variety of biological applications that have been a direct result of improvements in ToF-SIMS sensitivity. With the direct link between lipids and many cancers, the technology is a sensible choice for probing this disease.

In an early ToF-SIMS cancer cell study by Gazi et al. the capability to image molecular signals in freeze fractured single human prostate cancer cells using a Ga\(^+\) beam was first demonstrated [113]. Many of the ions reported are consistent with lipid species, including membrane phospholipids which have since been reported [112]. Subsequently, using a C\(_{60}\)\(^+\) primary ion beam for enhanced sensitivity, discrimination between different cell types was shown to be possible using ToF-SIMS and principal component discriminant function analysis (PC-DFA) [114]. Ions selected for analysis were those representative of specific cellular biomolecules such as phospholipids and amino acids in the range m/z 58 - 500. Discriminant function 1 in positive ion mode separated cell lines associated with bone and lymph node metastases and normal prostatic epithelial cells. This work suggested the potential of ToF-SIMS analysis for chemically discriminating cancers based upon their severity (bone metastatic prostate cancer being responsible for high mortality), which is important in a clinical context. This study also highlighted the often crucial role of statistical analysis such as principal component analysis (PCA) and PC-DFA in differentiating complex ToF-SIMS spectra from cellular systems which can often be weaker in total ion signal than that of tissue systems.

PCA was extensively used in the noteworthy study by Kulp et al in their analysis of human breast cancer cells by ToF-SIMS which was one of the earliest reports of its
type and paved the way for much of the cancer work currently reported [115]. Initially PCA was employed as a data reduction tool for seven pure protein standards with the resulting scores plot capturing 74% of the variance of the data set in PC1 and PC2. The corresponding loadings plot identified that the ions responsible for the largest variance in the spectra resulted from known biochemical properties of proteins such as iron content. PCA was then further employed to study three closely related human breast cancer cell lines, initially for homogenised cell lines and finally for individual cells. ToF-SIMS images were taken of six fractured cells of each cell line and a composite spectrum of the entire cellular region was used for PCA. The scores plot showed clustering for the different cell types for both homogenised and individual cells. The loadings of \(m/z\) values in principal components were however more difficult to interpret than for protein standards as many unassigned peaks at low \(m/z\) proved significant in the variance between the cell spectra. Some key ions that were identified as being important in the differentiation include \(m/z\) 184 of the phosphocholine head group and \(m/z\) 70 (arginine, asparagine, leucine, proline or valine), 81 (histidine) and 95 (histidine) which are present in pure spectra of amino acids [116, 117].

A similar study was later reported by Park et al, which employed ToF-SIMS and principal component analysis, and demonstrated the potential for ToF-SIMS to differentiate between normal colon mucosa tissue and colon cancer tissue based upon the observation of amino acid fragments from proteins [118]. Here a bismuth cluster liquid metal ion gun was used to obtain ToF-SIMS images of the tissue surface, and image-PCA analysis proved very powerful in highlighting differences between healthy and cancerous colon tissue surfaces.

In a recent metabolomic study by Armitage et al, multicellular tumour spheroids cultured from human colon cells with presence or absence of hypoxia inducible factors were investigated by ToF-SIMS [119]. Spectra were generated for metabolite standards of guanosine, thiamine, tyrosine and valine to assist with identifying correlated and anticorrelated peaks in the colon cellular data which are suggestive of hypoxic targets.

Cell response to drug treatment has also been explored in multicellular tumour spheroids in a pioneering study by Kotze et al [120]. The mechanistic modes of action of the chemotherapeutic drug doxorubicin were proposed following ToF-SIMS with a
C$_{60}^+$ primary ion beam and image PCA data analysis. Loadings plots were subsequently studied to deduce correlations between cholesterol and diacylglycerols acting as potential drug response markers at low and high drug treatment concentrations.

1.5.1.2 Probing the Biochemistry of Other Diseases with ToF-SIMS

A recent review article by the author highlights the suitability of ToF-SIMS as a powerful tool for the analysis of metabolites and lipids at the cellular and tissue level, which have implications in a wide variety of diseases [121].

1.5.1.3 Metabolomic and Lipidomic Applications of LC-MS

Species ranging from bacteria to algae and mice through to mammals have all been probed with LC-MS and there are thousands of examples in the literature. Key studies in this field are those from the Griffin group, Fiehn group and Dunn group [101, 122-126] which include not only plant and mammalian applications but platforms for improving workflow, data interpretation, reporting and addressing training needs. A noteworthy Nature Protocols paper describes chromatography-mass spectrometry platforms for large scale metabolic profiling, detailing all methodology from sample collection and storage, to preparation and analysis [127].

Lipidomics has been revolutionised by improvements in both LC separation and mass spectrometry resolution which now enable not just the characterisation of cellular lipids species but also an understanding of lipid substrates and products, which affords a more systems biology based analysis [42].

In the field of mammalian metabolomics and lipidomics of disease, cancer appears to attract the most interest for bioanalytical research.

Urine, plasma and cell extracts are all common matrices from which biomarkers can be identified. Wedge et al conducted a study on the comparisons of serum with plasma for inter-subject comparisons in metabolomic studies with the conclusion that neither fluid is more superior than the other but choice is primarily clinic led [128].
A recent study by Shao et al employed a pseudo-targeted LC-MS approach with a triple quadrupole linear ion trap mass spectrometer in the search for urinary biomarkers for the aggressive hepatocellular carcinoma for which early diagnosis is essential to improve survival rates [129]. The study looked at liver cirrhosis as well as liver carcinoma against healthy cells and found specific compounds such as certain amino acids and bile acids that were indicative of disease in the liver; furthermore, potential specific biomarkers, such as butyrylcarnitine, were identified to discriminate between hepatocellular carcinoma and cirrhosis.

Metabolomic profiling of tissue extracts from five pancreatic tumours and matched normal samples were analysed by UPLC-ToF-MS for biomarker discovery [130], one of a few untargeted studies searching for pancreatic cancer biomarkers [131]. Again, the drive for early detection is the fuel for such biomarker discovery. Uridine, succinate and malic acid were amongst those metabolites quantified to be significantly altered in pancreatic tumour tissue, all of which were identified with less than 1 ppm mass error. As with many of these studies, further validation will be required to test the reproducibility in larger cohorts and in this case the identification of a matrix that would be less invasive to the patient would also be desirable.

A non-targeted approach was recently reported by Sjostrom et al to identify potential protein biomarkers for breast cancer dissemination which ultimately led to the validation through a large scale clinical trial of five proteins as potential biomarkers for breast cancer recurrence [132]. This leads into the vast realm of proteomics, another –omics field in which larger biomolecules such as proteins are characterised in an attempt to aid clinical diagnosis and understand disease.

1.5.1.4 The Use of LC-MS in Understanding the Many Facets of Leukaemia

The metabolic features associated with AML have been profiled by LC-MS/MS in the search for potential diagnostic markers of disease response status in serum [133]. It is often a challenge to define complete remission according to morphology and blood cell counts, particularly when bone marrow samples taken from patients are poor quality. In this study, a combined technique of gel electrophoresis and nano-LC-
MS/MS was able to identify two significant proteins, complement factor H (CFH) and apolipoprotein H (ApoH), in the gel bands (resulting from prior separation and analysis) which showed a marked difference between pooled sera from patients who had reached complete remission (group 1) and those who were non-recovered (group 2), with a 2-fold increase in these proteins for patients in the latter group.

Phospholipids were biomolecules under investigation in a recent MS based study which quantified the relative amounts of 12 classes of phospholipids including peroxisome-derived lipids (a special class of phospholipid) in the thymus (mammalian), liver and spleen (mouse) and bone marrow (human) [134]. Lipids were extracted from bone marrow into chloroform-methanol (2:1) and analysed by ESI with MS and MS<sup>n</sup>-detection. Elevated levels of peroxisome-derived lipids were observed in all cases of leukaemia when compared to control which suggested a link to natural killer T (NKT) cells, perhaps in them being stimulated in the presence of leukaemia.

### 1.5.2 Spectroscopy in Metabolomics and Lipidomics

FTIR and Raman microspectroscopy have both developed significantly in the last decade as well utilised tools for probing the nature of biological species such as tissue, cells and serum [135-143]. They are complementary techniques with different selection rules, infrared spectra arising from the absorption of radiation according to functional groups present in a molecular species, and Raman spectra a result of changes in the polarizability of a molecule giving rise to scattering of the incident radiation (further details of which can be found in sections 2.3.1 and 2.4.1 respectively). The techniques are capable of providing a rapid, rich biochemical ‘fingerprint’ of all components of the species being analysed, which on interpretation are extremely informative in both a research and more recently, a diagnostic setting [144-148]. Probing drug-cell interactions with spectroscopic techniques has become increasingly popular and can contribute to the understanding of the mode of action of the drug at a cellular level [149, 150]. To date, biochemical and morphological classification of healthy versus diseased cell lines has been well demonstrated with FTIR and Raman spectroscopy [151, 152].
1.5.2.1 Metabolomic and Lipidomic Applications of Raman Spectroscopy

Despite Raman spectroscopy being widely available since the 1970’s, it is only emerging as a research tool for biological applications in more recent years. Its appeal is widespread as a non-invasive, label-free technique, and has been made more accessible for probing biomolecules by significant advances in computational power and sophisticated data processing software which have enabled the deconvolution of large data sets.

1.5.2.2 Cellular Raman Imaging

Raman spectroscopy has been employed for the dynamic characterisation and imaging of a wide range of single biological cells such as stem cells, bacterial cells and mammalian cancer cells [76]. One of the early applications of Raman spectroscopy for imaging live cells was reported by Hamada et al who imaged the molecular distribution of cytochrome c, protein β-sheets and lipids in the cytosol and nuclei of unstained live cultured cervical cancer (HeLa) cells [153]. Fixed HeLa cells had previously been imaged by Matthaus et al in a study that compared standard fluorescence staining microscopic methods with label-free confocal Raman microspectroscopy for visualising the distribution of mitochondria in cells. The adherent cells were cultured onto CaF$_2$ discs and fixed in 10 % phosphate buffered formalin solution and submerged in phosphate buffered saline for analysis. Hierarchical cluster analysis was performed on the lipid C-H stretching region between 2800 cm$^{-1}$ and 3020 cm$^{-1}$ and in the protein rich spectral range 1200 cm$^{-1}$ to 1800 cm$^{-1}$. Both Raman images obtained showed clear differentiation between cellular compartments which were not visible in the green fluorescent stained images due to the poorer spatial resolution of fluorescence microspectroscopy. Furthermore, the mitochondrial staining matched well with the Raman images of the same region thereby confirming that detailed cellular images can be constructed without sample pre-treatment with labels or dyes [154].

Subsequently, many Raman studies have imaged various cellular and even sub-cellular compartments of biological interest. The work of Schulze et al, reported to be the first study to specifically address Raman imaging of the nucleolus and its substructures,
focussed on imaging changes in the size, number and chemical composition of nucleoli as an indicator of malignancy in rapidly proliferating human breast cancer cells. As in previous Raman studies, PCA enhanced the interpretation of data obtained and allowed identification and mapping of the distribution of intranucleolar RNA [155].

1.5.2.3 Raman Spectroscopy in the Study of Cancer

Raman analysis of breast and prostate cancer cells and tissue are amongst the most common cancers reported in the literature, presumably due to the high prevalence of these cancers (breast cancer being the most commonly diagnosed cancer and the leading cause of cancer death in women worldwide and prostate cancer the second most frequently diagnosed cancer in men worldwide according to information complied it 2008 [156]) and also the ease of availability of in vitro cell lines for them. More than a decade ago, a pioneering study by Crow et al reported the use of Raman spectroscopy to categorise four different prostate cancer cell lines of varying degrees of differentiation [157]. PCA enabled the study of biological differences between cell lines, whilst linear discriminant analysis provided a diagnostic algorithm for cell classification with 98% sensitivity and 99% specificity.

Within the category of breast and prostate cancer (amongst others), biomarker characterisation and identification is a popular area of research for cancer diagnostics.

Raman spectral markers for breast cancer metastases in bone were explored by Ding et al [158]. Bones containing tumours were analysed for local changes in bone tissue that could indicate a change in bone composition leading to osteolysis (bone loss) and fractures. Peak ratios for phosphate and amide I stretch, phosphate and proline and carbonate and phosphate were calculated for 14 control and 30 tumour-bearing mouse tibia bones. Results suggested that the carbonate/phosphate parameter could potentially be used as a spectral biomarker since elevated carbonate levels have been reported in aged and metastatic bones and Raman data correlated well with this observation.

Lipid content in both breast and prostate cells upon treatment with synthetic hormones (MPA and androgen R1881) has been studied with micro-Raman and coherent anti-Stokes Raman scattering [159]. This offered a non-destructive alternative to that of the
more commonly used LC-MS, as well as providing spatial information on the distribution of lipids in formaldehyde fixed cells. Hormone treatment was found to stimulate lipid droplet formation and the composition of cellular lipids was well characterised according to the degree of saturation and chain length. The previously reported ratio of total unsaturated fatty acids at approximately 3015 cm\(^{-1}\) (TUFA), to the total amount of fatty acids at 2851 cm\(^{-1}\) (TFA) has been proposed as a molecular marker for healthy and tumour cells and in this study the TUFA/TFA indicated an increase in saturated lipid content after hormone treatment.

A more specialised application of Raman to breast cancer monitored the epithelial-to-mesenchymal transition (EMT) which is promoted by the activation of hypoxia-inducible factor (HIF) and indicates the lethal metastatic phenotype of breast cancer [160]. The scores plots generated by multivariate curve resolution (MCR) showed EMT to have an inverse correlation with keratin (5 bands in the fingerprint region) and proportional correlation with cell lipid content. Data also suggest that tryptophan shows a marked increase in the metabolic profile of EMT. The differences in fatty acid and lipid profiles observed between benign and highly metastatic cell lines analysed was of no surprise given previous breast cancer studies of tissue [161].

Most recently, Talari et al employed dispersive Raman spectroscopy to differentiate two breast cancer (MCF-7 and MDA-MB-436) and one healthy breast cell lines using PCA to identify variations between cell groups [151]. Interestingly, in this study, unusual cell preparation methodology was undertaken in which cells were prepared in agar plugs and fixed in formaldehyde before being sliced with a microtome prior to analysis. As well as considering the entire fingerprint region, PCA was performed across regions of specific biological interest: lipids between 3100 and 2680 cm\(^{-1}\), the amide I between 1800 and 1530 cm\(^{-1}\) and amide III between 1380 and 1190 cm\(^{-1}\) to specifically investigate these biomolecules. Differences in the concentrations of a variety of lipids, proteins and nucleic acids were observed between the cell lines and as such this work could potentially pave the way for determining new spectral biomarkers to differentiate types of breast cancer. In the same year, Abramczyk et al also employed the MCF-7 breast cancer cell line along with two other cell lines of differing malignancy to investigate the role of lipid droplets in cancer by comparing lipids in live cells in PBS; with those observed in adipocytes in malignant human breast tissue [162]. Lipid droplets have important roles in cellular metabolism and are widely
associated with cancer development, however the functional implications of these organelles is not fully understood. Raman microspectroscopy and imaging showed different chemical compositions between lipid droplets observed in epithelial breast cells and lipids in breast adipocytes suggesting they are formed via different biosynthetic routes. The Raman spectra of adipocytes were dominated by triglycerides of oleic and linoleic acid whilst arachidonic acid peaks were more prevalent in the lipid droplets of cell lines. There was an overall increase in lipid droplets with increasing malignancy of the cell lines which correlates with enhanced lipogenesis in cancerous cells and tissue.

In the study of drug-cell interactions, predication of chemotherapeutic response has been reported for a model system employing the well characterised anti-cancer drug cisplatin in lung adenocarcinoma cells [163]. Confocal Raman microspectroscopy was used to collect spectra for different cellular regions, the membrane, cytoplasm and nucleus for cells treated with increasing concentrations of drug. Regional spectral differences could be distinguished which were attributed to the chemical effect of drug in the cell and data were used to ultimately build regression models which could be employed in a predictive capacity. Prediction for regression against cisplatin dose was higher for individual regions of the cell rather than for data combined across the whole cell which demonstrates clear advantages of the high spatial resolution of this technique.

There are few examples in the literature of the use of Raman spectroscopy in the field of acute myeloid leukaemia. Cai et al report a fairly limited study in which HL60 cells are differentiated from normal human peripheral blood mononuclear cells by confocal micro-Raman spectroscopy according to averaged spectra [164]. Spectra were analysed in the region 500 – 2000 cm\(^{-1}\) and major differences between the two cell classes arose from total amount of nucleic acid, protein and other biomolecules.

Another cell differentiation study was applied to AML and myelodysplastic syndrome, which characterised cells from AML patients through Raman imaging and hierarchical cluster analysis [165]. This was the first time Raman had been applied to human AML cells and cellular morphology was well described by images in the fingerprint region (500 – 1800 cm\(^{-1}\)) allowing identification of myeloblasts, promyelocytes, abnormal promyelocytes and erythroblasts. Furthermore, the mean spectrum of each cell enabled
classification of cells into different subpopulations according to PCA, proving the methodology to be successful in characterising typical cells of haematopoietic tumours.

In perhaps the most interesting application of Raman spectroscopy within the context of this thesis, nuclei were extracted from AML HL60 cells which had been treated with two chemotherapeutic drugs, doxorubicin and vinblastine, to stimulate apoptosis [166]. Previous reports of whole cells studied with Raman spectroscopy indicated that changes at a nuclear level could not be observed, possibly due to the interference of other cellular cytoplasmic biomolecules. Employing the small spot size (approx. 1 µm) of a focussed 514 nm laser allowed spectra to be acquired for individual nuclei which were suspended in PBS during the analysis. Several nucleic acid bands were observed in the nuclei spectra, showing an early response to cytotoxic changes with drug treatment. This study demonstrated a more sensitive and rapid detection of cytotoxicity than conventional cytotoxicity assays commonly employed.

1.5.2.4 Raman Spectroscopy Translation into the Clinic

The ability of Raman spectroscopy to detect small biochemical changes in cells and tissue gives the technique scope to translate into the clinic for detecting cancerous and even pre-cancerous developments in patients.

Through numerous studies of in vitro, in vivo and ex vivo (extracted from patient) cervical cancer samples, Raman spectroscopy has advanced as a diagnostic tool for this disease [167]. Widespread applications are discussed by Ramos et al, who conclude that for advances in this field to continue, a greater body of supporting evidence is required. This would include standardisation in reporting data and increased number for cervical cancer clinical studies.

A range of cellular in vitro applications including therapeutics development, cell classification and live single cell imaging for prostate cancer cells are explored by Kast et al [145]. The review also extends to emerging clinical applications of Raman spectroscopy to prostate cancer screening, biopsy and tumour margin assessment.
In vivo Raman imaging is also explored in a review by Ellis et al, of particular interest are the developments in Raman endoscopy and Old et al discuss the clinical applications for a variety of cancers from colorectal cancer to lymphoma in a recent review [168]. Raman fibre-optic endoscopes are a powerful emerging technology to probe internal organs for malignancy or early stage changes, which could significantly reduce time spent by histopathologists for tissue diagnosis and decrease time for the patient from biopsy to results. Stone et al were one of the first groups to pioneer the fusion of Raman spectroscopy and clinical endoscopy in 2000 and first reported on the non-invasive discrimination of healthy and cancerous laryngeal tissue in vivo [169].

Since then, they and others, have continued to push the boundaries for such applications forward. It is an exciting and fast moving field, and such developments are essential if spectroscopic techniques are to be wholeheartedly taken up in the clinic. A more up to date review of medical diagnostics suggests that the accuracy and speed of the technique is now directly applicable to the clinic [170]. A number of challenges do, however, remain and at the forefront of these is to demonstrate the safety to both clinicians and patients through large-scale studies [168]. Krafft and Popp write openly about actively encouraging scientists to be early adopters and follow in the lead of the pioneering few stating that such studies are no less essential than the first ones establishing a method [135].

Despite promising developments for in vivo Raman spectroscopy in the clinic, there still remains a need for analysis of cells in culture, particularly in aiding the understanding of the spatial dynamics of biomolecules responsible for disease and in evaluating the action of drugs on cells. Spectropathology for the next generation has been recently discussed by Byrne et al, in which in vivo, ex-vivo and in vitro cell culture models are all highlighted as key areas for continuing development in the future for translation of spectroscopy to the clinic [171].

Schie et al have recently developed a line-scan Raman microscope to study the effects of the chemotherapy drug doxorubicin on living single lymphocytes over 24 to 96 hours [172]. This noteworthy study addresses the assumption made in many single cell imaging studies, that cellular changes occur at a local level, i.e. inside the cell, and whilst such sub-cellular analysis has proven successful for many applications, the total molecular content of biochemical species across a whole cell is often missed. Quantitative analysis revealed that biochemical changes were occurring at both a sub-
cellular level and a global level (total molecular cell content). This can provide additional, and potentially critical information about changes in a cell population happening at a global level and needs to be considered in future studies.

The use of 3D cell cultures to study in vitro cellular models is becoming more widespread due to their improved physiological relevance when compared to standard 2D cell cultures [173]. This can pose a problem for existing analytical techniques that have previously been successfully employed to probe cells in 2D cultures. Early work in this field suggest Raman microspectroscopy to be effective when spectra acquired at different penetration depths in spheroids of breast epithelial tumour cells all showed the same pattern, only varying in reduction of peak height with deeper penetration. Further investigations with Raman of chemically defined hydrogels provide a platform for 3D studies of tumour-stroma interactions [174].

1.5.2.5 Metabolomic and Lipidomic Applications of FTIR Spectroscopy

In comparison to other spectroscopic approaches and when compared to mass spectrometric techniques, FTIR is relatively cost effective and lends itself nicely to a rapid, initial screening method before further analytical method development is required. It can also be employed as a highly sophisticated tool for the chemical imaging of single cells and for well over half a century, IR spectroscopy coupled to microscopy has been recognised as a non-destructive and label free sensitive analytical method with great potential in biomedical research [175]. The emergence of commercial Fourier transform infrared spectrometers in the 1970’s and subsequent availability of commercial infrared microscopes in the 1980’s pioneered the way for a multitude of developments since then [58].

In 1997 an exceptionally detailed review was published, reporting the IR spectra of proteins and peptides in lipid bilayers which detailed specific vibration frequencies for the hydrophobic and interfacial regions of the lipid bilayer [176]. A vast amount of information on pure lipid systems such as vesicles, monolayers and lipid bilayers was available at that time and since then, a multitude of studies have further probed not just the cell membrane but sub-cellular compartments too.
More recently, FTIR has become widely employed in disease diagnosis and in monitoring drug-cell interactions and, not surprisingly, cancer research and diagnosis are at the forefront of these applications [56, 136, 148].

Baker et al have published a Nature Protocol article to encourage the standardisation of methods and procedures for collecting IR spectra and images from a range of biological samples from fixed cells to biofluids [138]. Such reports are essential in establishing robust protocols that are widely accepted in the field to assist in the adoption of IR into the heart of biomedical and clinical disciplines. Multivariate data processing techniques are also addressed, these being a large source of variation amongst current spectroscopists.

Wood et al are at the forefront of biomedical spectroscopy and have pioneered the diagnosis of malaria with FTIR in recent years. In one interesting study, the coupling of a focal plane array IR microscope to a synchrotron light source enabled the imaging and diagnosis of early stage malaria in blood smears based on unique lipid signatures of different malarial stages [147].

Identifying drug mode of action at a cellular level is of much interest to the scientific community and a recent review considers major contributions to this field for cancer cells and anti-cancer drugs [177]. The effect of anti-cancer drugs on seven cell lines as identified by FTIR was reported by Derenne et al and stimulated further spectroscopic studies [150]. One such study considers not only the effect of the drug on the cell, but also on the cell cycle which is an important question in studies of large populations of cell [178]. The outcome was that the overall effect of paclitaxel on prostate cancer cells (PC-3) was the dominating factor in the spectra and spectral changes arising from different stages of the cell cycle were negligible in comparison.

In another cell-drug interaction study, Kim et al probed the phenotypic response of cervical carcinoma cells exposed to the HIV anti-viral drug lopinavir [179]. Both the metabolic fingerprint from direct cellular analysis and the metabolic footprint from analysis of the growth medium were analysed to determine cellular responses to the drug. It is clear that such studies are of great value in understanding how a drug works inside the cell.
1.5.2.6 FTIR Applications for the Study of Leukaemia

In early spectroscopic leukaemia cell studies, the biochemical nature of apoptosis in 2 different leukaemic cell lines (K562 and CML) was investigated by Liu et al [180]. Cells induced with etoposide to induce programmed cell death were sampled at increasing time points for analysis of the dried cell pellet by IR. Spectra revealed a number of cellular changes including a shift in protein structure from beta sheet to random coil and an increase in lipid content. When the results obtained by IR were compared to those from traditionally used methods such as flow cytometry; IR was able to observe cellular changes at a much earlier stage in the apoptotic process (4 hours vs 6 hours). Zhou et al from the same group reported one of the early uses of FTIR in a similar study with HL60 cells which were induced by different drugs, one known to stimulate differentiation and the other apoptosis.

Apoptosis of HL60 cells was also investigated by Gasparri et al, this time with attenuated total reflection (ATR) FTIR as the sampling mode [181]. They reported a number of spectral changes with apoptosis progression, particularly in the nucleic acid region of the spectra and thus suggest that ATR-FTIR could be employed as a diagnostic tool for cell apoptosis.

More recently, Gaigneaux et al employed FTIR for cell discrimination of drug sensitive and multiresistant K562 cells. Daunorubicin-sensitive and resistant cells were able to be discriminated from each other by genetic algorithm or PCA analysis with differences arising from cellular lipid, protein and nucleic acid contents [182]. This study demonstrated the suitability of FTIR for classification of cells, following which the same group later went on to study cell discrimination with ATR-FTIR [183]. The biological principle was essentially the same as the previous study, but in this case emphasis was on the effect of different pre-processing methods on first and second derivative spectra. Neither study focussed on the biological interpretation of data in context of AML.

In a potential clinical application of cell classification, FTIR was employed to discriminate between T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cells [184]. Subtle spectral differences in biomolecular vibrations were observed between the different cell types and PCA showed spectral grouping for each of 5 cell lines. Classification results confirmed that spectral signatures from each cell line could
provide good classification with 74-100% sensitivity and 94-100% specificity. This suggests that this methodology could contribute to the development of a diagnostic algorithm for the future classification of haematological malignancies from blood and bone marrow clinical samples.

Bellisola et al have demonstrated the use of FTIR microspectroscopy for quantitative pre-diagnostic screening of drugs whose targets are specific molecular pathways that cause chronic myeloid leukaemia (CML) [56]. The rationale for this study was that CML patients often develop drug resistance due to increased phosphorylation of tyrosine which causes CML blasts to acquire growth factor independence and thus sustain proliferation. In this condition, DNA repair is also inhibited which causes genomic instability that further facilitates drug resistance. Spectra were acquired for formalin fixed single cells using Synchrotron-FTIR microspectroscopy and for a monolayer of approximately 15 cells using a globar source and IR data was cross-validated with western blotting analysis of untreated and drug treated cells. The focus of data reported here was on phosphorylated proteins within the cell and does not address other classes of biomolecules such as lipids. IR signatures of drugs, and of apoptosis, were observed and confirmed with standard molecular biology procedures, however it was acknowledged that a lot of research must be done (for example in compiling validated spectral libraries of drug signatures) for this methodology to be more routinely used in clinical diagnostics.

1.5.2.7 Considerations in FTIR Studies

FTIR analysis and imaging of biological samples, in particular cells, whilst proven to be a highly useful and powerful tool in cytology, is not without its problems. For the technique to be of profound clinical use, a pure absorption spectrum that is representative of the cellular biochemical composition alone needs to be recovered from a raw spectrum. The raw spectrum however is influenced not only by the biochemical composition of the cell but also by physical properties such as size, morphology and inhomogeneous optical density. As an example, the nucleus of a cell has been shown to exhibit non-Beer Lambert absorption which can cause mis-assignment of specific bands [185]. The term Mie-type scattering is given to this phenomenon which takes place when the wavelength of incident radiation is similar to
or substantially less than the size of the scattering particle. As a result of these scattering effects, a broad sinusoidal oscillation visually distorts the baseline of the spectrum which can affect both the intensity and position of bands, rendering interpretation of spectra challenging and sometimes unreliable [186].

Bassan et al in have developed an employed correction algorithm called the resonant Mie scattering extended multiplicative signal correction (RMieS-EMSC) which can be used to remove the baseline distortions in cells [187, 188]. RMieS correction was extended to tissue image analysis by Bambery et al who demonstrated that previously hidden spatial information on the biochemical contents of cervical biopsy and rat brain tissue could be revealed with removal of baseline distortions [189].

1.5.3 Applications of Multimodal Analytical Techniques to Metabolomics and Lipidomics

Despite the widespread application and increasing popularity of LC-MS, ToF-SIMS, Raman spectroscopy and FTIR spectroscopy for bioanalytical studies previously discussed, there are few examples in the literature of studies employing two or more of them. This is surprising given the potential range and power of data that could be obtained, for example, if the mode of action of a drug at a cellular level is the question, such a multi-modal analytical approach can only offer more pieces to completing the full picture. Metabolomic and lipidomic examples discussed in this section extend beyond cellular systems to encompass tissue and bacteria biochemistry due to the lack of multimodal applications to cell analysis reported in the literature. Such examples have been included to demonstrate the added value of employing a multi-modal approach in bioanalytical research.

An obvious choice of combining analytical approaches is with the complementary spectroscopic techniques of FTIR and Raman and there are a number of review articles on the topic of spectroscopy in cancer research and diagnosis which address both Raman spectroscopy and FTIR [148]. Of particular interest is that by Kendall at al who are keen to illustrate the advantages of using Raman and FTIR spectroscopies together [136].
Kast *et al* in reviewing the role of Raman spectroscopy in prostate cancer, explore the context of a multimodal system stating ‘it is needed to fully analyse subcellular components and to identify the ensemble effects that make up the cell and its function’, which the author feels is a sensible and realistic statement [145]. However, Kast goes on to say that ‘the system should also be non-destructive and provide comprehensive analysis to investigate viable live cells with little or no confounding interactive ligands or chemicals’. Which of course is desirable but such analysis would then be limited to non-destructive analytical techniques and the author feels this is overlooking the potential for such techniques to contribute holistically to the overall system. In this review, ToF-SIMS is brushed over for being disruptive and more discussion is given to laser-induced fluorescence and electrochemical methods which have limited selectivity.

The application of SIMS and Raman to tissue engineering studies have recently been reported [190]. Being ‘label free’, these techniques offer specific advantages for tissue engineering in that they are location specific approaches with the ability to contribute to efforts in developing artificial scaffolds for tissue regeneration through identifying substrate composition and the fate of individual cells within microscale culture platforms. However, this review handles the two techniques very separately and does not specifically address the overall benefits of using the two combined in one study.

A study that does exploit the combined strength of these two techniques investigates myelin membrane reorganisation initiated by extracellular ATP and adenosine in sciatic nerves of frogs [98]. The combination of the two methodologies allowed live cells to be analysed with Raman spectroscopy to study the overall effect of the applied stimulants and fixed cells to be analysed by ToF-SIMS to provide more detailed lipid composition information. Lipid standards were also prepared and analysed as films by ToF-SIMS to explore matrix effects and allow semi-quantitative data interpretation. The study concluded that it was difficult to propose robust interpretations of cellular lipid mechanisms with data from these two techniques alone, however results were important in beginning to piece together the mechanisms of peripheral myelin structure regulation for the first time.

A recent investigation of multimodal chemical imaging of molecular messengers in bacterial communities also employed confocal Raman microspectroscopy and SIMS
for *in situ* characterisation of metabolites and signalling molecules in bacterial communities [191]. Classification of microbial communities was possible with PCA image analysis of Raman data and SIMS images added to the biochemical information and spatial distribution of several key sensing molecules and metabolites across the surface of bacteria.

Utilising the complementary assets of ToF-SIMS and LC-MS facilitates a comprehensive understanding of a cellular system. In a study employing C$_{60}$ as the primary ion beam and tandem MS as the detector, over 50 lipids were identified in mouse leukaemic monocyte macrophage cell line *in situ* [175]. LC-MS analyses previously carried out and reported by the LIPID MAPS consortium [47] was used as a comparison of both data sets. In excess of 400 major lipid species were detected by LC-MS, a much larger number than with ToF-SIMS, most likely due to the greater signal to noise and reduced matrix effects of LC-MS technique over ToF-SIMS. However, lipids identified in both techniques showed excellent correlation and because of the agreement between two data sets, a calibration curve was plotted for the LC-MS data and compared to that obtained with SIMS. This provided evidence for the ‘quasi-quantitative’ potential for ToF-SIMS, but also illustrates the synergistic effect of employing these two techniques together. Unfortunately, in this example, the spatial resolution of SIMS was not employed, however in another study, imaging ToF-SIMS was combined with LC-MS to localise and quantify benzalkonium chloride (a preservative in eye drops) at the surface of rabbit eye tissue sections, and subsequently human samples [192]. This work exemplifies the quantitative ability of LC-MS and the high spatial resolution of ToF-SIMS to complete a wider picture of biochemistry occurring in tissue and suggests that these complementary techniques could be used to correlate the amount of xenobiotics with its localisation in tissue sections in further applications.

Synchrotron FTIR and ToF-SIMS were employed to characterise the biochemical composition of fatty liver (steatosis) in human liver tissue sections [99]. Spectra obtained with mid-IR synchrotron radiation showed a marked difference between fatty and non-fatty hepatocytes in that the amide I and II bands were absent in the fatty vesicles. Significant changes were also observed in the high wavenumber (3000-2800 cm$^{-1}$) lipid region. ToF-SIMS was therefore chosen as a useful technique to further explore the changes in lipid composition at the surface of tissue sections, without the
need for matrix deposition. The study was successful, with SIMS ion image overlays showing different localisation of diacylglycerols outside and inside fatty vesicles. The study concluded that a major benefit of ‘combining IR with MS is the possibility to establish a link between IR spectra and molecular composition of analytes’. This is indeed of great interest and shall be further explored in this thesis.

1.6 Acute Myeloid Leukaemia

1.6.1 Acute Myeloid Leukaemia Characterisation and Prognosis

Acute myeloid leukaemia (AML) is an aggressive cancer that causes life-threatening deficits of functional blood cells in humans. It is characterised by the uncontrolled proliferation, abnormal survival and maturation arrest of malignant cells that causes depletion of erythrocytes, platelets and neutrophils [193]. If untreated, patients can die within weeks of diagnosis and unfortunately the response to treatment and overall survival of patients post-treatment generally remains poor [194]. AML can occur in all ages but is most common in older patients with more than 75% of AML patients being over the age of 60 years when diagnosed. Unfortunately overall survival decreases with increasing age which contributes to unfavourable diagnostic outcomes in this particular cohort of patients [195], [196]. Figure 1.6 shows the average number of new cases of leukaemia per year reported by Cancer Research UK.
Despite the outlook for patients improving in the past 30 years (primarily due to improvements in supportive treatments, rather than the emergence of successful new therapies), around 90% of older patients still die from the disease [198, 199]. The extreme heterogeneity of AML is now widely recognised and patient treatment is beginning to move away from a ‘one fits all’ approach in a bid to improve clinical outcomes, for example by offering older patients access to potential new drugs via clinical trials [200].

AML is clinically characterised by a large amount of blast cells, which are morphologically immature blood cells in different stages of growth when compared to healthy cells present in the blood. This abnormal proliferation and differentiation of haematopoietic cells causes the accumulation of lymphoid or myeloid progenitors in the bone marrow and subsequently hinders the production of healthy blood cells, such as lymphocytes, which play an essential role in the bodies immunity. AML can also affect other organs of the body such as the liver and spleen, where white blood cells are also found.

The World Health Organisation (WHO) have classified AML according to subtypes to include factors such as genetic abnormalities, because there are many variables
which affect the prognosis and therefore the recommended treatment for one subtype of AML is often very different from another [201].

The human body makes a variety of blood cells in the bone marrow from stem cells as depicted in Figure 1.7.

Figure 1.7 Overview of the development of mature blood cells from haematopoietic stem cell.
1.6.2 Treatment of AML

Current treatments for AML are typically based upon a blood or platelet transfusion (where deemed necessary), followed by induction chemotherapy (initial chemotherapy administered to destroy leukaemia cells) which is highly cytotoxic and directly further reduces haemopoiesis (the formation of blood cells) in the patient. Frequently a drug called cytarazine (cytosine arabinocide) is administered with an anthracycline drug such as daunorubicin. The aggressive nature of this treatment is most poorly tolerated in the aforementioned elderly cohort of patients and whilst induction therapy can lead to remission in many patients, relapse is very common, with more than three quarters of elderly patients experiencing relapses after complete remission [202]. A study in 2015 highlights the need for diagnostic tools to identify residual AML in patients and presents data supporting the hypothesis that circulating extracellular vesicles can act as malignancy biomarkers reflecting changes in cell populations for AML patients in remission [203]. Whilst early indicators of the onset of disease are always of clinical benefit, alone they are not enough.

There is therefore an urgent clinical need for novel therapeutic approaches to the treatment of both primary AML and for the elimination of residual leukaemia in remission. In the 1990’s Pizer et al reported a study on highly proliferating AML HL60 cells which express high levels of FAS and synthesise fatty acid predominantly for membrane phospholipids [204]. HL60 cells provide a useful model for studying both activity of fatty acid synthesis and cellular adaptation to fatty acid synthesis inhibition. The study exposed these cells to an ester compound (12-O-tetradecanoylphorbol-13-acetate) upon which cells differentiated into macrophages, losing the ability to proliferate. Experimental data showed that fatty acid and phospholipid synthesis rapidly decreased in HL60 cells post exposure and continued to decrease during differentiation. This process involved downregulation of the fatty acid synthesis pathway which was demonstrated at levels of mRNA, protein expression and chemical activity. HL60 cells were also treated with a known FAS inhibitor, cerulenin, to allow assessment of cerulenin sensitivity in transformed and non-transformed in vitro HL60 cells. Cells of the differentiated phenotype acquired resistance to FAS inhibition by cerulenin indicating that cancer cells with an elevated expression of FAS have greater sensitivity to FAS inhibiton than in non-transformed fibroblasts which contain lower
levels of fatty acids [204]. Whilst results looked promising for paving the way for new treatments, progress to date has had little impact in the clinic.

New approaches are currently the focus of much research; a number of clinical trials have employed varying combinations of conventional chemotherapies but report little progress in offering significant benefits over the current recommended treatment [194]. One of the most recent developments in AML chemotherapy is the discovery of 2 new drugs which are both cytidine nucleoside analogues with DNA-hypomethylating properties; azacitidine and decitabine, the latter of which has recently been registered for AML treatment in Europe [205]. The results of an international randomised stage 3 clinical trial for azacitidine in elderly patients (median age 75 years) with newly diagnosed AML presenting with >30% bone marrow blasts were published in 2015. Azacitidine was observed to increase median overall survival by 3.8 months when compared with current common AML treatments and was generally well tolerated suggesting this as another treatment option for the elderly cohort of patients [206]. Both drugs, however, have a mode of action via incorporation into RNA and DNA (azacitidine) or just DNA (decitabine) which at certain prescribed doses crosses into cell cytotoxicity territory and inevitably gives rise to unwanted side effects [207, 208]. Many other treatments in current development are focussed around gene therapy, since genetics has been strongly implicated in the success of a patient’s treatment and overall patient prognosis. The drive towards target therapies is a challenge because of the inherent heterogeneity of AML both from a genetic complexity but also across a wide range of patients. The scientific community has been striving to further understand the molecular pathogenesis of AML and a recent proteomics review summarises developments in profiling the array of proteins present in cells, tissue or body fluids and their links to genetic mutations for the development of diagnostics and therapeutics in myeloid leukaemia [209]. A study in late 2015 has investigated specific mutations in AML associated oncogenes that often arise in haematopoietic stem cells and promote leukaemia stem cell phenotypes [210]. Their results suggest that myeloid differentiation is a prerequisite for leukaemia stem cell formation and subsequent AML development, a breakthrough which may pave the way for therapeutic developments.
1.6.3 Drug Redeployment

Drug redeployment represents an alternative to conventional chemotherapy and involves the use of existing drugs for medical conditions they were not originally designed for. There have been promising examples of this, such as the use of thalidomide in myeloma [211], and the potential for using neuropsychological drugs (specifically, neurotransmission modulators) in modulating brain cancer [212]. Few studies, however, have combined the pharmacological action of two existing drugs to give a greater effect from the drugs working synergistically than is observed from the individual drugs alone. Furthermore, by using a combination of drugs to disrupt more than one biochemical pathway in the cell, there is a vastly more complex work-around required for the cancer cell to evolve and proliferate.

1.6.4 Bezafibrate and Medroxyprogesterone Acetate (BaP)

In 2009, a novel study reported the in-vitro anti-leukaemic effect of a combination of the cholesterol lowering drug, bezafibrate (BEZ) and the female contraceptive medroxyprogesterone acetate (MPA) on a diverse range of AML cell lines and primary AML cells [213]. The combination therapy is denoted BaP. Figure 1.8 shows the structures of BEZ and MPA.
The anti-tumour activity of these drugs individually [214, 215], and combined [216] has been previously reported and is in part thought to be brought about by promotion of the endogenous synthesis of prostaglandin D$_2$ (PGD$_2$) and its potently antineoplastic derivative 15-deoxy prostaglandin J$_2$ (15d-PGJ$_2$). These prostaglandins contain unsaturated ketone moieties which allow for a wide range of potentially potent covalent modifications of cellular targets, one of which gives anti-tumour activity. However, the anti-tumour activities of these two drugs are complex and experimental results have suggested that actions of reactive oxygen species (ROS, which are generated in cells under oxidative stress) and activation of the lipid peroxidation pathway are also involved [213]. NMR was employed to further probe the metabolomic changes occurring with combined MPA and BEZ treatment across 3 AML cell lines. Amino acids valine, leucine and isoleucine were identified to be upregulated with drug treatment and tricarboxylic acid cycle intermediates, in particular the conversion of α-ketoglutarate to succinate and pyruvate into acetate were implicated. This suggested that combined MPA and BEZ action on AML cell lines was mediated downstream of the generation of ROS and that changes directly induced by ROS significantly contribute to the mechanism of action. The combined impact of the drugs were likely to influence broader aspects of mitochondrial function which was

---

Figure 1.8 Structures of bezafibrate (A) and medroxyprogesterone acetate (B).
implicated with a noticeable accumulation of metabolites essential for the biosynthesis of nucleotides, specifically glutamine and glycine, which could indicate hindered biosynthesis of purine and pyrimidine nucleotides [217].

Despite no clear understanding of the mechanisms behind the anti-leukaemic properties of combined BEZ and MPA, the fact that these two drugs had prior regulatory safety approval was exploited. Rapidly following on from the promising in vitro results a clinical study was performed to assess patient tolerance to, and in vivo performance of, the combination of BEZ and MPA, a combination now denoted BaP [193]. 20 patients who had been diagnosed with de novo secondary or relapsed AML (19 patients) or high risk myelodysplasia (1 patient), and for whom intensive chemotherapy had been ruled out, were recruited. The study demonstrated the safety of BaP in elderly patients (the median age of patient being 75 years) and also demonstrated both anti-AML activity and improved haemopoiesis in some of the patients. BaP treatment showed no haematological toxicity against non-cancer stem cells when administered continually and thus could pave the way for future AML therapy. Data did indicate that the daily doses administered to patients were lower than they could have been, and subsequently optimised drug concentrations have been reported and a further clinical trial of BaP at these higher concentrations is planned.

Recently, in 2015, the analysis by DIMS and protein expression measurements of key lipogenic enzymes of AML cells extracts was reported. The findings indicate that BaP treatment down-regulates fatty acid and phospholipid synthesis from C-13 labelled D-glucose. Moreover, a decrease in two key enzymes in the lipogenesis pathway, stearoyl CoA desaturase 1 (SCD1) and fatty acid synthase was observed across all cell lines measured. Cells were subsequently rescued with the addition of oleate, an SCD1 product, but not with the addition of palmitate, an SCD1 substrate and FAS product. This suggests that that one mode of action of BaP is via the lipogenesis pathway, specifically targeting SCD1 and the synthesis of monounsaturated fatty acids [218]. This research infers that further lipidomic studies could be useful in understanding BaP effects.
1.7 Project Aims

Metabolomic and lipidomic studies can yield powerful biochemical data at the cellular level, giving a snapshot view of the phenotype of a cell population under any given perturbation. From a clinical perspective, there shows clear promise for widespread devolution of BaP in the treatment of AML for the more difficult cohort of elderly patients. Further in vitro laboratory studies to understand the specific molecular targets of BaP and understand pathways which are affected at a cellular level would therefore be useful and could potentially pave the way for future AML therapies.

The author hypothesises that the effects of abiotic stress induced by BaP treatment in AML cells can be explored using a multi-modal analytical platform. This hypothesis led to a cross-disciplinary collaboration between the author at the University of Manchester and Andrew Southam and Farhat Khanim in the School of Biological Sciences at the University of Birmingham and forms the basis of this PhD thesis. The initial formalised collaboration plan can be seen in Appendix I.

The aim of this project is to employ FTIR, Raman spectroscopy, ToF-SIMS and UHPLC-MS to yield broad sets of complementary data which will shed new light on the mode of action of BaP on a cellular level. Applying these analytical techniques will provide novel insight into this new treatment protocol for which confidence in biological interpretation can be gained from comparing results across orthogonal analytical platforms. Information obtained will be qualitative, semi-quantitative (changes relative to cell groups) and spatial for which the imaging capabilities of Raman microspectroscopy will be employed where possible. Multivariate methods of data analysis will assist in data reduction and interpretation for the large and novel data sets that will be acquired.

Drug treated and control cells will be analysed in a variety of physical forms including live, fixed and extracted, the diversity of which will contribute a broad picture to drug-cell interaction. Both untargeted and targeted approaches to data acquisition and analysis will be taken to determine biomolecular perturbations at a cellular level that are consistent across both AML cell lines. A consideration of the two cell lines together, which represent different levels of cell maturity, will provide an extra level of confidence in interpretation of the anti-leukaemic activity of BaP.
Chapter 2 Analytical Techniques and Instrumentation

The development of analytical instrumentation for biological applications is under constant evolution. Many techniques have dramatically increased in sensitivity in recent years and continue to do so, with researchers in both academia and industry striving to push forward the boundaries of what can be detected and the chemical information that can be derived from analytes of interest. A meaningful interpretation of the data obtained relies on an understanding of the fundamental principles of the technique employed. This chapter describes the theory of four analytical techniques: ToF-SIMS, LC-MS, Infrared Spectroscopy and Raman Spectroscopy and the specific details of the instrumentation employed for the experiments described in this project.

2.1 Secondary Ion Mass Spectrometry

Secondary Ion Mass Spectrometry (SIMS) is a technique that probes the surface of a sample by bombarding it with a high energy primary ion beam and analysing the sputtered secondary ions by mass spectrometry. It first emerged as a surface analytical technique of some significance in the late 1960’s and early 1970’s when Benninghoven et al studied monomolecular metal substrate surfaces of Mo and Ag [219].

2.1.1 Secondary Ion Formation and Sputtering

The result of bombarding a surface with a high energy beam (typically between 10 and 40 keV of ions or neutrals), is that the particle energy is transferred to atoms of the solid. This occurs in a ‘billiard-ball-type’ collision process which induces a cascade of collisions between the atoms in the solid. [220]. During the cascade, some of the primary ion kinetic energy is transferred to the surface and gives rise to the emission (sputtering) of atoms and molecules. The majority of molecular ions are released from within 2-10 Å of the sample surface [221]. Ions and neutrals can be released from the surface as soon as their kinetic energy exceeds the binding energy to the substrate, and they can be released as single atoms, molecules or a large cluster [104]. The ions that are released in this sputtering process are subsequently detected by a mass spectrometer. Figure 2.1 shows a schematic of the sputtering process.
The sputtering process in ToF-SIMS. The depth of penetration of the primary ions and (to a lesser extent) the information depth from which molecular ions are derived is dependent on the energy and incident angle of the primary ions, and the nature of the sample. Typically, in biological ToF-SIMS the information depth is <10 Å meaning that analysis is specific to the cell/tissue surface exposed to the analysis.

Whilst a seemingly simple process, there are many factors that influence sputtering. In the last 10-15 years, a number of groups have utilised molecular dynamics simulations to understand and optimise the sputtering process [104, 220-224].

Even though the SIMS technique is inherently destructive, Benninghoven demonstrated that changes in surface composition, such as chemical reactions at the surface, could be observed by using a very low primary ion current density which increased the lifetime of the surface and therefore allowed mass spectrometric detection in the same timescale. This methodology is referred to as static SIMS as it is a true surface technique. Benninghoven also went on to demonstrate the utilisation of static SIMS in studying the oxidation of metals at the surface [225] and Vickerman pioneered many studies at this time to further demonstrate static SIMS emerging as a true surface technique [226, 227].
2.1.2 Static SIMS

The widest application of SIMS pre-1980’s was to analyse the elemental composition of a material as a function of depth [220]. This technique is referred to as Dynamic SIMS which today is extensively employed in the semi-conductor industry. In dynamic SIMS using atomic primary ions e.g. Cs+, the high primary ion flux induces significant surface damage, rendering this a fairly destructive technique which is therefore typically limited to the detection of elemental ions and some small fragment ions.

Static SIMS has seen much development in biological applications which were pioneered by groups such as Winograd et al at Penn State University. The emergence of such biological applications was largely driven by instrument developments in the last decade in which cluster primary ion beams (Au\textsuperscript{n+}, Bi\textsuperscript{n+}, C\textsubscript{60}+) have been employed instead of the original atomic ions such as Ga\textsuperscript{+} and Cs\textsuperscript{+}. These cluster ion beams are capable of generating higher secondary yields from molecular materials, particularly higher mass species, [228], and thus increasing the sensitivity of the technique making it more applicable to biological applications. Both the Au\textsuperscript{n+} and C\textsubscript{60}+ systems were developed to extend mass range and ion yield in static SIMS [229] and the aforementioned gold primary ion beam was employed in this study as detailed in section 2.1.6.

2.1.3 The Static Limit

When an atomic primary ion beam impacts the surface of a sample, most molecules within 5-10 nm from the point of primary impact are sputtered and/or fragmented by high energy collisions. In order for the analysis to reflect the nature of the undamaged ‘static’ surface, the sputtering process needs to be controlled so that statistically each primary ion hits at a new area of undamaged surface and ideally less than 1% of the top monolayer of the sample is sputtered. This is achieved by reducing the primary ion dose of the beam to a level where a maximum of one percent of the surface is sputtered. If we assume that the surface density of atoms is typically 1 x 10\textsuperscript{15} atoms/ cm\textsuperscript{2}, a one percent sampling of this would require a dose of primary ions to not exceed 1 x 10\textsuperscript{13} ions /cm\textsuperscript{2}. This value is known as the static limit and in practice to avoid exceeding this, the primary ion beam is usually pulsed in nanoseconds during operation.
The development of the aforementioned cluster ion beams has overcome this problem to a certain extent however, as the larger cluster ions such as $S\text{F}_5^+$ and $C_{60}^+$ leave significant less surface damage in many materials, allowing the static limit to be relaxed or overlooked. This means that it is theoretically possible to use the whole of the surface or even the sub-surface for analysis which is how improved signal and lower limits of detection are achieved.

2.1.4 The SIMS Equation

As previously discussed, one can study the process of sputtering in microscopic detail. However, regardless of the exact mechanisms involved, we can relate the number of secondary ions detected from a surface to both the sample they are emitted from and the experimental conditions which induce this. This means that the yield of secondary ions is heavily influenced by the electronic state of the material being analysed, which is termed the matrix effect in SIMS.

The SIMS equation relates some key parameters involved in this process:

\[
I_s^m = I_p y_m \alpha^{\pm} \theta_m \eta
\]

Where:

- $I_s^m$ is the secondary ion current of species $m$
- $I_p$ is the primary particle flux
- $y_m$ is the sputter yield
- $\alpha^{\pm}$ is the ionisation probability of species $m$ to positive or negative ions
- $\theta_m$ is the fractional concentration of species $m$ in the surface layer
- $\eta$ is the transmission of the analysis system
During SIMS operation, the primary ion flux is under user control and the transmission of the analysis system is fixed for that system. This means that the sputter yield and the ionisation probability of the species are the two main considerations for how a sample will behave during static SIMS analysis. The ionisation probability is strongly dependent upon the electron exchange processes between the emitted species and the surface. Thus the electronic state of the surface is crucial to the success of the experiment, this is the matrix effect previously referred to. Yields of secondary ions vary for elements across the periodic table depending upon the electron affinities for negative ion formation or ionisation potential for positive ion formation.

2.1.5 SIMS Instrumentation

The BioTof is a custom built SIMS instrument in the Surface Analysis Research Centre (SARC) lab at The University of Manchester. It was designed in collaboration with Pennsylvania State University and Kore Technology Ltd (Cambridge, UK) and the original build was described in detail in 1998 [231].

The original BioTof build has since been modified with two primary ion sources, a gold (Au\textsuperscript{+\textsubscript{n}}) gun, previously described by Davies [232] and a fullerene (C\textsubscript{60}\textsuperscript{+}) gun, previously described by Weibel [233]. The availability of two different primary ion sources offer the capability to perform a wide range of SIMS experiments as the two systems offer different performance benefits in terms of spatial resolution, mass range and ion yields [229]. A schematic diagram of the current BioTof system can be seen in Figure 2.2.
The sample to be studied is loaded onto a copper ‘stub’ which is placed onto the end of the transfer arm via the fast entry port. A sorption pump is cooled with liquid nitrogen to assist in the removal of water and to aid rapid pump down of the preparation chamber to vacuum. A gate valve is used to introduce the arm into the prep chamber. When pressures are stabilised, the sample is then transferred into the sample analysis chamber and mounted on the stage directly beneath the ion guns. View ports within the instrumental design allow visual inspection of this sample transfer. Once in place, a video microscope is used to find a region of interest for analysis and a secondary electron detector is available to provide topographic information on the sample.
2.1.5.1 Time of Flight Mass Spectrometry (ToF-MS)

ToF-MS was conceptually devised by Stephens in 1946 [234] and the first design for a ToF-MS instrument was described by Wiley and McLaren in 1955 [235]. It has since become a proven mass spectrometric technique in biological and clinical applications and further details on the evolution of ToF-MS from niche to mainstream can be found in a recent review [236]. It is a technique in which ions are accelerated into a field free drift region (flight tube), all with a very narrow range of kinetic energies (E_k), which are related to the mass and velocities of the ions as detailed in Equation 2.2.

Equation 2.2 Kinetic energy of ions in relation to ion mass and velocity

\[ E_k = \frac{1}{2} m v^2 \]

Where:

- \( E_k \) is the kinetic energy
- \( m \) is the mass of the ion
- \( v \) is the velocity of the ion

The ions are then separated on the basis of their mass to charge ratio (m/z) in the flight tube, with ions possessing the same m/z ratio arriving at the detector at the same time. Heavier ions will drift for a longer time than lighter ions, thus reaching the detector later, allowing ions of different m/z ratios to be separated as illustrated in Figure 2.3.
Figure 2.3 Separation of ions in the drift region of a ToF-MS. Heavier ions take a longer time to drift the fixed distance and therefore arrive at the detector later than lighter ions.

ToF is widely employed in LC-MS and GC-MS where the accurate mass of analytes is required and is ideally suited to SIMS experimentation, particularly when used with pulsed primary ion beams, as in a conventional ToF set up, ions are expelled from the source in bundles or packets. These bundles of ions are then accelerated towards the flight tube by a potential difference applied between an electrode and extraction grid. The m/z ratio can be calculated from measuring the time it takes the ions to move through the field free region from source to detector.

At the point where the ions are accelerated, a potential, $V$ is applied. The potential ($V$) is related to the $E_k$ of the ion as shown in Equation 2.3

**Equation 2.3 Relationship between kinetic energy of ions and potential applied**

$$E_k = V z$$

Where:

$V$ is the potential applied to the ions

$z$ is the charge on the ion
After initial acceleration, the ion travels in a straight line at a constant velocity to the detector. The time required to travel distance, \( L \) before reaching the detector is given by Equation 2.4.

**Equation 2.4 Relationship between time and distance travelled by ions**

\[
t = \frac{L}{v} \quad \text{or} \quad v = \frac{L}{t}
\]

Where:

- \( t \) is the flight time
- \( L \) is the length of the drift region
- \( v \) is the velocity of the ion

Combining Equation 2.3 and Equation 2.4 gives the working equation for the ToF mass spectrometer whereby the \( m/z \) for each ion can be calculated, Equation 2.5.

**Equation 2.5 Time of Flight equation**

\[
m/z = \frac{2Vt^2}{L^2}
\]

Where:

- \( m/z \) is the mass to charge ratio of the ion
- \( V \) is the potential applied to the ions
- \( L \) is the length of the drift region
- \( t \) is the flight time

In principle, there is no upper mass range for a ToF instrument, however early instruments had poor mass resolution because in practice ions can have different initial
velocities prior to acceleration into the flight tube [237]. This was overcome with the development of the reflectron by Mamyrin in 1973 which is a device designed to focus the ions on the basis of their energy [238].

2.1.5.2 The Reflectron

The reflectron which in its simplest form consists of a series of equally spaced grid or ring electrodes, is placed at the end of the flight tube, opposed to the ion source. The detector is positioned on the source side, to detect the arrival of ions after they are ‘reflected’. The detector is typically orientated at a slight angle to the direction of travel of the ion packets, allowing the detected ions to follow a path that is a different angle to incoming ions [239]. Figure 2.4 shows a schematic of a reflectron.

![Figure 2.4 Schematic of a reflectron positioned at the end of the drift region.](image)

The reflectron corrects the kinetic energy dispersion of ions with the same m/z ratio, as the ion with greater kinetic energy (illustrated as black in Figure 2.4) will penetrate the reflectron more deeply and therefore have a longer flight path. This allows ions of the same m/z to essentially catch up with each other and thus arrive at the detector at the same time. Because the reflectron decreases the ‘spread’ (peak width) of the ions
and increases the flight path by ‘reflecting’ ions back down the drift tube, mass resolution is significantly improved.

The BioToF mass analyser is an R-500 (Kore Technology, UK), mounted vertically with a 2.5° tilt to give an effective length of 3 m [240].

![Image of R-500 ToF mass analyser](image)

*Figure 2.5 A picture of an R-500 ToF mass analyser, reproduced from reference [240].*

### 2.1.5.3 The Detector

Ions striking the detector give rise to an electrical signal which is measured and reported. The most widely used ion detector in mass spectrometry is the electron multiplier in which ions from the flight tube strike a conversion dynode held at a high potential [241]. This causes secondary particles to be emitted which are converted to electrons and amplified by a cascade effect to produce a current. The microchannel plate (MCP) is a type of continuous dynode electron multiplier in which a potential gain is applied across a number of parallel cylindrical channels within the plate and upon which a semiconductive substance covers each channel, which gives off secondary electrons as it is struck by an incoming ion. The electron multiplication effect within a channel can multiply the number of electrons by $10^5$ and using two plates, amplification can reach $10^8$ [241]. This type of detector has a very fast response time because the secondary electron path inside the channel is very short and hence it is most widely used for ToF analysers which require precise arrival times and narrow pulse widths. The signal from the detector is then recorded by a time-to-digital convertor (TDC) which registers a hit on the computer each time a signal above a threshold voltage is detected.
2.1.6 Methodology for ToF-SIMS Experiments

Samples were analysed on the BioToF with a wien filtered Au$_3^+$ cluster primary ion beam which was operated at 20 keV. The wien filter is a device for mass filtering the ion beam and is comprised of a magnetic field and an electric field applied orthogonal to each other and to the incoming primary ion beam [242]. This allows only ions of a suitable mass to pass through the filter and is used for selecting the desired isotope of a cluster beam.

A primary ion fluence of $2 \times 10^{11}$ ions cm$^{-2}$ was used for each analysis to minimise sample damage. To optimise signal intensity and maintain image resolution, a pulse width of 100 ns was used. Images were acquired over a $256 \times 256$ μm field of view and delayed extraction was used to improve mass resolution. Ions were detected with a microchannel plate-scintillator-photomultiplier assembly operating at 8 keV post acceleration. Both positive and negative ion spectra were acquired for each sample in the mass range $m/z$ 0 – 1000.

2.2 Liquid Chromatography-Mass Spectrometry

The hyphenated technique liquid chromatography-mass spectrometry (LC-MS) has been extensively used for the past 40 years across a range of applications and industries from environmental, forensic, biomedical, pharmaceutical to clinical [243-246]. It is a highly sensitive and versatile technique, capable of separating out a wide range of complex mixtures of analytes (LC) and subsequently directing them into a mass spectrometer for detection and identification.

2.2.1 Liquid Chromatography

Chromatography is a separation technique in which there are two phases, a stationary phase (the column) and a mobile phase which is forced in one direction (liquid).

The principle underlying any chromatographic separation is described by the partition coefficient (K), which states that an analyte will have an interaction with the stationary phase and mobile phase at any given time. Analytes with a higher concentration in the stationary phase (C stationary) will be retained longer on the column and thus elute.
later than analytes having a preference for the mobile phase (C mobile).

\[ K = \frac{C \text{ (stationary)}}{C \text{ (mobile)}} \]

Ultra high performance liquid chromatography (UHPLC) employs stationary phases of very small particle size (e.g. 1.5 µm) to increase the efficiency (N) of the chromatography which reduces band broadening and manifests in sharper chromatographic peaks and often shorter run times than in conventional HPLC.

### 2.2.2 Coupling Liquid Chromatography to Mass Spectrometry

Once the analyte has eluted from the column in the mobile phase, the flow of liquid must be directed into a mass spectrometer. Given that MS only detects ions (positive or negative) and past the inlet the instrument is primarily under vacuum, this is no simple task. Electrospray ionisation (ESI) is the process by which the MS source (under atmospheric pressure) evaporates liquids into gases, ionises analytes and sprays them into the mass spectrometer.

Eluent from the column is pumped through a narrow capillary and a potential difference is applied between the capillary tip and a sampling plate. The charged aerosol is heated to evaporate the volatile solvents from the LC mobile phase and as this occurs, the shrinkage into smaller droplets induces ‘coulombic fission’ when the repulsion forces in the charged droplet are strong enough to overcome the forces that hold the droplet together. This promotes the release of analyte ions into the gas phase and an electrical field gradient directs the ions through a series of ion focussing and transferring devices, each at decreasing pressures until the ions eventually reach the mass analyser under vacuum.
LC-MS employed in this project was using the Thermo Scientific Orbitrap as the mass analyser, a unique technology based upon an ion trap mass spectrometer as described in section 2.2.4.

### 2.2.3 Ion Trap Mass Spectrometry

The ion trap device was first described by Paul and Steinwedel in 1953 and was modified for use as a mass spectrometer later that decade. It further evolved through the decades into a mass storage device and later mass ejector device and has been widely employed as a powerful analytical tool since the late 1980’s. The evolution of ion trap mass spectrometry and theory behind the technique is described in a comprehensive review by Todd [247]. In brief, an ion trap employs an oscillating field to store ions. An RF quadrupolar field traps ions in either a 2-D or 3-D trajectory and ions are then ejected by applying an RF potential voltage to the ring electrode, causing ions of different masses to be expelled from the trap. Tandem mass spectrometry can be performed in the trap with time-dependent MS/MS giving rise to all fragment ions of a selected precursor m/z contained within the trap. MS” is subsequently possible according to the same methodology, by selecting a fragment m/z in the trap and fragmenting it further.

### 2.2.4 The Orbitrap

Orbitrap technology was first described by Makarov in patents dating from 1996 and 2004, with the first commercial instrument available by Thermo Electron Corporation in 2005 [248]. This completely new technology employs an electrostatic ion trap and uses Fourier transform to obtain mass spectra. Ions oscillate around a central electrode at high voltage and adopt a spiral trajectory. An outer electrode representing a barrel is coaxial with the inner spindle-like electrode and the frequency of harmonic oscillations of trapped ions along the axis of the electric field are measured as mass/charge values. Mass spectra obtained have high mass resolution, mass accuracy to within 2-5 ppm with internal (2 ppm) or external (5 ppm) calibration and high dynamic range. The instrument consists of several multipole focussing devices, a linear ion trap with two detectors and a bent quadrupole (‘C-trap’) which stores ions
and ejects them into the orbitrap. A detailed description of the instrument is described by Hu et al [249]. A selection of applications and further details on LC coupled to the Orbitrap is discussed by Makarov and Scigelova [250].

### 2.2.5 Methodology for UHPLC-MS Experiments

For metabolic profiling, cell extracts were analysed using an Acquity UHPLC system (Waters, UK) coupled to an electrospray (ESI) LTQ–Orbitrap XL hybrid mass spectrometer (Thermo Scientific Ltd. Hemel Hempsted, UK). Chromatographic separations were performed on a Hypersil Gold C18 UHPLC (length 100 mm, diameter 2.1 mm, particle size 1.9 µm, Thermo Scientific Ltd. Hemel Hempsted, UK). The column was maintained at 50 °C and samples were eluted with a gradient mobile phase of water (+ 0.1% formic acid v/v (A)) and methanol (0.1% formic acid v/v (B)) as follows: the column was held at 100% A for 1 min and subsequently ramped to 100% B over 6.0 min, followed by a 3.0 min hold at 100% B before a rapid return to 100% A and an equilibration of 2.5 min (total run time 12.5 minutes) at a flow rate of 0.40 mL min⁻¹. Samples were maintained at 4 °C in the autosampler and sample injection volume was 10 µL (with all 10 µL transferred onto the column) and 100% of the column eluent directed into the mass spectrometer.

In total 48 extracts were analysed (24 methanol extracts and 24 chloroform extracts) and sample injection order was randomised to eliminate any bias.

Electrospray ionisation (ESI) was carried out in both positive and negative polarity for each sample. Centroided accurate mass spectra were acquired in the m/z range of 50 – 1000 using the Orbitrap mass analyser operating with a target mass resolution of 30,000 (Full Width Half Maximum as defined at m/z 400) and a scan time of 0.4 s. All samples were analysed within 80 hours of reconstitution. Mass calibration was performed according to the manufacturer’s guidelines. The first 10 injections were pooled QC cell extract samples, followed by triplicate injections of each cell extract.
2.3 Infrared (IR) Spectroscopy

Infrared spectroscopy is a widely employed analytical technique based on absorbance of electromagnetic radiation by matter which occurs due to different vibrational modes of chemical bonds. This allows a chemical signature to be obtained for a range of biological molecules present in a sample and as such IR spectroscopy is a highly informative, rapid and non-destructive tool for the analysis of biological material such as cells and tissue. For the purposes of this project, the mid-infrared region of the electromagnetic spectrum was collected which readily interacts with the molecular vibrations of organic and biomolecules. This spans the wavelength range 2.5 µm to 25 µm which correlates to a wavenumber range of 4000 cm\(^{-1}\) to 400 cm\(^{-1}\). In practice, the cut off for interpretation of reliable spectra in this region is achieved at approximately 10 µm (1000 cm\(^{-1}\)).

2.3.1 Principles of Infrared Spectroscopy

When a sample is irradiated with an infrared beam, functional groups within the sample can absorb radiation in accordance with quantum theory which describes energetic transitions between quantized vibrational states. A consideration of the vibrational nature of bonds provides an explanation of molecular absorption in the IR region. Relative positions of atoms in any molecule are fluctuating continuously with respect to each other as a consequence of different types of vibrations or rotations. Covalently bonded atoms have an electric dipole moment as a result of electronic rearrangement between bonded atoms and since molecules are in continuous movement, this dipole moment oscillates at a particular frequency. If the frequency of an incoming photon from IR radiation exactly matches the frequency of the dipole moment within a molecule, radiation will be absorbed and a change in amplitude of the molecular vibration will result. Thus, for a molecule to be infrared active, a change in dipole moment must occur.

The complexity of molecular motion is often simplified by considering the atoms involved as individual masses and considering the covalent bond which connects them.
as a spring performing a simple harmonic oscillation. Hooke’s law describes the movement of such a diatomic molecule according to Equation 2.7.

Equation 2.7 Frequency of oscillation of a spring in harmonic oscillation

\[ \nu = \frac{1}{2\pi} \sqrt{\frac{f}{\mu}} \]

Where:

\( \nu \) is the frequency of oscillation

\( f \) is the bond force constant

\( \mu \) is the reduced mass given by Equation 2.8.

Equation 2.8 Reduced mass of a two-body system

\[ \mu = \frac{m_1 m_2}{m_1 + m_2} \]

In practice, real molecules do not exactly obey the laws of simple harmonic motion since sufficient force applied will not only stretch, but, eventually break the bond (which equates to the dissociation energy). Figure 2.6 illustrates the Morse curve (in red) for a typical diatomic molecule undergoing anharmonic extensions and compressions, with a simple harmonic parabola illustrated (in black). The allowed vibrational energy levels are quantized and decrease in energy for anharmonic oscillations as the dissociation energy is neared.
Figure 2.6 The Morse curve illustrating energy of a diatomic molecule undergoing anharmonic oscillation, with allowed vibrational energy levels illustrated.

With the exception of homonuclear species, such as O₂ and N₂ (which have no dipole moment), all heteronuclear molecular species absorb IR radiation, making it a ‘universal technique’. The molecular motion of a polyatomic molecule encompasses rotational, translational and vibrational movements. For a molecule containing \( n \) atoms, there will be \( 3n \) degrees of freedom along the x, y and z axes of the Cartesian plane. 6 degrees of freedom are described by rotational and translational motion; therefore, the number of degrees of vibrational freedom for non-linear molecules can be determined by \( 3n - 6 \). A linear molecule can only rotate along 2 axes, therefore the number of degrees of vibrational freedom for a linear molecule is calculated by \( 3n - 5 \). Molecular vibrations can be simply categorised as stretching or bending. During a stretching vibration there is a change in the inter-atomic distance along the axis of a bond between two atoms, and during a bending vibration there is a change in angle between two bonds. Bending vibrations can be further categorised as scissoring, rocking, wagging and twisting which are primarily observed for CH₂ functional groups.
Different classes of chemical compounds will respond to varying IR wavelengths, the most useful for biological compounds being the mid-IR region (4000 to 400 cm⁻¹). Specific absorptions can be correlated directly to biochemical species and the IR spectrum produced can be described as a characteristic fingerprint of the biological or chemical species analysed [59]. A characteristic cellular FTIR spectrum with tentative spectral assignments can be seen in Figure 2.7, reproduced from reference [148].

\[\text{Figure 2.7 Typical biological FTIR spectrum and tentative peak assignments from a population of cells. Reproduced from reference [148].}\]
Spectral bands are typically narrow and easy to resolve allowing accurate and unambiguous assignments of bands to specific molecular groups. Bands in this project have been assigned according to previously described studies by Movasaghi et al in 2008 [251]. This comprehensive review of biological applications of FTIR includes a compilation of hundreds of FTIR bands from a wide range of biological applications that have been published, with assignments ranging from 472 cm\(^{-1}\) to 3611 cm\(^{-1}\).

The use of the mid-IR region by chemists for identifying organic compounds began in the late 1950’s with the development of cost effective and easy to use double beam recording spectrophotometers [252], which compensate for source and detector drifts often observed in single beam instruments. There have since been extensive studies in a wide field of biological application areas; for example diagnosis of disease states which are discussed in section 1.5.2.5.

The most common IR experimental configuration is that of a transmission experiment in which infrared light is passed through the sample of interest. Light absorbed is defined by Equation 2.9.

\[
A = -\log \frac{I}{I_0}
\]

Where:

- \(A\) is the absorbance of the species
- \(I\) is the intensity of transmitted light
- \(I/I_0\) is the intensity of incident light

The ratio \(I/I_0\) is termed the transmittance, \(T\).

The Beer-Lambert law describes the absorption of monochromatic light which is directly proportional to the path length through the medium and concentration of
absorbing species, according to Equation 2.10 which underpins all spectroscopic techniques.

Equation 2.10 The Beer-Lambert law

\[ A = \varepsilon c l \]

Where:

\( A \)  
is the absorbance of the species

\( \varepsilon \)  
is the molar absorptivity coefficient

\( c \)  
is the concentration in moles per litre

\( l \)  
is the path length in centimetres

This forms the basis for quantitative spectroscopic applications.

### 2.3.2 FTIR Instrumentation

Prior to the 1980’s, mid-IR region instruments were primarily of the dispersive type and used diffraction gratings.

Today, this type of instrument has been largely superseded by Fourier transform (FT) spectrometers for the mid and far-IR regions. FTIR is now a highly established technique which is desirable because of advantages in speed, signal-to-noise and reliability. The reproducibility and robustness of FTIR cannot be overstressed [59]. FTIR instruments allow high throughput up to thousands of samples per day via 96-384 well plates which are rapidly scanned and offer non-destructive analysis of a wide range of compounds [253]. Sample preparation is minimal, typically samples are spotted onto a sample plate made from IR transparent material such as zinc selenoid or silica and are dried prior to analysis. Removal of water from samples is crucial for successful FTIR analysis as the absorption of water is very intense. This is sometimes
perceived as a disadvantage of transmittance FTIR, but can be overcome by dehydration of samples, subtraction of the water signal from the spectra or by applying attenuated total reflectance methods (ATR). In 2004, Winder and Goodacre employed ATR FTIR for the novel study of the biochemical profile of bacteria without the need for sample drying for the first time [254].

The majority of commercially employed FTIR instruments are based on the Michelson interferometer design, for which A.A. Michelson was awarded the 1907 Nobel Prize in physics [252]. The interferometer is comprised of a source, three optical elements (a fixed mirror, scanning mirror and beam splitter) and a detector; and is a device that modulates optical radiation by initially splitting a beam of radiation into two beams of nearly equal power, one onto the fixed mirror and one onto a moving mirror. When these beams are recombined, the intensity variations of the combined beam can be measured as a function of differences in the path lengths of the two beams, ultimately allowing time-domain spectra to be acquired. The time-domain signal (or interferogram) is converted into a more useful frequency domain signal by applying Fourier transform, and this gives rise to the x-axis of the spectrum in either wavenumber (frequency or wavelength). Either transmission or reflection IR can be measured, with transmission being the most widely reported for biological applications, in particular with IR-microspectroscopy. Attenuated total reflection (ATR) spectroscopy is a method used for collecting reflection and is discussed in section 2.3.6.

Figure 2.8 shows a schematic of a basic FTIR spectrometer adapted from reference [252].
During operation, a reference signal is first obtained by scanning a reference (typically air), usually co-adding twenty or thirty scans. The sample is then inserted into the radiation path, scanned and then the ratio of the reference spectral data to that of the sample is computed to give the transmittance at the range of frequencies.

Whilst a powerful and widely employed technique, there are limitations of FTIR spectroscopy. Ideally, the spectra obtained from a biological sample will contain only signatures from biologically relevant species arising from the biochemistry of the sample. This would be true if, for example, cells were ‘flat’. Unfortunately, the heterogeneous morphology of real biological samples gives rise to scattering effects.
which occur when the light interacts with species of different morphology and in particular when the size of these species (e.g. a cell nucleus) is of the order of magnitude of the wavelength of light hitting it. This phenomenon was discussed in section 1.5.2.7.

2.3.3 Synchrotron-FTIR

Synchrotron radiation is produced when electrons are accelerated to reach almost the speed of light. The accelerated electrons are deflected in a magnetic field to produce a strong beam of light from a point source. The infrared component of this light is extracted and channelled as a highly collimated beam which is reportedly 100 to 1000 times stronger than that from a globar source, with brilliance greater at the longer wavelengths (far-IR). In practice, losses of the beamline through reflection, transmission and absorption of gases along the optical path result in a decrease of approximately an order of magnitude of this brilliance. But still there is a considerable advantage when compared to a conventional source, particularly when using the synchrotron radiation in FTIR-microspectroscopy. Further advantages of synchrotron radiation when compared to a globar source are increased stability of power delivered in time and the time structure of the radiation which is pulsed therefore allowing real-time experiments to be performed on a time scale of milliseconds in certain experimental configurations [255].

2.3.3.1 Production of Synchrotron Radiation

Details of the production and extraction of synchrotron radiation at the Australian Synchrotron facility in Clayton, Victoria are described in Figure 2.9 with information taken from reference [256]. The mid-IR beamline at this facility was the synchrotron source employed in this project.
The power of coupling an FTIR spectrometer to an optical microscope has been exploited for many years to probe the biochemical nature of small samples such as cells or specific areas of larger samples such as tissue. In this configuration, the detector of the FTIR spectrometer is typically a liquid nitrogen cooled mercury cadmium telluride (MCT) detector which offers high sensitivity for such applications [257].

There is a drive for apertures sampling mid-IR light to decrease in size affording an increased spatial resolution for biological imaging applications such as imaging cells and tissue. For IR radiation from a standard source, as the aperture size decreases and approaches the wavelength of mid-IR radiation (2.5 µm to 25 µm), diffraction effects...
can cause spectra obtained to become increasingly distorted and as such are problematic to interpret. The well-defined narrow beam of light emerging from a synchrotron source has low divergence and when focused through narrow apertures it has distinct advantages over a globar source.

Figure 2.10 shows a comparison of FTIR microscopy spectra obtained across a 10 μm x 10 μm tissue section acquired using a synchrotron source and globar source (reproduced from reference [257]). It is clear that the spectrum obtained using the synchrotron source has a higher signal to noise ratio and less distortion than that from a globar light source for this particular sample.

![Figure 2.10 FTIR spectra (offset) of a tissue section showing comparison of synchrotron radiation (upper) and globar-source (lower) acquired with a spectral resolution of 8 cm⁻¹, 1024 scans and from an area of 10 μm x 10 μm. Reproduced from reference [257].](image-url)
2.3.5 Methodology for Synchrotron-FTIR Experiments

Synchrotron-FTIR measurements (S-FTIR) were recorded using the FTIR microspectroscopy beamline at the Australian Synchrotron. For the microspectroscopy beamline the beam is directed to a Bruker Hyperion 3000 IR microscope (Bruker Optics GmbH., Ettlingen, Germany) equipped with a liquid-nitrogen-cooled mercury-cadmium-telluride (MCT) detector with a x 36 IR objective (NA = 0.5). The Hyperion 3000 microscope is coupled to a Bruker Vertex 80v spectrometer and data collection was carried out using Bruker’s *OPUS version 6.5* software, with an additional 3D analysis package (Bruker Optics). The Hyperion microscope and the sample were purged with dry nitrogen gas to minimize water vapour contributions in the spectra. Spectra were recorded in transmission mode in the region 4000-900 cm⁻¹ with a spectral resolution of 6 cm⁻¹ using an aperture size of 10 x 10 µm. This aperture size was chosen to reflect the typical cell sizes of approx. 10-15 µm (HL60) and 14-18 µm (K562). Background spectra of a cell free region containing only cell culture media were taken every 5 spectra to account for any non-cell specific contributions to the signal and beam current fluctuations and 64 interferograms were co-added per measurement in order to obtain a good signal-to-noise ratio. Spectra were collected from at least 350 single cells for each cell line, which accounts for replicates and analysis on different days.

*Figure 2.11 Bruker Hyperion 3000 IR microscope at the Australian Synchrotron facility.*
### 2.3.6 ATR Spectroscopy

Attenuated total reflectance infrared (ATR-IR) is a spectroscopic technique that probes the surface layer of a sample. A sample is deposited onto the surface of a material transparent to mid-IR radiation (e.g. diamond) with a higher refractive index than the sample (such as a prism). A schematic of this process is shown in Figure 2.12. The IR beam passing through the medium of denser refractive index ($n_1$) is internally reflected at the boundary with medium of lower refractive index ($n_2$) providing that the angle of the incident beam ($i$) is greater than a critical angle ($i_c$).

![Schematic of ATR Spectroscopy](image)

*Figure 2.12 A schematic of the process of internal reflection spectroscopy. Adapted from reference [257].*

An evanescent wave is produced which becomes attenuated and penetrates into the sample of lower refractive index. During this process, it is common for 10-20 reflections to occur during one passage of the IR beam and therefore generating a good quality spectrum for the sample. Penetration of the IR beam is limited to a depth of approximately 0.3 µm to 3 µm and as such this methodology is not limited to thin samples [257]. Providing that the sample remains in close contact with the ATR window, the sample can be a solid or liquid in a wide range of morphologies and therefore the technique offers an attractive alternative for samples that cannot be readily studied by transmission IR.
2.3.7 ATR-FTIR Methodology

ATR-FTIR analyses were performed on an Agilent 4500a portable FTIR spectrometer equipped with a triple reflection diamond ATR sample interface (Agilent Technologies, USA). Spectra were collected at 4 cm\(^{-1}\) resolution with 64 co-added interferograms corrected against a clean diamond background. For each cell line, 9 replicates were analysed for both drug treated and control cells.

2.4 Raman Spectroscopy

The phenomenon of Raman scattering was first experimentally demonstrated by C.V. Raman and K.S. Krishnan in 1928 [258]. In 1930 Raman was awarded a Nobel prize in Physics for his work on “the scattering of light and the discovery of the effect named after him” [259]. The technique is based upon the changes in polarizability of vibrating molecules and the scattering of light that is produced as a result. It is often described as a complementary technique to infrared spectroscopy and has found much use in biological applications in recent years as a non-destructive and rapid tool for understanding the molecular and biochemical information of cells or tissue without the need for sample preparation such as staining. The basic selection rule for Raman scattering is that there must be a change in the polarizability of the molecule.

2.4.1 Principles of Raman Spectroscopy

When an incident beam of radiation is passed through a substance, the majority is absorbed, giving rise to quantised vibrational changes in the molecule that have been previously discussed in section 2.3.1. This however, is not the only change occurring. When such radiation passes through a transparent substance, a small amount is scattered in all directions according to the molecules present in that medium. In the case of monochromatic incident electromagnetic radiation of a narrow wavelength, this will interact with the molecule and polarise the electron cloud that surrounds the nuclei. The extent to which the electron cloud is distorted depends upon the ability of the electrons to polarize (the polarizability, \(\alpha\)). A dipole is created along x, y and z Cartesian coordinates which can be described by Equation 2.11.
Equation 2.11 Dipole induced by oscillating electric field

\[ \mu = \alpha E \]

Where:

\( \mu \) = induced dipole moment

\( \alpha \) = the polarizability

\( E \) = strength of the electric field from incident radiation

The polarizability, \( \alpha \), is a function of the instantaneous position of atoms within a molecule which varies with time (the ability to perturb the electron cloud of a molecule being dependent upon the relative location of individual atoms which oscillate with frequency \( \nu_k \)), which can be expressed by Equation 2.12.

Equation 2.12 Polarizability changed by vibrational displacement

\[ \alpha = \alpha_0 + \frac{\delta \alpha}{\delta Q} Q_0 \cos(2\pi \nu_k t) \]

Where:

\( \delta Q \) = the physical displacement of atoms about their equilibrium position due to a particular vibration mode

\( \alpha_0 \) = polarizability of the molecule at equilibrium position

For the incident electromagnetic wave, the electric field can be expressed according to Equation 2.13.

Equation 2.13 Fluctuation with time of incident electromagnetic wave

\[ E = E_o \cos (2\pi \nu_o t) \]

Where:

\( \nu_o \) = the frequency of incident radiation
The oscillating electric field can induce a time-dependent induced dipole in the molecule as described by Equation 2.14.

*Equation 2.14 Time-dependent induced dipole moment*

\[ \mu = \alpha E_0 \cos (2\pi v_0 t) \]

Combining Equation 2.12 and Equation 2.14 gives Equation 2.15.

*Equation 2.15 Induced dipole moment in terms of polarizability change with vibrational displacement*

\[ \mu = E_0 [\alpha_0 + \frac{\delta \alpha}{\delta Q} Q_0 \cos(2\pi v_k t)] \cos(2\pi v_0 t) \]

Applying a trigonometric identity for the product of the two cosines gives Equation 2.16.

*Equation 2.16 Raman equation describing three distinct frequencies of induced dipole moment*

\[ \mu = \alpha_0 E_0 \cos(2\pi v_0 t) + \frac{\partial \alpha}{\partial Q} \frac{Q_0 E_0}{2} [\cos 2\pi(v_0 - v_k) t + \cos 2\pi(v_0 + v_k) t] \]

The induced dipole moment as described by Equation 2.16 is created at three distinct frequencies, \( v_0, (v_0 - v_k) \) and \( (v_0 + v_k) \) which result in scattered radiation at these frequencies. The majority of scattered light will be of the same frequency of the excitation light, with no energy lost, and will therefore scatter elastically (Rayleigh scattering) which is described by \( v_0 \). A much smaller proportion of the incoming radiation will cause a change in the nuclear motion of the molecule allowing transfer of energy from incoming photon to molecule or from molecule to scattered photon and in this case, light will scatter inelastically at a frequency below or above that of the incident radiation, which is Raman scattering. These processes are referred to as Stokes scattering \( (v_0 - v_k) \) and anti-Stokes scattering \( (v_0 + v_k) \) respectively.
The process of Raman scattering is weak, with approximately only one in every $10^6 – 10^8$ photons being Raman scattered but with modern instrumentation, this can be readily observed [252]. Figure 2.13 shows the energetic processes occurring in Rayleigh and Raman scattering.

*Figure 2.13* Energetic processes occurring in Rayleigh (light scattered from a material is of the same frequency (energy) as incident light) and Raman scattering. In Stokes scattering, the molecule starts out in a lower vibrational energy state and ends up in a higher vibrational energy state after scattering. In anti-Stokes scattering, the opposite is observed. The frequency (energy) differences between the Raman lines and incident line are solely characteristic of the scattering molecules and are independent of the excitation frequency.

The Raman shift depicted in a Raman spectrum, is specific to the molecules responsible for the scattering and therefore the position and relative intensities of bands in a Raman spectrum are characteristic of the molecular composition of a sample. A typical cellular Raman spectrum with tentative peak assignments is shown in Figure 2.14, reproduced from reference [148].
Figure 2.14 Typical cellular Raman spectrum with tentative peak assignments. Reproduced from reference [148].

<table>
<thead>
<tr>
<th>Wavenumber (cm$^{-1}$)</th>
<th>Raman Peak Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>621</td>
<td>C-C twisting mode of Phenylalanine (proteins)</td>
</tr>
<tr>
<td>642</td>
<td>C-C twisting mode of Tyrosine and Phenylalanine</td>
</tr>
<tr>
<td>670</td>
<td>Thymine, Guanine (DNA/RNA)</td>
</tr>
<tr>
<td>720</td>
<td>C-N stretching in Adenine and lipids</td>
</tr>
<tr>
<td>750</td>
<td>Symmetric breathing of Tryptophan (protein)</td>
</tr>
<tr>
<td>782</td>
<td>Uricil, Thymine, Cytosine (ring breathing modes in the DNA/RNA)</td>
</tr>
<tr>
<td>827</td>
<td>PO$_2$ stretching in DNA, Tyrosine</td>
</tr>
<tr>
<td>854</td>
<td>Ring breathing in Tyrosine and Proline (proteins)</td>
</tr>
<tr>
<td>935</td>
<td>C-C stretching mode of Proline and Valine</td>
</tr>
<tr>
<td>1003</td>
<td>C-C aromatic ring stretching in Phenylalanine</td>
</tr>
<tr>
<td>1030</td>
<td>C-H bending mode in Phenylalanine, C-N stretching in proteins</td>
</tr>
<tr>
<td>1080</td>
<td>PO$_2$ symmetric stretching (DNA/RNA)</td>
</tr>
<tr>
<td>1085</td>
<td>C-O stretching</td>
</tr>
<tr>
<td>1130</td>
<td>C-N stretching in proteins; C-O stretching in carbohydrates</td>
</tr>
<tr>
<td>1155</td>
<td>C-C and C-N stretching of proteins/lipids</td>
</tr>
<tr>
<td>1175</td>
<td>C-H in plane bending mode of Tyrosine and Phenylalanine; Cytosine, Guanine</td>
</tr>
<tr>
<td>1208</td>
<td>C$_4$H$_2$ symmetric stretching in Tryptophan, Phenylalanine</td>
</tr>
<tr>
<td>1220–1280</td>
<td>Amide III (C-N stretching, N-H bending, proteins), PO$_2$ asymmetric stretching (DNA/RNA)</td>
</tr>
<tr>
<td>1311</td>
<td>CH$_3$/CH$_2$ twisting mode of collagen and lipid</td>
</tr>
<tr>
<td>1340</td>
<td>Guanine (DNA/RNA), CH def. in proteins and carbohydrates</td>
</tr>
<tr>
<td>1430–1460</td>
<td>CH (CH$_3$) bending mode in proteins and lipids</td>
</tr>
<tr>
<td>1485</td>
<td>Amide II (N-H bending, C-N stretching, proteins); Adenine, Guanine (DNA/RNA)</td>
</tr>
<tr>
<td>1580</td>
<td>Adenine, Guanine (DNA/RNA); C=C bending mode of Phenylalanine</td>
</tr>
<tr>
<td>1615</td>
<td>C=O Phenylalanine, Tyrosine and Tryptophan</td>
</tr>
<tr>
<td>1620–1700</td>
<td>Amide I (C=O stretching, C-N stretching and N-H bending, proteins)</td>
</tr>
<tr>
<td>2890</td>
<td>CH$_3$ symmetric stretching (lipids)</td>
</tr>
<tr>
<td>2940</td>
<td>CH$_3$ and CH$_2$ symmetric stretching (lipids)</td>
</tr>
<tr>
<td>2980</td>
<td>CH$_3$ symmetric stretching (lipids)</td>
</tr>
</tbody>
</table>
2.4.2 Raman Instrumentation

The emergence of lasers in the 1960’s gave rise to commercially available modern Raman spectrometers in which a sample is irradiated with a powerful beam of visible or near-IR monochromatic radiation. Because the intensity of Raman scattering varies as the fourth power of the frequency, lasers with lower wavelengths emitting in the blue and green region of the spectrum such as argon and krypton ion are often employed to give improved signal. However, near-IR sources have more recently become widely used with the advantage that they can be operated at much higher power without causing photodecomposition of sample. Fluorescence, which can be a common interference of the inherently weak Raman signal, is also minimised by employing laser excitation wavelengths far removed from an absorption band of the sample of interest or by exciting with a laser in the less energetic near-IR region [252].

A high quality grating system is employed to select the weak Raman signal from higher scattering signals around it. This typically comes in the form of holographic ‘notch filters’ or dielectric edge filters which significantly increase the signal to noise by virtually eliminating interferences.

Gratings split the scattered radiation into component wavelengths which are then measured at an angle to the incident beam by sensitive detectors. Many commercial instruments originally used a photodiode array detector for the simultaneous collection of the entire Raman spectral region; however Fourier transform (FT) instruments became popular since 1987. Despite significant advantages such high spectral resolution capabilities, the disadvantages of FT Raman spectroscopy such as the strong absorption of aqueous samples mean that dispersive instruments are still heavily favoured. A more recent development which can be seen in the Renishaw inVia Raman Microscope, employs a charge-coupled device (CCD) as a multi-channel detector which has distinct advantages over Fourier transform instruments, such as wider range of laser options and capability for reduced noise and higher efficiency.
2.4.3 Raman Microspectroscopy/Microscopy

Coupling an optical microscope to a Raman spectrometer provides a powerful analytical tool. The Renishaw inVia microscope is an example of cutting edge technology in this field. A schematic of this can be seen in Figure 2.15.

![Schematic of Renishaw inVia Raman microscope.](image)

Initially the sample is illuminated by visible light and brought into focus on the area of the sample to be probed. There are possible adjustments in the x, y and z directions which can be set to the specific area of interest, e.g. a single biological cell. The light is then turned off and the laser beam is directed onto the sample, typically excitation wavelengths of 785 nm and 830 nm are employed to minimise fluorescence, however 532 nm has proven to be successful in enhancing the signal of specific chemical functional groups in the high wavenumber region arising primarily from biological lipid species. In the ‘Streamline’ configuration, optics within the microscope illuminate a line on the sample and the motorised stage then moves the sample beneath the objective lens of the microscope so that the line is rastered across the area of interest. The detector is synchronised with this so that data are collected and recorded simultaneously acquiring a high spatial resolution map of the sample (in the order of 1 µm) from which high quality spectral information from each pixel of the image can be extracted [260].
2.4.4 Methodology for Raman Experiments

Raman spectroscopic analysis was performed using a Renishaw inVia Raman microscope (Renishaw). Raman spectra and maps were obtained; using a 785 nm excitation beam (cell extract spectra) and 532 nm excitation beam (cell maps), ~30 mW at the sample, 50 x objective, exposure time of 10 s for point spectra and 6 accumulations of 1 s (6 s in total) for mapping. StreamLineHR mapping (which uses a laser-spot focus) was carried out in high confocal mode, at a step size of 0.5 μm.

2.5 Multivariate Analysis

The acquisition of biological SIMS, Infrared or Raman data results in vast data sets typically comprising of thousands of points. In order to extract useful information pertaining to the biological system/disease state being studied, data must be manipulated and interrogated using appropriate chemometric tools. In multivariate analysis of large spectral data sets, information regarding the relationships between spectra is sought. Techniques employed to analyse large data sets can generally be classified into one of two techniques: unsupervised analysis and supervised analysis.

2.5.1 Unsupervised Analysis

Unsupervised algorithms are based on cluster analysis and provide insights into how similar (or not) samples within a sample set are. Principal component analysis (PCA) is a well known technique for reducing the dimensionality of the data whilst preserving most of the variance.

2.5.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a multivariate projection method designed to extract and display the systematic variation in a data matrix. It is a technique in which each column of data within a matrix is assessed for its variance, and orthogonal lines of best fit (principal components) are fitted to the data. The first principal component (PC1) is the direction describing the highest degree of variance within the data set. The
second principal component (PC2) is then determined which describes the highest degree of variance for the remaining data and so on.

By representing multivariate data as a low dimensional plane, an overview of the data can be obtained which can reveal trends in the data set or outliers and also uncovers relationships between observations and variables and amongst the variables themselves. In the case of LC-MS and ToF-SIMS, the samples are spectra and the variables are mass to charge ratios \((m/z)\). In the case of FTIR, the samples are spectra and the variables are wavelengths corresponding to bands of IR absorption, and in the case of Raman the samples are spectra and the variables are wavelengths corresponding to light scattering.

### 2.5.3 Data Pre-Processing

Prior to PCA, data are usually pre-treated to re-shape the data into a more useful model for analysis. There are a number of pre-processing techniques widely employed, however, one has to be careful in their use and be aware that subsequent data interpretation such as quantitation can be skewed because pre-processing makes assumptions on the variance within the data set.

Scaling is carried out to ensure that variables within the space have the same contribution to the model. This overcomes the fact that a variable with a large range could have a large variance, whilst some variables within a smaller range, may have small variance. There are many ways to scale data but the outcome of them all is to standardise variables in the data set which means that the length of each co-ordinate axis in the variable space is regulated according to a pre-determined criterion (eg. variance). There are a number of types of normalisation, in which a common spectral feature is identified and each spectra of the data set is divided by a constant from that individual sample such as total ion/signal intensity or intensity within a specific \(m/z\) (MS/SIMS) or wavenumber (Raman) region. Sum normalisation, vector normalisation and PQN normalisation are all examples of this methodology. Most spectra undergo normalisation and in doing so, peaks in each spectra are modified to their own internal standard, thus eliminating changes in signal response arising from experimental artefacts such as fluctuating primary ion dose or sample charging (SIMS).
Mean-centering is usually the second part of the pre-processing, in which the average value for each variable is calculated and subtracted from each individual spectra (each sample). This shifts the origin of the data from zero to a position at the centroid of the data set and subsequent PCA analysis therefore describes the variance of the data from the mean and makes the data set easier to interpret. Mean-centering compensates for the fact that in multivariate analysis it is common to compare parameters that are measured on different scales.

Mean centering is frequently employed, in particular in the pre-processing of SIMS spectra, however it is not always of necessity and where it is carried out, it should be the last pre-processing step before statistical analysis [71]. A geometrical approach to PCA is illustrated in Figure 2.16.

![Image of PCA representation](image)

**Figure 2.16** Representation of PCA for principal components 1 and 2. A K-dimensional variable space is set up showing three variable axes (A). Data points are plotted in the variable space (B). The vector of averages is calculated shown here as a red point (C). Mean-centering moves the average point to the origin (D). The first principal component, PC1 represents the maximum variance direction in the data (E) and the second principal component, PC2, reflecting the second largest variation in the data and orthogonal to PC1 (F). Figure adapted from reference [261].
For a matrix, X, with N observations and K variables, a variable space is set up with as many dimensions as there are variables. Figure 2.16A shows three variable axes. Into the K-dimensional variable space, each observation (which is a row of the X matrix), is plotted. Figure 2.16B shows the cloud of data points plotted in this variable space. Mean centering then determines the vector of averages, which is illustrated by a point situated in the middle of the data cloud in Figure 2.16C (red). The subtraction of the averages from the data cloud then causes a shift so that the average point (red) is now the origin as shown in Figure 2.16D.

The first principal component (PC1) is now computed and is represented as a line in the K-dimensional space that best accounts for the shape of the data cloud, i.e. represents the maximum variance in the data. The line goes through the average point as can be seen in Figure 2.16E. Each observation in the data cloud can then be projected onto this line which yields a coordinate value along the PC line which we describe as a score. Scores data for specific PC’s are typically plotted against one another to give a scores plot, from which evidence of clustering and correlations can be seen if there are any common characteristics between data points. One principal component is rarely enough to model the variation of a data set and therefore a second principal component is calculated. PC2 is plotted as a line orthogonal to PC1. PC2 also passes through the average point and reflects the largest source of variation in the data after PC1. This can be seen in Figure 2.16F.

The meaning of a scores plot is given by the loadings which inform how the variables are linearly combined to form the scores. Loadings display the magnitude of the correlation (large or small) and the manner of the correlation (positive or negative) in which the measured variables contribute to the scores.

PCA is thus a very powerful technique which relies on no prior information regarding groups that are trying to be distinguished.

2.5.4 Supervised Analysis

In contrast to unsupervised analysis which identifies patterns in data by seeking evidence of natural groupings or clustering with no a priori knowledge; supervised analysis (often referred to as classification or discriminant analysis) has information at
the outset regarding both the number of parent groups and representative samples from each group (which can be used as a training set). This allows assignment of any unclassified data sets to one of the parent groups.

2.5.5 Discriminant Analysis (DA)

Discriminant analysis is one of the most powerful and widely used methods for supervised pattern recognition and can be used for data sets in which groups display a normal distribution of data (discriminant analysis is a parametric method).

The first step is to assign a classification rule to the data set, for example, by assigning a sample to the group with the nearest mean value. Once assigned, this rule then needs testing to confirm how good the classification is. There are a number of methods by which the effectiveness of the model can be tested such as by using data that were not included in the training set at all, by employing a new training set or using the ‘leave one out’ method.

2.5.6 Principal Component - Discriminant Function Analysis (PC-DFA)

Principal Component-Discriminant Function Analysis (PC-DFA) is a Discriminant Function Analysis (DFA) which employs PCA to reduce the dimensionality of the data followed by discriminant analysis which discriminates between groups based on the amount of each input variable that a group has in common. In actuality, the model tries to maximise the variance between groups whilst minimising the variance within each group [71]. This relies on a priori knowledge of which group samples analysed belong to and as such is termed a supervised technique.

The two main objectives of DFA are to understand the features that contribute to group differences and to generate a prediction of the likelihood that an unknown (e.g. sample) will belong to a particular class or group. The outcome of DFA is a set of discriminant functions (variates) which are the combination of two or more independent variables that best describe the pre-defined groups’ structure. Scores and loadings plots from DFA are more well defined than those from PCA because additional information relating to class structure allows better understanding of groups.
It is common to validate PC-DFA data independently by projecting subsets of the data into the PC-DFA ordinate space. Initially each class of data would be randomly split into a training set and test sample set. The PC-DFA model would then be generated for the training data set and tested by projecting the test set into the same ordinate space. A close alignment between training and test data validates the model [71].

2.5.7 Partial Least Squares - Discriminant Analysis (PLS-DA)

Partial least squares - discriminant analysis (PLS-DA) is a classification method [262] widely employed for large spectroscopic or spectrometric data sets to predict the classification of samples within a population studied and to facilitate biomarker selection in metabolomics studies.

In the PLS1 model, a dependent variable (y) is categorical (rather than continuous) and is chosen to represent class membership, for example, y can be a vector with values -1 and +1 where -1 represents samples belonging to a control group of samples and +1 representing each sample belonging to a drug-treated group. By using information of group classes, PLS-DA can improve the separation between these two groups of samples. The relationship that develops between the categorical variate y and the data, means that categorical variable values can be subsequently predicted for samples of unknown origin by using the spectral data.

Rigorous validation of PLS-DA models is required because of the tendency for the method to overfit the data and permutation testing is widely employed to confirm the suitability of the PLS-DA model to correctly classify [263]. Permutation tests evaluate whether the classification of two assigned groups is significantly better than a random classification of two arbitrary groups. By randomly assigning class labels to different samples, and calculating a classification model based on ‘incorrect’ assignments, the newly calculated classification model should not be able to predict class membership very well. Such permutation tests are repeated many times to yield a distribution of classifications that are expected to not be significant or not.

A widely used variant of PLSDA is orthogonal PLSDA (oPLS-DA), in which the first components orthogonal to the dependent variable are removed from the data [264]. OPLS enhances the interpretation of PLS by forcing all classification information into
a single component (the other components describing the variation that is orthogonal to the class information), however, the prediction or classification power of both models is the same.

2.6 Summary

The application of spectroscopic and mass spectrometric techniques employed in this project provide the ability to probe a broad spectrum of biomolecular species within the cell. This chapter has explored the technological aspects of each analytical platform and Table 2.1 provides a qualitative comparison summary of these multimodal spectroscopic and mass spectrometric techniques for the analysis of cells. It is clear that no one sample preparation or analytical method is suitable for cell analysis across all techniques and it is therefore essential that experimental design, incorporating sample integrity and instrument performance, are optimised to gain high quality data for subsequent biological interpretation. These considerations are addressed in Chapter 3.
<table>
<thead>
<tr>
<th>Sample prep considerations</th>
<th>ToF-SIMS</th>
<th>LC-MS</th>
<th>Mid-FTIR spectroscopy</th>
<th>Synchrotron-FTIR</th>
<th>Raman spectroscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cells dehydrated or frozen</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Deposition of monolayer of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cells for analysis can be</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>challenging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cells must be extracted,</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>potentially losing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biochemical information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample not too thick (&lt; 15</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>µm) and must be</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dehydrated for analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little sample prep</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>required – cells can be</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little sample prep</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>required – cells can be</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in vivo applications</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No/Limited</td>
</tr>
<tr>
<td>possible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Destructive | Yes | Yes | No | No | No |
| Real time    | No  | No  | Yes| Yes| Yes |
| Signal to noise | Medium | High | Medium | Medium-High | Low |
| Imaging Spatial Resolution | High (1-10 µm depending on SNR) | No spatial information possible | Low (due to diffraction limit, approx. 10 µm therefore sub-cellular detail poor) | Medium (approx. 2-5 µm) | High (approx. 800 nm–1 µm) |
| Influenced by matrix | Yes - Matrix effects can alter ion intensities | Yes - Matrix effects from solvent in ESI causes interference from adducts etc. | No, as long as no water present | No | No |
| Influenced by water | No | No | Strong vibrations of water (O-H), disadvantage for biological systems | No | No/Limited |
| Throughput | Low | High | High | Medium | Medium |

Table 2.1 Qualitative comparison summary of multimodal spectroscopic and mass spectrometric techniques employed in this project for the analysis of cells.
Chapter 3 Experimental Principles and Methodology

3.1 Introduction

Prior to obtaining results presented in Chapters 4, 5 and 6, extensive method development was carried out for this project. AML cell lines were cultured in-house, and analysed both as whole cells and cell extracts across multiple techniques. This required sample preparation to be fit for purpose for each technique employed, from obtaining sufficient biomass for cell extracts to the use of appropriate fixatives for single cell analysis. Typically, many cell lines studied in drug-cell interaction situations reported in the literature are adherent cells, which are amenable to growth onto a substrate for analysis and are therefore somewhat easier to manipulate than suspension cells for which data is presented here. Deposition of cells and extracts onto different substrates for each instrumental technique was therefore investigated.

3.2 Acute Myeloid Leukaemia In Vitro Cell Lines

3.2.1 Cell Line Origin

Two AML cell lines have been employed in this project, HL60 and K562. They were selected based on the fact that they represent different stages in differentiation (HL60 being a more immature and less differentiated cell than K562); therefore, a consideration of biochemical changes of both could potentially yield more confident information on the mechanistic action of BaP. Secondly, previous BaP studies, discussed in section 1.6.4, have employed both cell lines, so to build on information already obtained is a desirable outcome of the project.

The HL60 cell line originated from a 36 year old female patient with AML in 1977 [265]. The original diagnosis was that of acute promyelocytic leukaemia but a re-classification in 1988 indicated that AML was a more appropriate description [266]. Culturing peripheral blood leukocytes from the patient in the presence or absence of conditioned growth medium led to active growth cells in suspension after 3 weeks and resulted in a growth factor independent immortal cell line [267]. HL60 cells resemble promyelocyte morphology, being large blast-like cells with large, round nuclei and...
containing azurophilic granules. They are desirable cells for many studies of
differentiation as they can readily be induced to further differentiate *in vitro*, with some
reagents causing differentiation to granulocyte cells and others to monocyte or
macrophage cells [268].

The K562 cell line was derived from leukaemic cells obtained from a 53 year old
female patient in blast crisis in 1970, who had been suffering with chronic myeloid
leukaemia for approximately 4 years [269]. They too were established as liquid
cultures and have been characterized as a highly undifferentiated cell of the
granulocytic series [270] and are now widely classified with the morphology of AML
cells. Unlike HL60 cells which undergo differentiation with BaP treatment, K562 cells
enter necrosis, eventually resulting in apoptosis (cell death).

### 3.2.2 Cell Culture

The use of *in vitro* cell cultures has been reported for many decades as way of
providing a regulated environment through which the biochemistry of mammalian
cells can be studied outside of the body. This presents an ideal way to study drug-cell
interactions without the need for invasive clinical procedures. Mammalian cells exist
in culture in one of two forms, adherent or suspension. Adherent cells are anchorage-
dependant and grow as a monolayer on the surface of the cell culture flask. For
passaging, they require cleavage with trypsin which releases cells from the surface.

Both HL60 and K562 cell lines readily grow as suspension cells and proliferate
without attaching to a substrate. At high densities of cells in suspension, the growth
medium becomes exhausted and cell cultures require splitting to provide fresh
nutrients for continued proliferation.

For all experiments presented in this thesis, acute myeloid leukaemia (HL60 and
K562) cell lines were cultured in RPMI 1640 medium (+ L-Glutamine) supplemented
with 10% (v/v) foetal bovine serum (FBS), 1% (v/v) penicillin-streptomycin (Life
Technologies, ThermoFisher) and maintained in exponential growth (between 0.25-
1.5 x 10⁶ cells/mL) at 37 °C with 5% CO₂. Penicillin and streptomycin were used to
safe-guard the culture from potential infection from gram-positive bacteria.
Establishing a culture of high integrity is essential to ensure the cultured cells continue to represent the primary cell line from which they were established. Where possible during this project recommended guidelines were followed [271] to ensure that cells were fit for purpose. Aseptic techniques were applied at all times and cells were handled in a designated laminar flow microbiology safety cabinet to prevent cell contamination from external sources and to avoid cross-contamination between cell lines.

3.2.3 Cell Counting

A number of cell passages were carried out prior to experimental drug treatment to ensure that cells had reached the expected growth rate for healthy cells which was typically a doubling rate of approximately 20-22 hours for HL60 cells and 24-26 hours for K562 cells.

Cell counting was carried out to determine the population of cells prior to passaging. This was particularly useful in the early stages of cell culture to gain a feel for the rate at which cell populations were growing and therefore decide on what split ratio to apply when passaging with fresh media to maintain cells within the aforementioned exponential growth concentration. Cells were pipetted into disposable haemocytometers and visualised under an optical microscope. A manual clicker counter employed for determining cell populations and Table 3.1 shows an early cell count for each cell line as representative of typical behaviour of the AML cells employed in this project. During this project there were periods of time during which cell lines were frozen down in DMSO and stored in liquid nitrogen. When bringing the cells back up to establish a new culture, it was observed that K562 cells grew at a much slower rate than HL60 cells and initially required 72 hours before reaching a population for passaging. Once established, both cell lines were typically passaged every 2 days.
<table>
<thead>
<tr>
<th>Date</th>
<th>Cell Type</th>
<th>Cell Count Squares: 1, 2, 3, 4</th>
<th>Average Cell Count</th>
<th>X 10⁴ volume factor</th>
<th>X 2 for Trypan blue staining dilution factor = Total cells, M mL⁻¹</th>
<th>Split for passaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>19/2/14</td>
<td>K562</td>
<td>48, 57, 61, 49</td>
<td>53.75</td>
<td>537500</td>
<td>1.070</td>
<td>1 in 4</td>
</tr>
<tr>
<td>19/2/14</td>
<td>K562</td>
<td>41, 48, 47, 42</td>
<td>44.5</td>
<td>445000</td>
<td>0.890</td>
<td>1 in 3</td>
</tr>
<tr>
<td>24/2/14</td>
<td>HL60</td>
<td>44, 65, 42, 50</td>
<td>50.25</td>
<td>502500</td>
<td>1.050</td>
<td>1 in 4</td>
</tr>
<tr>
<td>24/2/14</td>
<td>HL60</td>
<td>53, 46, 70, 92</td>
<td>65.25</td>
<td>652500</td>
<td>1.300</td>
<td>1 in 4 &amp; 1 in 5</td>
</tr>
</tbody>
</table>

Table 3.1 Typical cell populations counted using a haemocytometer during routine cell culture of AML cell lines.

### 3.2.4 Cell Viability Staining

A dye exclusion test was performed with Trypan Blue solution (Sigma Aldrich, UK) to determine the population percentage of viable cells at any given time. This was carried out at regular intervals during the project to confirm the integrity of cell populations in culture. Cells with compromised membranes readily take up the dye and are stained blue whereas viable cells with intact membranes remain unstained. Typically, no more than approximately 10% of cells in culture were showing significant alteration in cell membrane viability when tested at regular intervals throughout the project.

### 3.2.5 Cell Morphology and Characteristics

Both cell lines were visualised and photographed using a Nikon Eclipse 90i digital microscope (Nikon Instruments, Tokyo, Japan) with a x 40 objective. Images can be seen in Figure 3.1 for HL60 and K562 cells respectively. HL60 cells are noticeably smaller than K562’s and because they are earlier in lineage they are less granular in appearance. All HL60 cells in the image appear in-tact and healthy as depicted by a ‘halo’ effect surrounding the outer membrane. K562 cells are larger than HL60 cells and being more mature they are visibly more granulated. The K562 image depicts both healthy and compromised cells where the cell membranes are blebbing and these cells would be expected to stain positive with trypan blue. Within a population of cells in a
culture flask, some will naturally exist in various stages of cell growth and other cells may present at stages of necrosis or apoptosis.

![Image of HL60 and K562 cells](image)

**Figure 3.1** Optical images of HL60 and K562 cells taken with a x 40 objective. A visual difference in morphology can be seen between the two cell lines.

### 3.2.5.1 Mycoplasma Testing of Cell Lines

Mycoplasmas are microscopic prokaryotes that present a significant threat to cell culture communities. Because they are slow growing and tend not to ‘cloud’ the culture as other infections might, they can go undetected for long periods of time and reports suggest that this infection can significantly alter cell metabolism and growth. Because this has potential to affect metabolic data derived from experiments in this project,
mycoplasma testing was carried out at regular intervals of approximately every 1-2 months [272]. Both HL60 and K562 cell lines tested negative for mycoplasma infection and results confirming this can be seen in Appendix II.

3.3 Drug Compounds for Cell Inoculation

3.3.1 Drug Standards Preparation
Bezafibrate and medroxyprogesterone acetate in a drug redeployment situation (combination treatment denoted ‘BaP’ [213]) are the pharmaceutical compounds of interest employed in experiments presented in this thesis for reasons discussed in section 1.6.4. Bezafibrate (2-[4-[(4-Chlorobenzamido)ethyl] phenoxy]-2-methylpropanoic acid) was purchased as a white solid of molecular weight 361.82 g/mol and medroxyprogesterone acetate was purchased as a white solid of molecular weight 386.52 g/mol (Sigma-Aldrich, UK). The compounds were stored at room temperature prior to use.

Stock solutions of the drug were prepared to concentrations of 0.5 M bezafibrate in DMSO and 5.0 mM medroxyprogesterone acetate in ethanol to allow for a 1 in 1000 dilution when administered to cells in culture (eg. 4 µL of drug solution added to 4 mL of cell culture).

3.3.2 Raman of Drug Standards
Raman spectra were obtained for both bezafibrate and medroxyprogesterone acetate in solution and solid form. Strong spectral interference was observed from solvent peaks in drugs in solution; therefore, spectra are reported for microcrystalline drug powder on a quartz substrate (enclosed due to the toxic nature of drug powder if inhaled). Figure 3.2 shows spectra obtained for bezafibrate at 532 nm and 785 nm, and for medroxyprogesterone acetate at 785 nm with tentative peak assignments.
Raman spectra of quartz substrate for live cell imaging has been previously reported [273] and there appear to be no contribution from quartz to the drug spectra obtained. Peaks have been tentatively assigned in Figure 3.2. Notable drug peaks observed are found in the region of 1320 cm\(^{-1}\) (CH deformation), 1590 cm\(^{-1}\) (C=C stretch) and 3068 cm\(^{-1}\) (CH aromatic stretch) for bezafibrate with the peak at 1590 cm\(^{-1}\) being previously reported as the highest intensity Raman peak for this drug [274]. Medroxyprogesterone acetate has an interesting Raman signature with intense peaks at 1606 cm\(^{-1}\) (C=C stretch) and 1673 cm\(^{-1}\) (C=O stretch) which have also been reported in the literature, with the aforementioned peak previously used in the quantitative determination of MPA in pharmaceutical preparations [275, 276].

A knowledge of the position of drug signature bands is useful when studying drug-cell interactions as it can facilitate the mapping of drug uptake and distribution within a single cell and is also important to consider the possible interference of drug bands with those arising from cellular biological species.
3.3.3 Drug Treatment Protocol

HL60 and K562 cells were seeded at $0.5 \times 10^6$ cells/mL in either T-25 or T-75 Falcon flasks (vented) and treated with 0.5 mM bezafibrate in DMSO and 5 µM medroxyprogesterone acetate in ethanol. Control flasks (from the same seeding flask as drug treated cells) received equal volumes of ethanol and DMSO. Cells were incubated for 24 hours at 37 °C with 5% CO₂.

The dose of each drug administered was selected according to previously reported data investigating the IC50 and clinically relevant dose for these drugs determined from work carried out by collaborators at The University of Birmingham on the same cell lines [213].

Replicates for both cell lines were accounted for in all experiments conducted and specific details of these can be found in the relevant results sections.

3.3.4 Cell Cycle Considerations for Studying Drug-Cell Interactions

The impact of cells within a given population being at different stages of the cell cycle when administered drug treatment, was considered for studies carried out in this project. Whilst there is the potential for cells to respond differently to a drug depending upon which stage within the cycle it is in when treatment is administered, a recent report employing FTIR to study drug action as a function of cell cycle found that inherent cell cycle signatures from control and drug treated cells (Caki-2 cell line treated with 5-FU and Paclitaxel) could not be observed [277]. Cell-cycle dependent features were masked by more pronounced biological changes occurring as a result of drug treatment. This supported an earlier report that the cell cycle does not account for the major source of variability in FTIR spectra when cells are exposed to an anti-cancer drug [178]. For these reasons, cell cultures here were not synchronised prior to drug administration during this project. The 24 hour treatment protocol prescribed ensured that all HL60 cells in a given population (20-22 hour doubling rate) and the majority of K562 cells (24-26 hour doubling rate) would have undergone a complete cell cycle.
3.4 Method Development for Single Cell Analysis

The initial aim of the project was to establish protocols for deposition of whole cells and cell extracts onto appropriate substrates for Raman spectroscopy and ToF-SIMS, whilst maintaining sample integrity for subsequent analysis.

3.4.1 Deposition of Cells on Substrate for Analysis

Substrates employed in this project were chosen according to widely reported successful methods for each analytical technique. Silicon wafers were used for ToF-SIMS analysis because the conductivity of the Si wafer facilitates sample ionisation and Raman grade calcium fluoride windows were used for Raman microspectroscopy as they do not give rise to any interfering spectral bands in the biological region of interest in the spectra. A number of attempts were performed to establish the optimal parameters for cell coverage on substrates in order to avoid cell clumping or too sparse a distribution on the surface.

3.4.1.1 Method Development to Adhere Cells to Substrate for Raman Microspectroscopy

The Renishaw inVia Raman microscope described in section 2.4.2 is equipped with a cell culture chamber in which cells can be maintained at 37 °C with 5% CO₂ to represent the conditions of an incubator. This offers the potential for Raman mapping of live cells in situ without the need for any sample preparation and allows ‘real-time’ monitoring of cellular drug uptake. This attractive option was investigated with an alternative AML cell line, Kg1a, cultured in-house prior to analysis. This cell line was chosen specifically for this study as some reports suggest Kg1a cells can be more adherent than HL60 and K562 cells (despite them all being suspension cell lines) and therefore these cells were more likely to remain stationary during analysis.

The live cell Raman study was not successful with Kg1a suspension cells. When irradiated with the laser beam, cells began to slowly migrate away from the source and did not stay in the same place long enough for good quality spectra to be acquired, furthermore, it was impossible to locate the same cell more than once. Kg1a cells were
therefore not further employed in this project and all other studies reported are with
the aforementioned HL60 and K562 cell lines for biological reasons previously
discussed.

In an effort to overcome this, two methods were explored.

1. CaF$_2$ discs were placed into 6 well-plates containing different volumes of
HL60 and K562 cell cultures immediately after passaging to determine if any
cells would naturally ‘sit down’ on the substrate. This method was
unsuccesful. Initially upon removing substrates from the cell cultures at 24
hours, it appeared to the naked eye that cells were present on the surface.
Unfortunately, it was then necessary to wash with PBS to remove any cell
culture media residue prior to analysis and in doing so cells were easily
dislodged and washed away from the surface.

2. Poly-l-lysine which is a polymer frequently used for promoting cell attachment
to solid surfaces, was investigated at different concentrations. Poly-l-lysine
was pipetted onto the surface of a cleaned CaF$_2$ window and allowed to air dry
for 10, 20 and 30 minutes before substrate was placed in cell cultures according
to method 1. At high concentrations of poly-l-lysine some cell adherence was
observed at 24 hours but again, washing steps removed the cells easily.
Furthermore, concerns over altered phenotypes of cells growing whilst
artificially adhered on a substrate were an issue.

For these reasons, whole cells studied in this project were analysed as cells deposited
on CaF$_2$ windows for Raman microspectroscopy according to the method described in
section 3.4.1.2.

3.4.1.2 Preparation of Whole Cells for Raman Microspectroscopy

Cell suspensions of HL60 and K562 in growth media were removed from the
incubator and centrifuged at 1000 rpm for 5 minutes. The cell pellet was re-suspended
in Dulbecco's phosphate buffered saline solution (PBS), 500 µL (Sigma-Aldrich). The
solution was agitated well and transferred into an eppendorf tube. This was spun down
rapidly in a small bench-top centrifuge, the supernatant removed and pellet re-
suspended again in PBS, 500 µL. This wash procedure was repeated 3 times, in the
final step, leaving a small amount of PBS to re-suspend the cell pellet (approximately
50 µL containing approximately 3 million cells).

Two approaches were initially evaluated to determine the best method for transferring
cells onto substrate for subsequent Raman analysis.

1. Cell suspension, 50 µL was cytospun onto CaF₂ discs to give an even coating of
cells.

2. Cell suspension, 20 µL, was spotted onto CaF₂ and at the same time, excess
suspension was drawn off the disc to leave just a smear of cell extract remaining.

Both preparations on CaF₂ discs were allowed to air-dry before analysis.

An initial comparison of HL60 cells spotted onto CaF₂ discs versus cells cytospun onto
the same substrate yielded little difference between Raman spectra and white light
microscopy images obtained. Due to incorrect fitting of CaF₂ discs into the cytospinner
some sample was lost during the coating onto substrate and thus sample preparation of
choice for work reported here was that of spotting onto discs according to the
procedure described in method 2 above.

3.4.2 Raman Microspectroscopy Method Development

Preliminary experiments were performed on HL60 control and BaP treated cells for
Raman method development with a 785 nm laser, to optimise parameters such as laser
power and scan duration. Initial data can be observed in Figure 3.3 which shows
overlaid spectra for control (blue) and BaP treated (red) whole cells (500–3200 cm⁻¹)
with main peaks of biological interest assigned. Spectra were pre-processed using the
noise filter algorithm in WiRE software (Renishaw plc).
Figure 3.3 Representative noise filtered Raman spectra for control (black) and BaP treated (red) HL60 cells.

The measured spectra contain peaks that have been previously reported for biological samples. Throughout this project, all spectroscopic assignments for both Raman and FTIR spectroscopy were cross-referenced against a number of literature sources, which include [74, 76, 140, 159] for Raman band assignment here.

3.4.2.1 Raman Pre-Processing Considerations

Initial HL60 control cell maps were analysed after spotting onto CaF$_2$ discs with no fixation. Figure 3.4 shows data obtained for a representative single HL60 control cell.
Figure 3.4 Raman whole cell extracted images (500–1800 cm⁻¹) for a typical HL60 control cell showing effects of data pre-processing. White light image is shown in (a). Heat maps portrayed in (b) and (c) depict regions of high intensity in red with lower intensity in blue. The spectrum (d) is an average spectrum which has undergone noise filtering and background subtraction depicting all biological material across the whole cell shown in (c).

Figure 3.4 (a) shows the white light microscopic image in which a number of single cells can be observed and the cell that was chosen for imaging is indicated in the white box. Cells were observed to be on average 16-18 µm in size. Figure 3.4 (b) shows the Raman image (post cosmic ray removal) extracted for the region 500 – 1800 cm⁻¹ which depicts the majority of biological contents of the cell. This is confirmed by the contrast in cell against its background media from which it can easily be distinguished. Raw data was then subjected to pre-processing algorithms which can improve the quality of spectra and thus quality of image produced. Figure 3.4 (c) depicts the same cell after noise filtering and background subtraction. The noise filter function employs PCA to filter out the noise and thus gives an apparent enhancement of signal from the cell against background noise, in a similar fashion to smoothing. For background subtraction, 75% of the background spectrum was subtracted from each spectrum of
the whole image data file in each instance. There is wide scope for Raman and FTIR data pre-processing and bio-spectroscopy studies reported in the literature span a variety of approaches which are summarised in a recent review [278].

Data pre-processing algorithms were thus explored at length for Raman data presented in this thesis to ensure that pre-processing was fit for purpose. The data analysis workflow employed for final Raman images presented is described in Figure 4.2.

### 3.4.2.2 Raman Excitation Lasers: 785 nm and 532 nm

Two lasers were employed for Raman work in this project, a 785 nm laser and a 532 nm laser. Initially the 785 nm laser was selected as a non-destructive, biological friendly wavelength to routinely probe cells due to it reportedly inflicting less cell damage and having reduced background fluorescence than a 532 nm laser. Biological signal was observed to be strong in the fingerprint region between 600 and 1800 cm\(^{-1}\) when cells were irradiated at 785 nm but a weaker signal was obtained for the high wavenumber lipid region between 2900 and 3100 cm\(^{-1}\) (Figure 3.3 illustrates representative spectra obtained at 785 nm showing low signal in the high wavenumber region). Conversely, the 532 nm laser gave enhanced resonance for lipid C-H vibrations observed in the high wavenumber region (which can be observed in Figure 3.8). Data are reported in section 4.3.7 for formalin fixed cells probed with a 532 nm excitation wavelength.
3.4.3 Cell Fixation Method Development for Raman Microspectroscopy

Initial analysis of cells described thus far has been on dehydrated cells spotted onto substrate. Cell fixation was subsequently investigated in an attempt to maintain the integrity of cells during analysis. The merits and disadvantages of cell fixation are discussed in section 1.4.2.2, and here both methanol and formalin fixatives have been studied.

3.4.3.1 Methanol Fixation Protocol

Cell suspension from the washed cell pellet described in section 3.4.1.2 was spotted onto CaF₂ discs and excess suspension was drawn off the disc to leave a sparse monolayer of cells remaining. Once the surface of the disc had dried (after approximately 5 minutes) a large drop of cold methanol (from a 1 mL pipette) was spotted onto the surface to completely cover the visible cell smear. The methanol was left on the disc surface for 5 minutes then removed by a flicking movement to shake off any excess. The cells were now considered fixed in methanol by this procedure.

Figure 3.5 shows a comparison of the white light images obtained by microscopy for HL60 control cells with no chemical fixing and post methanol fixation.
A visual inspection of the images suggests increased heterogeneity in the morphology of the methanol fixed cells, in particular there appears to be a swelling of the cell membrane causing cell enlargement and blebbing.

### 3.4.3.2 Formalin Fixation Protocol

Cell suspension from the washed cell pellet described in section 3.4.1.2 was spotted onto CaF$_2$ discs and excess suspension was drawn off the disc to leave a sparse monolayer of cells remaining. Once the surface of the disc had dried, excess formalin (10% formalin solution, buffered pH 7.0, VWR) was spotted onto the surface to fix the cells. The formalin was left on the disc surface for 15 minutes then removed by gently shaking off the excess and the fixed cells on the surface of the disc were subsequently washed with 3 x PBS, 500 µL and air dried prior to immediate analysis. Figure 3.6 shows a comparison of the white light images obtained by microscopy for HL60 control cells with no chemical fixing and post formalin fixation.
A number of formalin fixed cells on substrate appeared to show residual formalin crystals on the surface, which were not removed by washing. This was noted and is shown in Figure 3.7 but not considered a problem for Raman microspectroscopy due to the confocality of the technique. Prior to single cell mapping, a depth profile was acquired in which spectra were obtained at 1 µm intervals spanning 12 µm of the z axis. Subsequently, a depth was selected for mapping according to the spectra that gave the strongest biological signal for each cell.

A series of Raman maps were acquired for methanol and formalin fixed cells and a comparison of spectra obtained did not appear to show a significant difference between cells analysed in both states. Figure 3.8 shows representative raw spectra obtained for
HL60 drug treated cells at 532 nm, overlaid for Raman maps extracted at 2955 cm\(^{-1}\) representative of the asymmetric methyl stretch of lipids.

*Figure 3.8* Representative raw Raman spectra for HL60 BaP treated cells obtained at 532 nm excitation.

The strong Raman signal at 532 nm for C-H stretches assigned to lipid species in the high wavenumber region suggests that any changes in cellular lipids with drug treatment could potentially be readily observed when employing this excitation wavelength. When considering the most appropriate fixative for this study, the numerous reports of methanol potentially causing lipids to migrate and leach from cells was a concern and this phenomenon explains the use of formalin as the fixative of choice for a number of cell studies discussed in Chapter 1. Furthermore, Raman spectra obtained from formalin fixed cells (when compared to methanol-acetic acid fixed cells) have been reported to show the closest resemblance to live cells [83]. For this reason, Raman results presented in Chapter 4 are derived from cells that were fixed in formalin according to the protocol described in section 3.4.3.2.
3.5 Method Development for Cell Extract Analysis

The analysis of cell extracts contributes a significant part to this project. Data presented for ToF-SIMS, Raman spectroscopy and LC-MS are all obtained from exactly the same cell extracts to afford the most reliable way of data interpretation in a multi-disciplinary setting. Extracts were initially obtained from collaborators (for ToF-SIMS method development) and subsequently generated using extraction protocols described in this section.

3.5.1 Preparation of Cell Extracts for ToF-SIMS Analysis

Silicon wafers (Agar Scientific Ltd, UK) were prepared by sonicating for 10 minutes each in hexane, chloroform, water and methanol sequentially to remove any residual contamination, and allowed to dry at room temperature in a clean vessel. All solvents were HPLC grade (Sigma Aldrich, UK).

For this initial sample preparation study, dried HL60 and K562 cellular lipid extracts, 60 µL (equivalent to approximately 9 mg of wet mass containing around 3-5 million cells) were obtained from collaborators in the School of Biosciences at The University of Birmingham. Extracts were prepared in Birmingham according to Bligh and Dyer extraction protocol [73], and the lipid extract was stored at -80 °C. Cell extracts were thawed at room temperature and reconstituted in HPLC grade 2:1 methanol:chloroform, 100 µL, and sonicated for 5 minutes prior to deposition.

Samples were prepared by pipetting the cell extracts directly onto the surface of each silicon wafer in varying volumes and allowed to dry at room temperature.

A sample preparation study was carried out which involved pipetting varying volumes of HL60 cell extract onto the wafers (10 µL, 20 µL (in 4 x 5 µL) and 30 µL (in 3 x 10 µL) and comparing results obtained for signal intensity of significant peaks observed in the ToF-SIMS spectra. Each aliquot of sample on the wafer was allowed to dry fully in air before applying the next one to assist in the building up of a homogeneous sample surface on the wafer. Following this, a further study was carried out to determine the merits and procedure for carrying out a de-salting wash of the extracts on the wafers.
For washing, a 0.15 M solution of ammonium formate was prepared in HPLC grade water. Silicon wafers with dried cell extract on the surface were washed in this solution by holding the wafer in the ammonium formate solution for 30 seconds and rinsing in a fresh ammonium formate solution for a further 5 seconds. Wafers were allowed to air dry fully before sample transfer and analysis.

### 3.5.2 Results: ToF-SIMS Sample Preparation Study

Initial analyses employed 2 primary ion beams, \( \text{Au}_3^+ \) and \( \text{C}_{60}^+ \) by way of comparison to determine which projectile was capable of giving the most intense ion yield for cell extracts. Following the selection of a \( \text{Au}_3^+ \) primary ion beam for giving overall increased signal, positive ion spectra were acquired on the BioToF for both control and BaP treated HL60 cell extracts, which were loaded in varying volumes according to the protocol stated in section 3.5.1. Aliquots of 10 µL, 4 x 5 µL (20 µL total) and 3 x 10 µL (30 µL total) were analysed and the peak at \( m/z \) 184 (from the phosphocholine (PC) head group) in the positive ion SIMS spectra was chosen as a representation of total amount of biological material present on the wafer for this comparison (data from further peaks were tabulated but are not shown). Peak areas were extracted for each sample volume and duplicate replicates plotted. Figure 3.9 shows a plot of peak areas against the cell extract volumes prepared onto silicon wafers.

![Graph to show m/z 184 signal as a function of sample loading](image)

**Figure 3.9** Plot to show \( m/z \) 184 peak area as a function of sample loading for HL60 cell extracts.
As well as determining optimum sample coverage on a silicon wafer, a study was performed to determine linearity of SIMS signal with increasing number of primary ion beam shots (equivalent to ion dose). Again peak areas were extracted for peak at \( m/z \) 184 for primary ion beam shots of 1 M, 2 M and 4 M \((n = 3)\) and average peak areas can be seen plotted in Figure 3.10, with error bars representing standard deviation.

![Graph to show \( m/z \) 184 signal as a function of primary ion beam shots](image)

*Figure 3.10 Plot to show \( m/z \) 184 average peak area \((n=3)\) as a function of primary ion beam shots.*

Having determined a suitable sample loading volume of 30 µL (loaded in 3 x 10 µL aliquots) and crudely estimated the linearity of signal with primary ion beam dose, fresh cell extracts were reconstituted and prepared at 30 µL on wafer and were left either unwashed or were washed with 0.15 M ammonium formate solutions as described in section 3.5.1. Analysis was taken from two different regions of each wafer, with 2 M shots each time. Peak areas were extracted from notable peaks of interest that were concurrent across each analysis and tabulated. Despite replicates across each wafer being taken for this analysis, peak areas were not averaged so that reproducibility across the wafer could also be assessed for the washing protocol.

An example raw positive ion ToF-SIMS spectra for BaP treated HL60 cell extracts (post-optimised wash procedure) is displayed in Figure 3.11 \((m/z \ 0–200)\). Peaks at
$m/z$ 43 and 55 were chosen to represent organic material (hydrocarbons, $C_3H_7^+$ and $C_4H_7^+$ respectively), $m/z$ 184 to represent cellular biological material (PC lipid head group) and peaks at $m/z$ 73 ($\text{Si(CH}_3\text{)}_3$ polymer unit) and 147 ($\text{Si}_2\text{C}_5\text{H}_{15}\text{O}$) to look at the effects of washing on the common contaminant polydimethylsiloxane (PDMS).

![Figure 3.11 Positive ion ToF-SIMS raw spectra showing $m/z$ 0 – 200 for HL60 BaP treated cell extract after washing with 0.15 M ammonium formate according to final wash procedure developed.](image)

Figure 3.12 and Figure 3.13 show the results for the comparison sample preparation washing study for HL60 control cell extracts and HL60 BaP treated cell extracts respectively.
Figure 3.12 Plot to show ToF-SIMS peak signals for HL60 control cell extracts as a function of washing with 0.15 M ammonium formate solution.

Figure 3.13 Plot to show ToF-SIMS peak signals for HL60 BaP treated cell extracts as a function of washing with 0.15 M ammonium formate solution.

An increase in area obtained for peaks m/z 184, 43 and 55 for both control and BaP treated cell extracts after washing can be observed. This indicates that the washing protocol for 0.15 M ammonium formate has a positive effect of enhancing the ionisation efficiency of molecular species of interest, most likely by reducing ion
suppression from background salts such as potassium and ammonium which are washed away during the procedure. A further outcome is the small but desirable apparent decrease in the PDMS signals at $m/z$ 73 and 147 observed for both cell extracts after washing.

Duplicate analyses across a single silicon wafer show the same trend for each molecular species indicating that the sample preparation method in its entirety has a reproducible effect across the surface coverage, although there are some differences in peak areas observed which indicates a possible non-homogeneous coverage of cell extract on the wafer.

In summary, the final protocol for subsequent large scale study of cell extracts (described in section 3.5.4) by ToF-SIMS was a sample loading volume of 30 µL (loaded in 3 x 10 µL aliquots) followed by washing with 0.15 M solution of ammonium formate solution for 30 seconds and a second rinse in fresh ammonium formate solution for a further 5 seconds (for which an example spectrum is shown in Figure 3.11).

3.5.3 Optimisation of Cell Extraction Procedure for Metabolomic and Lipidomic Studies

The aim of this part of the project was to identify an efficient extraction protocol which provided sufficient biomass for the multidisciplinary analysis of BaP treatment on the metabolome and lipidome of AML cells. This experiment was designed to provide enough biomass of cell extract to allow comparative studies across multiple analytical techniques from one single large-scale 24 hour drug treatment which is described in section 3.5.4.

At 24 hours, control and drug treated cells were immediately quenched with cold methanol to stop the biochemical processes occurring in the cell. This procedure has been previously described [279] and is essential when comparing multiple replicates with each other.

Post-quenching, cells were extracted immediately for lipidomic and metabolic profiling. Three different extraction protocols were investigated in this study and
Figure 3.14 shows the final extraction protocol developed and selected for experimentation within this project, which was adapted from the widely reported Folch [280] and Bligh and Dyer [73] extraction methods for lipids, and optimised for this analysis.

**Figure 3.14** Workflow developed for AML cell extraction, adapted from the Folch and Bligh and Dyer methods for lipid extraction.

### 3.5.4 Experimental Procedure for Large Scale Metabolomics and Lipidomics Study Employing ToF-SIMS, UHPLC-MS and Raman Spectroscopy for the Analysis of Cell Extracts

A key aim of this project was to be able to obtain high quality analytical data from multiple instrumental platforms for the investigation of BaP treatment at a cellular level. To enable this, one large-scale cell-drug treatment protocol was developed which incorporated 6 replicates for each of control and BaP treated cell populations. The protocol described in Figure 3.15 ensured that cell populations at 24 hours post-
treatment were large enough to provide enough biomass (when extracted according to the protocol described in Figure 3.14) for analysis across multiple analytical techniques on the same cell extract.

![Diagram of cell extraction process]

Repeted for each cell line, HeLa and K562 in parallel

Pool 45 mL from each (270 mL total) Count 2 aliquots from each pool

Aliquot 16 mL from pooled flask into each of 15 flasks (1 in 3 dilution = cell population ~ 0.5 M/mL each flask)

Fill 15 T-225 flasks with 32 mL fresh media

n = 6 DRUG treated n = 6 CONTROL n = 3 routine cell culture

Add 48 µL 0.5 M Bez in DMSO Add 48 µL 5 mM MPA in ethanol
Add 48 µL DMSO Add 48 µL Ethanol

Culture as normal

Quench and extract according to extraction protocol

Centrifuge 12 flasks for 5 min. @ 1000 g. Carefully discard majority of supernatant and reconstitute in 8 mL PBS

Incubate @ 37°C / 5% CO₂ for 24 hours

Figure 3.15 Preparation and treatment of cells in culture for large scale metabolomics experiment.

3.5.5 Preparation of Cell Extracts for Raman Analysis

Stainless steel plates were selected as the Raman substrate for cell extract analysis as the larger surface area of the plates allowed for the deposition of 4 cell extracts in each corner of the plate thereby increasing throughput to allow for replicate analysis in as short a time frame as possible. Plates were cleaned thoroughly with ethanol twice prior to analysis. Re-constituted non-polar (chloroform) cell extracts were deposited onto plates as 5 µL drops with a micropipette to ensure as small a drop size as possible and extracts were deposited and analysed in a randomised fashion. All six replicates
prepared according to the protocol in section 3.5.4 were analysed for control and drug treated cell extracts for both HL60 and K562 cell lines.

Despite the small volume/drop size employed, dried spots of cell extract are known to exhibit the ‘coffee ring’ effect, whereby sample deposited concentrates further at the outer edges of the deposited droplet with drying [281]. There is little that can be done to eliminate this effect other than to keep the spot size deposited as small as possible. In order to address this heterogeneity in concentration across the deposition, triplicate analyses were taken from inside the coffee ring and triplicates from the outside of the ring as illustrated in the example extract in Figure 3.16. Data for this experiment are presented in section 4.3.6.

![Figure 3.16 Images to show deposition of cell extract (outside and inside dried spots) for Raman analysis.](image)

Summary of extracts analysed:

2 cell types HL60 and K562, drug treated and control conditions, non-polar (chloroform) extracts, 6 biological reps equalling a total of 24 samples. Data acquired from a minimum of 6 areas of each deposited extract (3 from outer edge of dried extract, 3 from inside dried extract spot as a minimum). A total of > 144 Raman spectra acquired for analysis.
3.6 Conclusions

Cell culture models for HL60 and K562 cell lines were established in house to enable the mode of action of BaP at a cellular level to be studied and cell preparation methodologies were explored to ensure samples were fit for subsequent analysis with the specific analytical techniques employed in this project. Formalin fixation of whole cells for Raman microspectroscopy and ammonium formate washing of cell extracts prior to ToF-SIMS analysis were found to be suitable cell preparation methods. An efficient protocol for the extraction of intracellular metabolites and lipids for large scale metabolomics investigations was established which ensured a yield of adequate biomaterial for analysis across multiple analytical platforms. As well as cell preparation considerations, there is large scope for manipulation of instrumental parameters in the quest to yield high quality data and parameters were optimised here for ToF-SIMS (primary ion dose) and Raman microspectroscopy (laser wavelength and acquisition time).
Chapter 4 A Spectroscopic Toolbox for Probing the Action of BaP Drug Therapy in AML Cells

4.1 Introduction

Probing drug-cell interactions with spectroscopic techniques has become increasingly popular and can contribute to the understanding of the mode of action of the drug at a cellular level as discussed in section 1.5.2. The majority of spectroscopic cellular studies reported to date are on cells which have been chemically fixed [77, 84, 141, 142, 147, 165] and are therefore in a dehydrated state. Whilst fixation has its place and is widely accepted in the spectroscopic community, there are limitations to the useful information that can be derived from this approach. Problems encountered include a change in the conformation of DNA caused by dehydration and also fixatives which interfere with inherent biochemical signature of the cell meaning that a cautious approach to data interpretation must be taken.

There are significant benefits to be had from probing a cell in its hydrated state for a more accurate view of the nature of intracellular biochemical species in the physiological condition. Employing the superior properties of a synchrotron beam enables such in vivo analysis to be performed in real time with cells remaining in their growth medium, thereby eliminating the requirement for any time consuming and potentially chemically detrimental sample preparation. The ability to probe a single cell, one by one is extremely desirable, ensuring information obtained is specific to the living cell in question, rather than that averaged over a heterogeneous cell population, which can also include cell debris and cells that are undergoing cell death.

This chapter presents a spectroscopic interpretation of the action of BaP drug therapy in two AML suspension cell lines, HL60 and K562 as discussed in Chapter 3. Cells reported here were analysed live (synchrotron FTIR), dehydrated (ATR-FTIR), as extracts (Raman) and fixed (Raman); a diversity which contributes to a broad spectroscopic picture of drug action and provides an in situ interpretation of the effect of BaP treatment on AML cell. The large number (>800) of single cells reported here gives further confidence to the power of the techniques, as described by statistical models for data interpretation.
4.2 Materials and Methods

4.2.1 Preparation of Cells and Measurement Details

Cells were cultured *in situ* at the Australian Synchrotron (Clayton, Victoria) for one week prior to Synchrotron and ATR experimentation. Cells were cultured *in situ* in the Manchester Institute of Biotechnology for Raman experimentation.

Control and BaP treated cell suspensions of HL60 and K562 in growth media were removed from the incubator at 24 hours and centrifuged at 500 $\times$ g for 5 min. The supernatant was discarded, leaving a small volume of liquid in the cell pellet.

*For Synchrotron-FTIR analysis*, minimal sample preparation was required; the solution was agitated gently to release the cell pellet with a pipette tip and 2 $\mu$L of the concentrated cell suspension was removed and gently placed on a CaF$_2$ disc seated in a purpose-built microfluidics sample accessory device as previously described [282]. A photographic image of the microfluidics disc seated in the sample holder can be seen in Figure 4.1. The device held cells in position and ensured cells were in a fully hydrated state for analysis, while the microfabricated gasket thickness was selected to ensure the cells were gently held between the upper and lower windows of the chamber and remained stationary during measurements.

![Microfluidics sample accessory device](photograph-courtesy-of-Robbin-Vernooij)

*Figure 4.1 Microfluidics sample accessory device for use in the synchrotron FTIR Bruker Hyperion 2000 microscope (photograph courtesy of Robbin Vernooij).*
For ATR-FTIR analysis, the cell pellet was washed twice with physiological saline solution with centrifugation performed between washes. The remaining cell pellet was treated as stated above with 2 µL of the concentrated cell suspension being placed directly onto the diamond ATR window and allowed to air dry thoroughly before recording spectra. Cells were in a dehydrated state for analysis.

For Raman extract analysis, a large scale experiment was performed as described in section 3.5.5 with 6 replicates for control and drug treated flasks for each cell line. After 24 hours, cells were extracted according to the protocol described in section 3.5.3.

For Raman mapping analysis, the cell pellet was re-suspended in PBS and washed according to the protocol described in section 3.4.1.2. The washed cell pellet was spotted onto a CaF$_2$ disc and then fixed in formalin according to the protocol described in section 3.4.3.2.

**Synchrotron-FTIR Measurements:** Synchrotron-FTIR measurements (S-FTIR) were recorded using the FTIR microspectroscopy beamline at the Australian Synchrotron on a Bruker Hyperion 3000 IR microscope (Bruker Optics, Germany) according to the parameters described in section 2.3.5.

**ATR-FTIR Measurements:** ATR-FTIR analyses were performed on an Agilent 4500a portable FTIR spectrometer (Agilent Technologies, USA) according to the experimental details described in section 2.3.7.

**Raman Measurements:** Raman spectroscopic analysis was performed using a Renishaw inVia Raman microscope (Renishaw plc., Wooton-under-edge, UK) with excitation laser of 532 nm according to experimental parameters described in section 2.4.4.
4.2.2 Spectroscopic Data Processing

4.2.2.1 FTIR Data Analysis

Pre-processing of the spectra was carried out in MATLAB (The Mathworks Inc. Natik, U.S.A.). For both cell lines, all spectra acquired were used in the data modelling. Raw spectra were corrected for baseline fluctuations and any scaling effects due to different pathlengths, using extended multiplicative signal correction (EMSC) [283]. To avoid introducing any further artefacts into the data set, no further pre-processing algorithms were employed. Following EMSC correction, the spectra were cropped to the specific regions of interest to be analysed (3100 – 2800 cm\(^{-1}\) for the lipid region and 1800 – 1100 cm\(^{-1}\) for the fingerprint region). ATR-FTIR spectra were processed as the second derivate to enhance spectral features. All spectral assignments have been made according to widely cited literature references [57, 143, 251].

Statistical Analysis:

Orthogonal Partial least squares-discriminant analysis (oPLS-DA).

Prior to the oPLS-DA, the synchrotron single cell spectra data matrix was pre-processed in the following order: (a) Savitzky-Golay filtering based second derivative (order: 2, window: 9 points) was applied first, followed by (b) standard normal variate spectra normalization, following which (c) Pareto scaling variable pre-processing was carried out and finally (d) mean centering was performed. The synchrotron nuclear spectra data were subjected to the same spectral pre-processing regime with the exception of step a) for which baseline correction with 15 point smoothing was performed prior to normalisation. Data processing was performed employing MATLAB from using the routines available on the PLStoolbox package from Eigenvector Research Inc. (Manson, USA).

Bootstrapping Validation.

In order to validate the models generated by partial least squares-discriminant analysis (PLS-DA) bootstrapping with replacement was used for the synchrotron FTIR spectra. In this resampling process, 1000 bootstraps were undertaken and in each bootstrap each training set contained on average 63.2% of all samples, and each test set included the remaining 36.8% samples. Statistics were then generated for the 1000 models in
terms of correct classification rates and confusion matrices. This technique gives well represented estimates of the average model and is therefore less likely to be biased [97].

4.2.2.2 Raman Data Analysis

For cell extracts: Raman spectra for control and drug treated cell extracts were pre-processed in WiRE software (Renishaw plc.) for cosmic ray removal and further manipulated according to in-house scripts in MATLAB. Spectra were baseline corrected then normalised according to the standard normal variate (SNV) method. Hierarchical cluster analysis (HCA) was performed to select the most reproducible spectra in each class of data set obtained from both inside and outside the dried extract deposits and PCA was then performed on two sets of data: triplicate spectra obtained from inside each droplet, and triplicate spectra from outside.

For single cell maps: Raman maps for 5 control cells and 6 drug treated cells were pre-processed in WiRE software (Renishaw plc.) for cosmic ray removal and further manipulated according to in-house scripts in MATLAB. Signal to noise ratio (SNR) assessment was performed using Morphological Scores (MoS) [284]. Fuzzy c-means clustering algorithm (FCM) was employed to image features of the cell and subsequently PLS-DA (using baseline corrected and SNV normalised spectra) was able to differentiate between the two cell classes. The PLS-DA models were validated by using a k-fold cross double cross-validation (CV) [263] where k is the number of cells, i.e. the spectra of an individual cell were held out as the test set and the model was built on the remaining spectra. The model was then applied to the test set spectra to predict their class membership. This procedure was repeated k times until each cell had been held out as test set for once. Within each CV procedure, the number of PLS-components was chosen by employing another k-fold CV on the training set and the one that generated the best results on the internal validation set was chosen.

A summary of Raman image data analysis can be seen in Figure 4.2.
4.3 Results and Discussion

4.3.1 Mean Spectra Obtained from S-FTIR and ATR Analysis

Data are presented here for both ATR and S-FTIR analyses. Cells analysed by ATR-FTIR were sampled from a heterogeneous cell pellet which predominately consisted of viable cells, but was also likely to contain cells in varying stages of necrosis and cell debris from cells which may have undergone apoptosis in culture. A significant advantage of ATR-FTIR is that it is easily accessible, portable and inexpensive; however, cells were dehydrated, thus limiting the biochemical information that can be derived from resulting spectra. Exploiting the superior properties of the synchrotron beam (S-FTIR) [285] allows transmission of light through aqueous environments and the two windows of the sample holder, thus enabling the analysis of live single hydrated cells. Cells were clearly visible under the microscope to confirm their viability and a randomised selection of single cells was selected from each aliquot of cells in this study. In total, 441 individual HL60 cells and 385 K562 cells were analysed by S-FTIR. Averaged spectra for HL60 and K562 cells analysed with both techniques are shown in Figure 4.3. The C-H stretching region (3100-2800 cm\(^{-1}\)) as well as the cell fingerprint region (1800-900 cm\(^{-1}\)) are overlaid for control and 24 hour
BaP treated cell spectra for both ATR-FTIR and synchrotron-FTIR analysis of each cell line, respectively. To enhance spectral features observed in the high wavenumber region (where prominent lipid bands are observed), second derivative spectra were plotted and are shown in Figure 4.4.

Figure 4.3 Normalised, mean spectra from 9 replicates for ATR analysis of HL60 and K562 control and BaP treated cells and 3 replicates for each of 2 trials for S-FTIR analysis, incorporating a total of 441 individual cells (HL60) and 385 individual cells (K562).
Figure 4.4 Normalised, mean second derivate S-FTIR spectra from the high wavenumber region for control and BaP treated K562 and HL60 cells.

Biological interpretation of spectral features is discussed in section 4.3.3.

4.3.2 Statistical Analysis and Validation of S-FTIR and ATR Spectra

Orthogonal partial least squares-discriminant analysis (oPLS-DA) was performed using the 3059-2715 cm\(^{-1}\) and the 1819-1223 cm\(^{-1}\) regions for the synchrotron data and using the 3133-2806 cm\(^{-1}\) and the 1840-862 cm\(^{-1}\) regions for the ATR data for control and 24 hour drug treated cell spectra (regions selected for biological material present in the fingerprint region and high wavenumber lipid region, within the stable range of the detector). Figure 4.5 shows the scores of the first oPLS-DA latent variable (LV) for each cell type, K562 (Figure 4.5a) and HL60 (Figure 4.5b) with LV1 clearly distinguishing drug treatment (red) from control cells (green). Although the LV1 calibration values are not good indicators of the classification performance due to overfitting issues, they are necessary to interpret the direction of the oPLSDA first loading vector. A more reliable estimation of the model quality was obtained by the
cross validation results (Venetian blinds, 10 splits), which indicated a good classification performance in the four cases, given cross validation error rates of 15% (4LVs), 0% (3LVs), 15% (2LVs) and 11% (1LV).

Figure 4.5 Scores plots for the first oPLS-DA component for S-FTIR spectra of control (green) and BaP treated (red) cells. Each point represents a K562 single cell (a) and HL60 single cell (b).

The same trend was observed for ATR spectra across 9 replicates for each cell line as shown in Figure 4.6.
Further validation of the oPLS-DA model was obtained by using partial least squares discriminant analysis (PLS-DA). The classification model for the PLS-DA computed across the spectral region 3100 – 1100 cm$^{-1}$ for all S-FTIR cells analysed, with no outliers removed, can be seen in Figure 4.7. Bootstrapping was performed 1000 times to permute whether the cell was drug treated or not. The output shows two discrete distributions; a ‘real’ (in blue) and ‘null/random’ (in red) model, these models have good correct classification rates (CCR) of an average of 94.36% for K562 and 78.14% for HL60 cells.

*Figure 4.6 Scores plots for the first oPLS-DA component for ATR spectra of control (green) and BaP treated (red) cells. Each point represents a replicate of cell pellet of K562 cells (a) and HL60 cells (b).*
Figure 4.7 Classification model for PLS-DA of S-FTIR analysis of whole cell populations for K562 (a) and HL60 (b) cells showing classification rates for real (blue) and random (red), with a CCR of 1 being 100% correct classification. The embedded tables are the average confusion matrices with rows representing predicted classification and columns representing experimental.

When comparing data acquired between the two cell lines, it is apparent that HL60 cells show a greater distribution in spectra than K562 cells. This is understandable, given that the aperture size employed for analysis and the average size of HL60 cells are comparable, thereby a portion of HL60 cells presenting at smaller than 10 µm in diameter were likely to contribute as outliers (arising from the non-cell region sampled from smaller cells).

### 4.3.3 Changes in Vibrations Arising from Lipid Species with BaP Treatment

A visual inspection of the ATR spectra (Figure 4.3) reveals an apparent change in the intensity of signals in the high wavenumber (3100 – 2800 cm\(^{-1}\)) region, predominated by the \(v_{as}(\text{CH}_3)\), \(v_{as}(\text{CH}_2)\), \(v_{s}(\text{CH}_3)\) and \(v_{s}(\text{CH}_2)\) stretches arising from lipid species within the cell, and located at 2955, 2920, 2870 and 2850 cm\(^{-1}\). Specifically, it appears that there is an overall increase in signal intensity for both the asymmetric and symmetric methylene \(v(\text{CH}_2)\) stretches with BaP treatment and a corresponding decrease in intensity for the methyl \(v(\text{CH}_3)\) asymmetric and symmetric stretches. This trend is also observed in the synchrotron data as shown in Figure 4.4 in which the

<table>
<thead>
<tr>
<th></th>
<th>K562</th>
<th>BAP</th>
<th>CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAP</td>
<td>0.9322</td>
<td>0.0678</td>
<td></td>
</tr>
<tr>
<td>CON</td>
<td>0.0463</td>
<td>0.9537</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>K562</th>
<th>BAP</th>
<th>CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAP</td>
<td>0.7660</td>
<td>0.2340</td>
<td></td>
</tr>
<tr>
<td>CON</td>
<td>0.2027</td>
<td>0.7973</td>
<td></td>
</tr>
</tbody>
</table>
second derivative spectra show a peak minimum corresponding to the maxima intensity in the underivatised spectrum exhibiting an overall increase in methylene intensity. The fingerprint region of the spectra is dominated, as expected, by the amide bands I (1639 cm\(^{-1}\)) and II (1535 cm\(^{-1}\)) arising from \(\nu(C=O)\) stretching mode (Amide I) and \(\delta(N-H)\) bending and \(\nu(C-N)\) stretching modes of peptide linkages (Amide II), which typically represent total cellular protein. The fingerprint region also contains more subtle features, often appearing as small inflections of a peak more prominent in second derivative spectra, for example a change in DNA conformation can be determined by considering the position of the asymmetric phosphate stretching vibration observed at 1220 cm\(^{-1}\) in B-DNA and 1235-40 cm\(^{-1}\) in A-DNA, which is discussed later.

A consideration of the first loadings vector from the o-PLS-DA confirms that the main contribution to the separation observed in the scores is from the aforementioned methylene moieties, with prominent features observed in the high wavenumber region at 2920 cm\(^{-1}\) and 2850 cm\(^{-1}\). The first loadings vector from the oPLS-DA, with asymmetric and symmetric \(\nu(C-H)\) stretches annotated can be seen in Figure 4.8. Here, the loadings from both synchrotron and ATR spectra are overlaid showing consistent loadings and good correlation between data sets obtained with these two techniques. Loadings shown for the synchrotron spectra are noisier than spectra obtained with ATR likely due to differing amounts of biomaterial analysed. Synchrotron data are averaged over a cell population of a few hundred cells for each cell line, however thousands of cells would have been present in the aliquot sampled and dried on the ATR element. Despite cells being hydrated for S-FTIR analysis and dehydrated in the case of ATR, because the effect of the drug treatment on DNA appears to be negligible and the predominant spectral changes arise from disruptions of lipid bands within the spectra, both techniques produce concordant results. There is less of an observable contribution to the discrimination between control and drug treated cells in the lower wavenumber spectral region of the loadings (data not shown).
Figure 4.8 oPLS-DA loadings vectors overlaid for S-FTIR (red) and ATR-FTIR (blue) loadings in the high wavenumber lipid region, for K562 (a) and HL60 (b) cells.

Figure 4.9 oPLS-DA loadings vectors overlaid for S-FTIR (red) and ATR-FTIR (blue) loadings in the fingerprint region, for K562 (a) and HL60 (b) cells.

The apparent change in methylene to methyl ratios in data acquired for drug treated cells was further explored by extracting spectral peak areas for all single cells analysed with S-FTIR. Figure 4.10 shows peak area ratios for asymmetric and symmetric stretches of methylene and methyl moieties for K562 (Figure 4.10a) and HL60 (Figure 4.10b) cell populations. The box plots include all synchrotron spectra acquired, each
point representing a cell; the spread therefore representing inter-cell variability across replicates spanning two trials. Data clearly show an increase in the methylene to methyl ratio with drug treatment for both asymmetric and symmetric stretches. A concurrent trend is also observed for ATR spectra acquired from cell pellets from 9 replicates, with data shown in Figure 4.11.
Figure 4.10  Box and whisker plots for significant changes (p < 0.05, see Table 4.1) in peak area ratios of methylene to methyl stretching vibrations between control and 24 h drug treated cells analysed by S-FTIR. a. shows peak area ratios 2920:2955 cm\(^{-1}\) for asymmetric C-H stretching and 2850:2870 cm\(^{-1}\) for symmetric C-H stretching for K562 cells and b. shows the same ratios for HL60 cells. The median of each box plot can be compared to describe the changes in lipid peak area ratios with drug treatment; with the maximum values, minimum values and outliers showing the spread in inter-cell variability within population sizes of 385 cells for K562 and 441 cells for HL60. Extreme outliers (n<10) were removed from the plot for visual appearance, however all cells analysed were included in the data analysis.
Figure 4.11 Box and whisker plots for significant changes (p < 0.05, see Table 4.1) in peak area ratios of methylene to methyl stretching vibrations between control and 24 h drug treated cell pellets analysed by ATR. a. shows peak area ratios 2920:2955 cm⁻¹ for asymmetric C-H stretching and 2850:2870 cm⁻¹ for symmetric C-H stretching for K562 cells and b. shows the same ratios for HL60 cells. The trend observed is concurrent with synchrotron FTIR single cell data shown in Figure 4.10.
PLS-DA confirms the observed grouping according to methylene to methyl peak area ratios for BaP treated and control cells and p-values for each are detailed in Table 4.1. The CH$_2$:CH$_3$ ratios proving to be significant in both cell lines ($p < 0.05$) for both synchrotron and ATR FTIR analyses. (Ratios reported for 1240 cm$^{-1}$:1220 cm$^{-1}$ are discussed in section 4.3.4).

<table>
<thead>
<tr>
<th>S-FTIR (raw, heights)</th>
<th>K562</th>
<th>HL60</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_2$/CH$_3$ from 2920 cm$^{-1}$ : 2955 cm$^{-1}$</td>
<td>1.0757×10$^{-20}$</td>
<td>6.2607×10$^{-5}$</td>
</tr>
<tr>
<td>CH$_2$/CH$_3$ from 2850 cm$^{-1}$ : 2870 cm$^{-1}$</td>
<td>2.4078×10$^{-13}$</td>
<td>2.1×10$^{-3}$</td>
</tr>
<tr>
<td>1240 cm$^{-1}$ : 1220 cm$^{-1}$ ratio</td>
<td>0.6660</td>
<td>0.7681</td>
</tr>
<tr>
<td>ATR-FTIR (2nd deriv, heights)</td>
<td>K562 (HL60)</td>
<td></td>
</tr>
<tr>
<td>CH$_2$/CH$_3$ from 2920 cm$^{-1}$ : 2955 cm$^{-1}$</td>
<td>0.0259</td>
<td>0.0148</td>
</tr>
<tr>
<td>CH$_2$/CH$_3$ from 2850 cm$^{-1}$ : 2870 cm$^{-1}$</td>
<td>0.0539</td>
<td>0.0331</td>
</tr>
<tr>
<td>1240 cm$^{-1}$ : 1220 cm$^{-1}$ ratio</td>
<td>0.1981</td>
<td>0.7738</td>
</tr>
</tbody>
</table>

*Table 4.1  Student t-test p-values to quantify the group separation in the PLS-DA scores of each cell line analysed by S-FTIR and ATR-FTIR.*

A biological interpretation of these data is interesting since an increase in CH$_2$:CH$_3$ ratio indicates that lipids in drug treated AML cells show an overall change in saturation state with BaP treatment when compared with control cells. Lipid composition in drug treated cells may consist of a greater number of acyl chain double bonds at a chain end position or data could suggest increased lipid saturation in mid acyl chain positions. To gain further insight into this picture, Raman microspectroscopy was employed to image control and drug treated single cells and data for this is presented in section 4.3.7.

### 4.3.4 Investigating the Effect of BaP on Cellular DNA with S-FTIR

Drawing conclusions on drug mode of action based on a consideration of just the lipid region of the spectra can be somewhat short-sighted. Similar changes in lipid biochemistry can also be observed in cells undergoing various stages of cell death from necrosis through to apoptosis. The increase of lipids in this case is reportedly due to a
number of cell membrane processes such as blebbing and vesicle formation, which can give rise to an increase in CH$_2$ absorption observed in the spectra [286]. This implies that many drug-cell interaction studies, which report lipid changes, might in fact just be monitoring the death of a cell due to drug treatment and this has recently caused some concern in the spectroscopy community.

In this study, the previously described BaP treatment is known to further differentiate HL60 cells and induce apoptosis in K562 cells, but not within the 24 hour prescribed drug treatment [213]. When a cell enters apoptosis, the DNA shifts to a more disordered state, and the DNA signal in an infrared spectrum typically decreases [52]. A shift in the DNA band can also be observed in cells which have been treated with a drug known to intercalate with DNA, whose mode of action is such that base pairing is interrupted and unravelling from orderly B-DNA to the more disordered A-DNA form occurs [54].

Again, care must be taken in interpreting the data; do the spectra show a drug induced biochemical change other than cell death?

Here, the major DNA peaks have been explored to assess if one of the targets of BaP could in fact be DNA and to confirm that what was being monitored in this study was not simply cell death. One of the vibrations as a marker for the latter hypothesis is the carbonyl vibration for protein linking in the amide I band at 1657 cm$^{-1}$ which has been reported to significantly decrease with apoptosis [277]. A visual inspection of the spectra in Figure 4.3 indicates no obvious change in amide I signal intensity with BaP treatment. Perhaps more diagnostic though is the position of the asymmetric phosphate stretch of DNA which is reported to shift from 1220 cm$^{-1}$ to 1240 cm$^{-1}$ with a conformation change from B-DNA to A-DNA [79], a phenomenon which can occur during drug treatment and also during cell dehydration. Fully hydrated DNA exists in the B-conformation and one of the main markers for a B- to A-DNA transition is the shift of $\nu_{\text{asym}}$(PO$_2^-$) to a higher wavenumber. As such, this change can only be confidently observed in synchrotron spectra for which cells are hydrated during data acquisition. Figure 4.12 shows peak area ratios between control and drug treated cells for the asymmetric phosphate stretch of DNA, to look for a shift in peak maxima indicating a change in DNA conformation. For this scenario we have a null hypothesis, with DNA signals not responding in a manner indicative of cell death.
The absence of a significant change in this peak area ratio is further confirmed in Table 4.1, with the 1240:1220 peak ratios showing very high $p$-values, thereby not being significant. This information, coupled with the apparent absence of change for the amide I peak intensity strongly suggests that DNA is not a target of BaP therapy. This is of key significance from an *in vivo* biological perspective, because it could explain why the combination drug BaP showed little or no haematological toxicity to healthy cells in a small cohort of patients recruited for clinical trial [193]. If BaP was interfering with the DNA of a cell, one would expect that to be a fairly non selective process and thus signs of DNA disorder in control cells might be expected.

4.3.5 S-FTIR Analysis of Single Nuclei Extracted from Control and Drug Treated AML Cells

Within the last decade it was demonstrated for the first time that S-FTIR spectra could be obtained from isolated nuclei [287]. Here a preliminary study was carried out to investigate whether effects of BaP treatment could be observed in AML cells on a nuclear level using S-FTIR.
4.3.5.1 Nuclei Extraction Methodology

Cell nuclei were extracted from HL60 and K562 control and drug treated cells using a Nuclei EZ Prep Nuclei Isolation Kit (Sigma, Australia). Cells were harvested from T-75 culture flasks with cultures at a high confluence. Approximately $8 \times 10^6$ cells were harvested by centrifugation at $500 \times g$ for five minutes at 4 °C and washed in 10 mL of ice-cold Dulbecco’s phosphate buffered saline (PBS), followed by further centrifugation at $500 \times g$ for 5 minutes at 4 °C. The supernatant was discarded and the remaining cell pellet was vortexed and completely re-suspended and incubated on ice with 4 mL of cold Nuclei EZ lysis buffer for 5 minutes. Following this incubation period, nuclei were collected by centrifugation at $500 \times g$ for 5 minutes at 4 °C and the supernatant containing cytoplasmic components was discarded. The process of incubation with lysis buffer was repeated a second time to further purify the nuclei extracted. The final pellet of isolated nuclei was resuspended in $200 \mu L$ of cold Nuclei EZ storage buffer and stored at – 80 °C prior to analysis.

Nuclei were analysed in the microfluidics sample device previously described in section 4.2.1 and according to S-FTIR methodology for single cell analysis described in section 2.3.5 with an adjustment of the aperture size to 5 x 5 µm to allow for the smaller size of nuclei in comparison to single cells.

4.3.5.2 Nuclei Extract Results

Nuclei isolated from HL60 cells were smaller than those isolated from K562 cells and were difficult to analyse as single organelles, particularly since they presented as clumped together in many areas. With the selected aperture size of 5 x 5 µm, spectra obtained for HL60 nuclei were poor quality and data are not presented. Nuclei extracted from K562 cells were amenable to analysis and mean spectra for single K562 nuclei extracted from control cells ($n = 50$ nuclei) and BaP treated cells ($n = 43$ nuclei) are shown in Figure 4.13.
Figure 4.13 Normalised mean S-FTIR spectra of nuclei extracted from K562 control (light blue) and BaP (dark blue) treated cells.

PLS-DA was performed for control and 24 hour BaP treated nuclei spectra. Figure 4.14 shows the scores of the first PLS-DA latent variable for K562 nuclei in which LV1 clearly distinguishes BaP treated (green) from control cell nuclei (red).

Figure 4.14 PLS-DA scores plot for the first PLS-DA component showing discrimination between control (red) and BaP treated (green) cell nuclei analysed by S-FTIR.
Data indicate that S-FTIR can detect BaP induced biological changes within single nuclei of AML cells, revealing that BaP acts at the nuclear level. This is interesting given that spectral DNA bands previously observed in the cells (data presented in section 4.3.4) show no significant change with drug treatment which therefore suggests that biological differences giving rise to the observed classification here are arising from other nuclear changes. Cell nuclei contain lipids which play a role in differentiation, proliferation and apoptosis [288]. PC lipids and sphingomyelin which are present in high quantities are known to form lipid-protein complexes, and the existence of nuclear microdomains (with similarity to intracellular lipid-protein complexes) was recently reported for the first time [289]. Furthermore, it was reported that nuclear microdomain lipid composition changes with proliferation.

The first loadings vector from the PLS-DA shown in Figure 4.15, clearly shows spectral features arising from nuclear lipids in the high wavenumber region, as well as further biological features in the fingerprint region, which include bands arising from proteins (Amide I and II) and lipids (CH deformations and lipid carbonyl stretching).

![Figure 4.15 PLS-DA loadings vector showing key bands in the spectra contributing to the discrimination between control and BaP treated cell nuclei.](image-url)
Further work needs to be carried out for data interpretation which would include a larger sample set (this was not possible due to limited synchrotron beam time), however results from this preliminary study are promising and further contribute to our understanding of cellular targets of BaP.

4.3.6 Raman Spectroscopy of Cell Extracts for Lipidomic Study

Non-polar cell extracts were prepared onto stainless steel plates and analysed in a random sequence as described in section 3.5.5. Due to the aforementioned coffee ring effect with drying [281], spectra were obtained in triplicate as a minimum from both the inside and outside of the dried extract droplet for each sample, thus generating a large data set of more than 144 Raman spectra in total. Spectra were processed as described in section 4.2.2.2.

Figure 4.16 shows the HCA analysis for each sample class for spectra obtained from inside and outside each dried extract spot. The HCA cut off was designated as the first big “jump” in between-cluster similarity.
Poor reproducibility between spectra was observed for replicate analysis from both regions of the dried cell extract. HCA selected 51.9% of spectra (41 out of 79 spectra in total) for the final analysis from the inside of each extract spot and 39.47% (30 out of 76 spectra) for the final analysis from the outer region of each coffee ring. Data therefore suggest that dried cell extract from the inside of each spot could have a more homogenous distribution overall than the extract at the outer edges of each spot. This is expected, given the coffee ring effect of a higher concentration distribution of biomolecules expected at the outer edges of each spot with drying.

**Figure 4.16** Hierarchical cluster analysis for cell extract spectra selection from inside (a) and outside (b) of each dried sample spot. The x-axis is an arbitrary scale (0-1) of distance based on dissimilarity (closer to 0 equating to more similar).
A further contributing factor to poor reproducibility of AML cell extracts from both regions of the dried extract spots can be explained by a consideration of the visual appearance and corresponding spectra for regions within the dried extracts where dark spots are clearly visible. An example of this can be seen in the optical image of HL60 control extract biorep 3, captured from inside the cell extract shown in Figure 4.17.

Figure 4.17  Optical image from inside dried cell extract spot showing regions from where replicate Raman spectra shown in Figure 4.18 were obtained. Region 2 highlighted with white circle indicates sample heterogeneity.

Figure 4.18 shows corresponding raw spectra overlaid from regions 1 (blue), 2 (red) and 3 (black). Spectra from regions 1 and 3 have a similar profile to each other but are clearly different from the spectra acquired from region 2 (where there is an obvious dark area within the dried cell extract, identified with white circle). The spectra from region 2 shows a prominent band at the wavenumber region of 1006 cm\(^{-1}\) which is absent in regions 1 and 3 and can be assigned to the symmetrical ring breathing mode of phenylalanine [145]. Phenylalanine is an aromatic amino acid which is a strong contributor to the spectral features of protein. This sharp peak is observed in all spectra taken from the visibly ‘dark’ areas which appear to be randomly distributed both inside the dried spots and outside at the edges and contribute to the heterogeneity of dried cell extracts throughout this extract study.
Despite Figure 4.17 and Figure 4.18 exemplifying the biochemical heterogeneity (and therefore spectral irreproducibility) within dried cell extracts, spectra retained after HCA were deemed appropriate for further analysis.

Figure 4.19 shows PCA scores plots for extracts analysed from inside (a) and outside (a) dried extract spots. For both plots, the first principal components represent the most significant source of spectral variability in each data set (13.66 % of the total variance for inside spectra and 24.07 % for the outside spectra) which describes the variance between the two cell types, HL60 and K562.
Figure 4.19  PCA scores plots taken from inside (a) and outside (b) regions of dried sample for HL60 and K562 control and BaP treated cell extracts. Principal components 1 and 2 are shown, with cell type clearly separated along PC 1.

Principal component 2 accounts for 7.21% of the intra-cell variance for spectra obtained from the inside of dried extract drop and in the case of HL60 cell extracts, suggests some discrimination between control and BaP treated cells (red). Principal component 2 does not obviously discriminate between drug treatment for K562 cell extracts (blue), however PC 3 shown in Figure 4.20 does appear to (for spectra taken from both inside and outside sampling regions).
Figure 4.20  PCA scores plots taken from inside (a) and outside (b) regions of dried sample for HL60 and K562 control and BaP treated cell extracts. Principal components 1 and 3 are shown, with cell type clearly separated along PC 1.

The small variance along PC3 which appears to discriminate between K562 control and BaP treated cell extracts for sampling regions both inside and outside dried extracts, can be further interrogated. The loadings for both PC1 (which discriminates cell type) and PC3 are shown in Figure 4.21.
The loadings for PC1 which describes variance between HL60 and K562 cell types, are not consistent for the different regions of dried extract sampled. There are clear differences between loadings obtained for inside (Figure 4.21a) and outside (Figure 4.21b) regions of the samples. This significantly limits biological data interpretation for this study as it pertains to the cell extract being too heterogeneous to yield significant data, even after HCA was used to filter the data matrix. Furthermore, loadings for PC3 which appears to describe the variance between K562 control and BaP treated cell extracts, are noisy and cannot be interpreted with confidence.

Data presented here suggest that Raman spectroscopic analysis of AML cell extracts is not fit for purpose for probing the action of drug within cells due to sample heterogeneity in cell extracts when prepared for analysis. Raman microspectroscopy of whole cells is therefore an attractive alternative for probing perturbations to intra-
cellular species and offers a distinct advantage in that data obtained contains spatial information for biological species. This methodology is explored in section 4.3.7.

4.3.7 Raman Microscopy to Map the Spatial Distribution of Lipids in Control and Drug Treated Cells

Raman mapping was employed to further probe the drug-induced biochemical changes observed with FTIR and to provide spatial information regarding these changes. K562 formalin fixed cells were mapped according to methodology described in section 2.4.4. The resulting data matrix for each cell map obtained was refined by removing background pixels by evaluating the signal-to-noise ratio of the Raman spectra. A cut-off threshold of 2 efficiently removed all the background pixels as shown in Figure 4.22. Single cell images of cell features were derived using FCM analysis on all spectra obtained from K562 formalin fixed control and drug treated cells (n=6 and n=5 respectively). In this process, each pixel is associated with a fuzzy membership value so that the gradient from one cell compartment to another can be well presented. Figure 4.22 shows a representative K562 control cell (Figure 4.22a) and drug treated cell (Figure 4.22b) with the cell membrane clearly defined (yellow) around the outer perimeter of the cell in Cluster 1 and internal cell structures clearly observable (yellow) in Cluster 2.
Figure 4.22 FCM images of a representative control and BaP treated K562 formalin fixed cell.

Table 4.2 shows the confusion matrices for the PLS-DA computed across the entire spectral region for all 11 K562 formalin fixed cells, with correct classification rates of 95.23% for BaP treated cells and 89.11% for control cells of combined clusters (all cell spectra). Despite there being no obvious changes by eye when comparing the images with each other, the VIP PLS-DA scores plots shown in Figure 4.23 clearly highlight three distinct Raman peaks as key to this classification which all arise from cellular lipid changes with BaP treatment.
Table 4.2 PLS-DA K-fold-cross-validation model on 11 K562 cells. Rows represent predicted classification and columns represent experimental.

<table>
<thead>
<tr>
<th></th>
<th>BaP</th>
<th>CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BaP</td>
<td>88.38%</td>
<td>11.62%</td>
</tr>
<tr>
<td>CON</td>
<td>23.66%</td>
<td>76.34%</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BaP</td>
<td>98.38%</td>
<td>1.62%</td>
</tr>
<tr>
<td>CON</td>
<td>7.46%</td>
<td>92.54%</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BaP</td>
<td>95.23%</td>
<td>4.77%</td>
</tr>
<tr>
<td>CON</td>
<td>10.89%</td>
<td>89.11%</td>
</tr>
</tbody>
</table>

Figure 4.23 VIP scores for PLS-DA of K562 control and drug treated cells showing three Raman lipid peaks.

The averaged spectra for clusters 1 and 2 (overlaid for drug and control in Figure 4.24) can be used to probe this further. In both clusters, BaP gives rise to an apparent increase in the CH₂ symmetric stretch from lipid acyl chains observed at 2845 cm⁻¹ which is consistent with changes observed by S-FTIR and ATR-FTIR. Furthermore, the enlarged spectra of the fingerprint region in Figure 4.24 show an apparent decrease in the peak at 1440 cm⁻¹ which arises from the CH₂ scissoring mode and is reportedly
proportional to the amount of saturated C-C lipid bonds [74, 159, 290]. This suggests that overall lipid saturation decreases with BaP treatment.

Mapping the spatial distribution of spectral features across single cells has identified that the aforementioned lipid changes appear to be consistent across the whole cell, with no obvious regional differentiation between membrane lipids and intra-cellular lipids.
Figure 4.24 Average spectra from cluster 1 and cluster 2 regions, overlaid for control and BaP treated K562 cells. Black arrows in the enlarged spectra show the lipid regions that contribute to classification of control and drug treated cells.
4.3.8 Biological Interpretation of the Effect of BaP on Cellular Lipids

All three spectroscopic techniques employed in this study revealed an increase in the CH$_2$:CH$_3$ ratio of lipid peaks in the high wavenumber spectral region of drug treated cells which indicates a change in cellular lipid saturation state with BaP treatment and demonstrates good consistency between the analytical techniques. From the FTIR data alone it is not possible to infer specifically what lipid saturation change this refers to. However, in the case of the Raman spectra the peak observed at 1440 cm$^{-1}$ decreases (reportedly proportional to total lipid saturation) indicating an increase in lipid unsaturation with BaP treatment.

One novel aspect of this data is the visualisation of increased lipid unsaturation at the single cell level. Data suggest that the BaP-induced change in saturation state is uniform across the cell. Increasing lipid unsaturation typically renders these biomolecules more fluid which has implications for cancer cell biology and survival. The phenomenon of high rates of *de novo* lipogenesis observed in cancer cells (discussed in section 1.2.6) enables them to synthesise new cell membranes during rapid proliferation and it also allows cancer cells to be more independent in terms of their energy supply and biochemical need for growth. Enhanced lipogenic phenotype is characterised by lipids with saturated or mono-unsaturated acyl chains in a number of different cancer cell lines which is significant in terms of protecting them from lipid peroxidation and apoptosis [30]. During lipid peroxidation, free radicals remove electrons from cell membrane lipids, producing a number of oxidised lipid species and small molecule reactive oxygen species (ROS), which could jeopardise the stability and longevity of a tumour cell [30]. Membrane rigidity (arising from increased lipid saturation) is also implicated in cell resistance to anti-cancer drugs by impairing uptake of drugs through the membrane [291] and any increasing lipid unsaturation observed within cells could be a mechanism to preventing drug resistance.

Considering the important role that saturated lipids play in cancer cell function and survival, spectroscopic data reported here suggests that the BaP-induced increase in lipid unsaturation observed is likely to play a role in its’ anti-cancer activity.
4.3.9 Conclusion

Employing a multi-platform spectroscopic approach to probe the action of BaP on HL60 and K562 cell lines affords a comprehensive picture of the effect of drug therapy on the biochemical nature of the cells and allows us to further understand the targets of BaP. Data reported for S-FTIR, ATR and Raman microspectroscopy strongly indicate that lipid biochemistry is a significant target of BaP, furthermore the ability of S-FTIR to probe cells in a hydrated state enabled ‘drug-DNA’ interaction to be probed and indicated that DNA is unlikely to be a target of this therapy.

A significant increase in methylene functionality with drug treatment indicates that AML cells with increased unsaturation post-BaP treatment are representative of the more typical biochemistry of non-cancerous cells. Raman microspectroscopy supported the findings of FTIR and added a further dimension to the study by providing spatial information of lipid distribution which suggested that BaP-induced saturation change is uniform across a single cell. Individual cell nuclei were also observed to respond to BaP in a preliminary study on a small cohort of K562 nuclei which was suggestive of intra-nuclear lipids also responding to drug treatment in a similar manner, a hypothesis which needs to be further tested.

A mass spectrometric study described in Chapter 6 will further probe the response of cellular lipid species to BaP treatment and will investigate saturation changes in targeted lipid species to provide orthogonal information to the understanding of BaP mode of action.
Chapter 5 A Metabolomics Investigation into the Anti-Leukaemic Action of BaP In Vitro by UHPLC-MS

5.1 Introduction
Metabolomic profiling can yield powerful biochemical data at the cellular level, giving a snapshot view of the phenotype of a cell population under any given perturbation. There have been vast numbers of reports in the literature of UHPLC-MS being widely used to probe metabolic changes arising from drug-cell interactions as well as for the discovery of new biomarkers and for disease diagnosis as discussed in section 1.5.1.3.

This chapter presents the findings of an untargeted MS-based metabolomics study to investigate the drug-cell interactions between BaP and AML cell lines. In order to contribute to the current understanding of the mode of action of BaP, compositional changes in intracellular components were investigated in control and BaP treated cells and resulting metabolic patterns were projected onto biochemical pathways to elucidate potential targets of BaP within the cell.

5.2 Materials and Methods

5.2.1 Cell Culture and Drug Treatment
Acute myeloid leukaemia cell lines (HL60 and K562) were cultured as described in section 3.2.2. Drug treatment with BaP was administered as described in section 3.3.3.

Replicates (reps) were prepared for this metabolomics study. Six reps of each cell line were cultured in T-75 flasks along with six control flasks. Each flask was seeded from the same pooled cell suspension as illustrated in Figure 3.15.
5.2.2 Sample Preparation

Cells were extracted at 24 hours for UHPLC-MS metabolic profiling according to the protocol described in Figure 3.14.

Data presented in this chapter are from the analysis of polar (methanol) extracts obtained from this extraction procedure.

5.2.3 UHPLC-MS Methodology

Cell extracts were analysed using an Acquity UPLC system (Waters, UK) coupled to an electrospray (ESI) LTQ–Orbitrap XL hybrid mass spectrometer (Thermo Scientific Ltd. Hemel Hempsted, UK) according to the conditions detailed in section 2.2.5.

5.2.4 Data Quality

A total of 128 injections were performed for this large-scale metabolomics study, totalling 80 h of instrumental analysis. These included six replicates for each of the two cell lines for control and BaP treatment for both polar and non-polar extracts (24 methanol extracts (data presented here) and 24 chloroform extracts (data presented in Chapter 6)). Technical replications of pooled QC cell extract were performed at the beginning of the run (5 x QC injections) and after every five sample injections, totalling 22 QC injections and all analysis was carried out in both positive and negative ESI (i.e. not with polarity switching in a single run). Sample injection order was randomised to eliminate any bias. Data analysed for this study represents half of the total, being from polar extracts and consisting of 11 QC samples.

5.2.5 UHPLC-MS Data Processing

Raw data files were converted in to the universal NetCDF format via the software conversion tool within Xcaliber software supplied by the manufacturer (Thermo Scientific Ltd. Hemel Hempsted, U.K.). Subsequently, in house peak deconvolution software containing the XCMS algorithm was used for peak picking as described previously [128, 292]. The output from this system resulted in a MS Excel based data
matrix of mass spectral features with related accurate $m/z$ and retention time pairs. Data from the internally pooled QC samples were then used to align for instrument drift and quality control. The data matrix was also signal corrected to remove peaks with greater than 20% RSD threshold and with no more than two missing values within QC samples across the analytical run. Spectra were then normalised using probabilistic quotient normalisation (PQN) [293] for peaks that occurred in all samples. Normalisation of metabolomics data is often required because metabolite concentrations can often span several orders of magnitude (typically <1 μM to mM concentrations) which can cause small but significant metabolite changes to be obscured. Furthermore, normalisation can reduce technical variation or systematic bias [294].

5.2.6 UHPLC-MS Metabolite Identification

Data were aligned and processed through the Taverna workflow as described by Brown et al [295] for putative peak identification and MetaboAnalyst 3.0 [296, 297] for further statistical and pathway analysis.

Ambiguity arising from the same $m/z$ ratio can lie within metabolite and lipid identification due to multiple identifications (based on accurate mass) and differing points of unsaturation for lipids. Multiple adducts of the same species can also occur due to the presence of different charged (composite) species (i.e. protonated and sodiated ions).

5.3 Results and Discussion

Metabolomic profiling of six replicates of HL60 and K562 control and BaP treated cell extracts yielded a large data matrix containing 5390 metabolite features in positive mode and 3643 features in negative mode. Processing these data through the Taverna workflow gave an output of 2116 annotated features in positive and 1744 features in negative mode which were subjected to further statistical analysis.
5.3.1 Multivariate Analysis of UHPLC-MS Profiling Data from Drug Treated Cells

The features annotated by Taverna were imported into MetaboAnalyst for multivariate and univariate analysis to initially assess biological differences between the two cell lines and to investigate apparent trend with drug treatment. Data here were subjected to $\log_{10}$ transformation and Pareto scaling prior to statistical analysis to give all annotated features a normal distribution prior to statistical analysis. The output from these pre-processing algorithms can be seen in Figure 5.1 for metabolite features observed in positive mode.

![Figure 5.1](image.png)

**Figure 5.1** Representative example of UHPLC-MS data pre-processing showing selected features before and after $\log_{10}$ transformation and Pareto scaling performed in MetaboAnalyst 3.0 prior to further statistical analysis.

The resultant PCA scores plots from 2-way comparisons between classes are shown in Figure 5.2 for both positive and negative ESI modes.
Figure 5.2  PCA scores plots obtained for HL60 and K562 control and BaP treated cells showing all replicates included in the polar UHPLC-MS metabolomics study. Coloured ellipses represent 95% confidence limits calculated from scores.
Principal component 1 (PC1) clearly shows discrimination between the two cell lines, with 32.1% of the total explained variance (TEV) assigned to inherent biological differences between HL60 and K562 cells for features observed in positive mode and 37.4% for features found in negative mode.

The trend between control and BaP treatment is the same for both cell lines, with drug treated cells observed in the positive (top half) of the PCA space, and accounting for 18.3% of the discrimination along PC 2 for data obtained in positive and 23.7% for negative modes. The plot suggests that the BaP treatment applied has a similar overall impact for both cell lines given the mirrored separation in the PC space.

The exceptionally well-defined grouping observed along PC2 in Figure 5.2 could simply be due to the presence of drug within the cell extracts analysed rather than drug effect on intracellular metabolites. The data matrix was therefore further refined by removing any obvious $m/z$ values that could directly pertain to either drug (bezafibrate and medroxyprogesterone acetate) administered. Rationale for refinement was based upon the molecular weight of these drugs and a consideration of possible dimer and adduct formation as well as spurious proposed identifications for suspected drug peaks. This reduced the total number of features to 2085 for positive and 1697 for negative ion modes, representing the removal of 31 potential drug related features observed in positive and 47 potential features observed in negative. PCA was performed on this refined data set and PCA scores plots are shown in Figure 5.3.
Figure 5.3  PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of features directly associated with drugs administered. Coloured ellipses represent 95% confidence limits calculated from scores.
Again, the two cell lines clearly separate along PC1 with a similar contribution to discrimination as previously observed in Figure 5.2 (here, 35.0 % TEV for positive features and 41.7 % for negative features). BaP treated cell extracts are also present in the same PC space as previously observed, however, now there is a visual reduction in the distance between group distinctions for control and drug treated cell extracts. A projection of the groups within the PCA space can be further observed in the 3-dimensional scores plots shown in Figure 5.4. Two particular replicates, namely K562 BaP 3 and HL60 CON 1 do not group well with their respective classes and are therefore suggestive of outliers in the PCA scores plots for both positive and negative, with the latter causing a slight overlap of grouping between control and drug treated HL60 cell extracts in positive mode. The design of the study to include 6 replicates for each cell extract class compensates for there being one potential outlier in 2 of the 4 classes observed and these are therefore not considered to be a problem.
In order to assess instrument variation over the course of the experiment, QCs were included in the data matrix for the scores plots and are shown encircled in pink in Figure 5.5, which represents a data matrix with the removal of drug peaks (data presented in Figure 5.3) and the inclusion of QCs, shown for both positive and negative ionisation modes.
Figure 5.5 PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of features directly associated with drugs administered and including pooled QCs in the analysis. Coloured ellipses represent 95% confidence limits calculated from scores.
QCs are observed to be tightly grouped towards the middle of the PC space as one would expect (since they consist of a representative aliquot from all samples analysed) which suggests good instrument precision for the entire run duration.

PCA scores plots clearly suggest potential to further explore the relationships between control cells and BaP treated cells through interpretation of the PCA loadings. Interrogating the loadings and correlating these with output from univariate statistical analysis provides insights into perturbations of cellular biochemistry with drug treatment.

5.3.2 PCA Loadings Analysis for UHPLC-MS Feature Selection

Loadings plots were constructed to determine variables (metabolite features) that were strongly contributing to the class separation between control and drug treated cells observed by PCA. In the representation of loadings shown in Figure 5.6, variables with loadings in a given position in the loadings plot contribute heavily to observations whose scores are found in a similar position in the complementary scores plot from Figure 5.3. Despite visual interpretation of these loadings being difficult, they provide a reliable summary of the variables influence on the model assuming normalisation for scores and loadings are on the same scale [298] and were used here to select data for further interpretation.
Figure 5.6 Loadings plots obtained for HL60 and K562 polar cell extracts analysed by UHPLC-MS showing m/z features contributing to class separation between control and drug treated extracts along PC2 shown in Figure 5.3.

To interrogate these data further, the top 80 m/z loadings were taken from each of the positive and negative regions along loadings 2 (along which class separation in PCA scores plots was suggestive of drug treatment) and tabulated. These 160 potentially key features were then cross checked with significant variables selected by univariate one-way analysis of variance (ANOVA), with $p < 0.05$ set as the level of statistical significance and post-hoc analysis according to the Fishers LSD test. Selecting a low $p$-value reduces the likelihood that the association between a metabolite feature and its related function is by chance. Values of $p < 0.05$ are generally used to indicate non-random association and statistical significance [299], although this does not mean that features with a $p$-value greater than this do not have a genuine related function to the data set and conversely neither does this cut-off guarantee significance [300].

ANOVA reduced the data set from 2085 features to 778 significant features ($p < 0.05$) in positive mode and from 1697 features to 751 significant features in negative mode. The top 20 features from each ionisation mode with the lowest $p$-values that were also present in the key features selected from multivariate analysis were tabulated and are shown in Table 5.1.
<table>
<thead>
<tr>
<th>m/z</th>
<th>Ionisation mode, positive (+) or negative (-)</th>
<th>p-value</th>
<th>-log10(p)</th>
<th>Present in PC 2 loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>432.0874</td>
<td>+</td>
<td>1.63E-50</td>
<td>49.788</td>
<td>y</td>
</tr>
<tr>
<td>298.0579</td>
<td>+</td>
<td>5.85E-46</td>
<td>45.233</td>
<td>y</td>
</tr>
<tr>
<td>762.6698</td>
<td>+</td>
<td>6.06E-45</td>
<td>44.217</td>
<td>y</td>
</tr>
<tr>
<td>514.0423</td>
<td>+</td>
<td>7.39E-45</td>
<td>44.132</td>
<td>y</td>
</tr>
<tr>
<td>416.0342</td>
<td>+</td>
<td>2.34E-44</td>
<td>43.63</td>
<td>y</td>
</tr>
<tr>
<td>438.1057</td>
<td>+</td>
<td>3.13E-44</td>
<td>43.504</td>
<td>y</td>
</tr>
<tr>
<td>452.0713</td>
<td>+</td>
<td>3.94E-43</td>
<td>42.404</td>
<td>y</td>
</tr>
<tr>
<td>845.1220</td>
<td>+</td>
<td>5.83E-46</td>
<td>41.234</td>
<td>y</td>
</tr>
<tr>
<td>761.6707</td>
<td>+</td>
<td>8.48E-42</td>
<td>41.072</td>
<td>y</td>
</tr>
<tr>
<td>583.1144</td>
<td>+</td>
<td>1.84E-41</td>
<td>40.735</td>
<td>y</td>
</tr>
<tr>
<td>403.0631</td>
<td>+</td>
<td>2.49E-41</td>
<td>40.603</td>
<td>y</td>
</tr>
<tr>
<td>420.695</td>
<td>+</td>
<td>3.40E-41</td>
<td>40.469</td>
<td>y</td>
</tr>
<tr>
<td>224.1281</td>
<td>+</td>
<td>5.55E-41</td>
<td>40.256</td>
<td>y</td>
</tr>
<tr>
<td>434.0820</td>
<td>+</td>
<td>7.15E-41</td>
<td>40.146</td>
<td>y</td>
</tr>
<tr>
<td>767.1866</td>
<td>+</td>
<td>4.35E-40</td>
<td>39.361</td>
<td>y</td>
</tr>
<tr>
<td>383.1494</td>
<td>+</td>
<td>8.70E-40</td>
<td>39.061</td>
<td>n</td>
</tr>
<tr>
<td>465.0405</td>
<td>+</td>
<td>9.10E-40</td>
<td>39.044</td>
<td>y</td>
</tr>
<tr>
<td>515.0447</td>
<td>+</td>
<td>1.67E-39</td>
<td>38.778</td>
<td>y</td>
</tr>
<tr>
<td>869.4436</td>
<td>+</td>
<td>1.69E-39</td>
<td>38.771</td>
<td>n</td>
</tr>
<tr>
<td>773.6540</td>
<td>+</td>
<td>3.03E-39</td>
<td>38.518</td>
<td>y</td>
</tr>
<tr>
<td>474.0806</td>
<td>-</td>
<td>1.27E-36</td>
<td>35.897</td>
<td>y</td>
</tr>
<tr>
<td>559.0487</td>
<td>-</td>
<td>1.66E-36</td>
<td>35.78</td>
<td>y</td>
</tr>
<tr>
<td>632.0497</td>
<td>-</td>
<td>2.39E-35</td>
<td>34.621</td>
<td>y</td>
</tr>
<tr>
<td>610.0551</td>
<td>-</td>
<td>1.30E-34</td>
<td>33.885</td>
<td>y</td>
</tr>
<tr>
<td>542.0678</td>
<td>-</td>
<td>1.33E-33</td>
<td>32.876</td>
<td>y</td>
</tr>
<tr>
<td>458.0847</td>
<td>-</td>
<td>1.33E-33</td>
<td>32.875</td>
<td>y</td>
</tr>
<tr>
<td>941.1351</td>
<td>-</td>
<td>8.06E-33</td>
<td>32.094</td>
<td>y</td>
</tr>
<tr>
<td>499.0749</td>
<td>-</td>
<td>9.53E-33</td>
<td>32.021</td>
<td>y</td>
</tr>
<tr>
<td>544.0641</td>
<td>-</td>
<td>9.03E-32</td>
<td>31.044</td>
<td>y</td>
</tr>
<tr>
<td>875.1453</td>
<td>-</td>
<td>1.25E-31</td>
<td>30.904</td>
<td>y</td>
</tr>
<tr>
<td>457.0911</td>
<td>-</td>
<td>3.46E-31</td>
<td>30.46</td>
<td>y</td>
</tr>
<tr>
<td>470.1035</td>
<td>-</td>
<td>7.83E-31</td>
<td>30.106</td>
<td>y</td>
</tr>
<tr>
<td>493.0581</td>
<td>-</td>
<td>1.16E-30</td>
<td>29.934</td>
<td>y</td>
</tr>
<tr>
<td>880.1678</td>
<td>-</td>
<td>5.87E-30</td>
<td>29.231</td>
<td>y</td>
</tr>
<tr>
<td>444.0614</td>
<td>-</td>
<td>8.06E-30</td>
<td>29.093</td>
<td>y</td>
</tr>
<tr>
<td>512.0383</td>
<td>-</td>
<td>1.62E-29</td>
<td>28.79</td>
<td>y</td>
</tr>
<tr>
<td>634.0461</td>
<td>-</td>
<td>2.89E-29</td>
<td>28.539</td>
<td>y</td>
</tr>
<tr>
<td>678.0427</td>
<td>-</td>
<td>3.04E-29</td>
<td>28.517</td>
<td>y</td>
</tr>
<tr>
<td>561.0453</td>
<td>-</td>
<td>3.56E-29</td>
<td>28.448</td>
<td>y</td>
</tr>
<tr>
<td>805.1603</td>
<td>-</td>
<td>4.84E-29</td>
<td>28.315</td>
<td>y</td>
</tr>
</tbody>
</table>

*Table 5.1* Correlation between ANOVA and PC 2 loadings for significant features found in cell extracts.
All 20 of the top ANOVA features (lowest \(p\)-values) were present in the top 80 features selected from loadings 2 for negative mode and 18 out of 20 were present for positive mode. It is interesting to note that all features in Table 5.1 are present in the positive loadings (Figure 5.6), which indicates that the most significant features found are upregulated with drug treatment (higher concentrations in drug than control). This is explored further in Table 5.2.

Whilst ANOVA is a widely reported univariate method for finding significantly altered metabolites this can still result in a high chance of false discovery, even with apparently low thresholds of \(p\)-values (typically considered significant at \(p < 0.05\)). A false positive presents when the \(p\)-value is lower than the cut off suggesting a significant change in metabolite, when in reality no difference actually exists. In this case there is a 5 % chance of the wrong answer, i.e. 5 % of the reportedly significant data will in fact be false positives. With the large data matrices in this experiment, there could therefore be hundreds of false positives from applying ANOVA to all individual features found.

The false discovery rate (FDR) approach can estimate the chance of false discovery at a selected threshold. A \(q\)-value is assigned for each feature which is the lowest FDR at which the feature is called significant and reduces the number of false positives being obtained for large data sets. The significance analysis of microarrays (SAM) method was employed here to further assign confidence to significant features (variables) found. SAM is an established statistical method for determination of differentially expressed features (genes or metabolites) within high-dimensional data and directly estimates \(q\) values using a re-sampling approach [301, 302]. Outputs from SAM are a \(t\)-statistic value (SAM score, denoted \(d\)) and a \(q\)-value for each feature which vary inversely proportional with each other (as \(d >0\) increases, the \(q\)-value decreases). The \(d\)-value (SAM score) is a statistic based on the ratio of change in the variable to the standard deviation of repeated measurements for that feature. It is termed the ‘relative difference, \(d(i)\) in feature expression [303]. Variables with scores greater than an adjusted threshold (delta) are compared to their relative difference to the distribution estimated by random permutations of the class labels. For each threshold, a portion of the variables in the permutation set will be found to be significant by chance and this number is used to calculate the FDR [297]. The \(q\)-value is therefore computed by finding the smallest delta value for which a feature is called
significant and then assigning a FDR value to it corresponding to delta [304]. Figure 5.7 shows the SAM plots obtained for each ionisation mode. The choice of delta value is user dependent, based upon how many false positives the user is comfortable with. For the positive mode data matrix, a delta of 1 was selected which resulted in an output of 350 significant features and for negative mode a delta value of 1.5 was selected which yielded 305 significant features.
Figure 5.7 Significant features identified by SAM (Significance Analysis of Microarray) in positive and negative ionisation. For features in green, the further the deviation from linearity, the greater their significance.
For significant features identified by SAM (indicated by green circles in Figure 5.7), the further that a feature deviates from the linear ‘$d$’ line of observed vs expected, the more significant that feature is [305], therefore those with higher observed $d$-values contribute more significantly.

Output values from SAM were tabulated against the top 10 variables selected for each ionisation mode according to the aforementioned multivariate (PCA) and univariate (ANOVA) statistical analysis.

Table 5.2 shows these top 10 features with the corresponding SAM position by significance with $d$- and $q$-values reported from SAM. All previously determined significant features correlated with SAM significance, with all 10 negative features coming within the top 50 of 305 significant features identified by SAM, and all 10 positive features falling within the top 125 of 350 significant features identified by SAM. False discovery rates are reported as 0.004 for positive mode features and < 0.0005 for negative mode features.
<table>
<thead>
<tr>
<th>m/z (positive mode)</th>
<th>Position within significant 350 features identified by SAM</th>
<th>SAM d-value</th>
<th>SAM q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>298.0579</td>
<td>41</td>
<td>4.7001</td>
<td>0.0</td>
</tr>
<tr>
<td>416.0342</td>
<td>11</td>
<td>5.7580</td>
<td>0.0</td>
</tr>
<tr>
<td>432.0874</td>
<td>13</td>
<td>5.6541</td>
<td>0.0</td>
</tr>
<tr>
<td>438.1057</td>
<td>73</td>
<td>3.9205</td>
<td>0.0</td>
</tr>
<tr>
<td>452.0713</td>
<td>76</td>
<td>3.8914</td>
<td>0.0</td>
</tr>
<tr>
<td>514.0423</td>
<td>42</td>
<td>4.6480</td>
<td>0.0</td>
</tr>
<tr>
<td>583.1144</td>
<td>125</td>
<td>3.3828</td>
<td>3.0189 E-5</td>
</tr>
<tr>
<td>761.6707</td>
<td>52</td>
<td>4.3379</td>
<td>0.0</td>
</tr>
<tr>
<td>762.6698</td>
<td>37</td>
<td>4.7542</td>
<td>0.0</td>
</tr>
<tr>
<td>845.1220</td>
<td>91</td>
<td>3.7462</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m/z (negative mode)</th>
<th>Position within significant 305 features identified by SAM</th>
<th>SAM d-value</th>
<th>SAM q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>458.0847</td>
<td>3</td>
<td>8.2821</td>
<td>0.0</td>
</tr>
<tr>
<td>474.0806</td>
<td>7</td>
<td>7.2502</td>
<td>0.0</td>
</tr>
<tr>
<td>499.0749</td>
<td>34</td>
<td>5.3533</td>
<td>0.0</td>
</tr>
<tr>
<td>542.0678</td>
<td>16</td>
<td>6.2198</td>
<td>0.0</td>
</tr>
<tr>
<td>544.0641</td>
<td>50</td>
<td>4.9746</td>
<td>0.0</td>
</tr>
<tr>
<td>559.0487</td>
<td>20</td>
<td>5.9034</td>
<td>0.0</td>
</tr>
<tr>
<td>610.0551</td>
<td>27</td>
<td>5.5662</td>
<td>0.0</td>
</tr>
<tr>
<td>632.0497</td>
<td>12</td>
<td>6.3056</td>
<td>0.0</td>
</tr>
<tr>
<td>875.1453</td>
<td>24</td>
<td>5.7972</td>
<td>0.0</td>
</tr>
<tr>
<td>941.1351</td>
<td>28</td>
<td>5.5171</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 5.2 Numerical correlation between top 10 significant features according to ANOVA and PCA with SAM significance for cell extracts. The higher the SAM d-value, the more likely it is to be significant.

A strong correlation between significant features found by univariate and multivariate data analysis along with an imposed cut off for estimation of false discovery rates using a re-sampling approach (SAM) has allowed determination of the top 10 significant features for each ionisation mode. For each of these features, box plots were generated to explore the relative concentrations changes of each m/z between control and drug treated cell extracts. Table 5.3 shows the top 10 significant features selected by statistical analysis with box plots to illustrate relative feature concentrations for each feature.
<table>
<thead>
<tr>
<th>m/z</th>
<th>Top 10 significant m/z features from correlation of PCA loadings 2 and ANOVA (&amp; confirmed by SAM) in positive mode</th>
</tr>
</thead>
</table>
| 298.0579 | ![Graph](image1)
| 416.0342 | ![Graph](image2)
| 432.0874 | ![Graph](image3)
| 438.1057 | ![Graph](image4)
| 452.0713 | ![Graph](image5)
| 458.0847 | ![Graph](image6)
| 474.0806 | ![Graph](image7)
| 499.0749 | ![Graph](image8)
| 514.0423 | ![Graph](image9)
| 542.0678 | ![Graph](image10)
| 544.0641 | ![Graph](image11)
| 559.0487 | ![Graph](image12)
Table 5.3 Box plots for the 10 most significant features identified by univariate and multivariate analysis contributing to differences between control and BaP treated cell extracts. Red and navy blue box plots indicate relative amounts of each particular compound in HL60 and K562 BaP treated cell extracts respectively, with green and light blue box plots representing relative amounts of compound in HL60 and K562 control cell extracts respectively.

All features were found to be upregulated in drug treated cells for both cell lines analysed as indicated by the higher relative concentrations of each feature indicated by the red (for HL60 BaP) and navy blue (for K562 BaP) bars. This explains the location of these features in the positive (drug) region of the PC 2 space in Figure 5.3 and Figure 5.4.
5.3.3 Putative Identification of Significant Features

One of the key challenges in metabolomics studies is the identification of metabolites and typically mass-based database searching is employed which searches against molecular ions and adduct species for thousands of metabolites. Here, the 20 m/z features identified as being significant were searched against 3 key databases: KEGG, HMDB and LIPIDMAPS which are integrated into the LC-MS workflow previously described [295]. Putative identifications identified for the features are reported in Table 5.4 with a suggestion of the biological role of the tentatively assigned metabolite within cellular function.

<table>
<thead>
<tr>
<th>m/z (positive mode)</th>
<th>Putative ID from integrated KEGG/HMDB/LipidMaps search</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>298.0579</td>
<td>Hydroxyketamine</td>
<td>Unlikely, usually present after ketamine exposure</td>
</tr>
<tr>
<td>416.0342</td>
<td>N-Succinyl-amino-ketopimelate</td>
<td>Intermediate in lysine biosynthesis</td>
</tr>
<tr>
<td>432.0874</td>
<td>PC[4:0]</td>
<td>A diacylglycerophosphocholine class of lipid of the category glycerophospholipid. Systematic name: 1,2-dibutyryl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>438.1057</td>
<td>Arginyl-Phenylalanine; Phenyllalanyl-Arginine</td>
<td>Dipeptides composed of arginine and phenylalanine. Incomplete breakdown product of protein digestion or protein catabolism. Some dipeptides are known to have physiological or cell-signalling effects although most are simply short-lived intermediates</td>
</tr>
<tr>
<td>452.0713</td>
<td>Kinetin-N-glucoside</td>
<td>The product of the reaction between kinetin and UDP-D-glucose</td>
</tr>
<tr>
<td>514.0423</td>
<td>alpha-Hydroxyethyl-thiamine diphosphate</td>
<td>Hydroxyethyl-ThPP is involved in Glycolysis, Gluconeogenesis, Alanine and aspartate metabolism, Valine, Leucine and isoleucine biosynthesis, Pyruvate metabolism, and Butanoate metabolism</td>
</tr>
<tr>
<td>583.1144</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>761.6707</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>762.6698</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>845.1220</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>m/z (negative mode)</td>
<td>Putative ID from KEGG/HMDB/LIPIDMAPS search</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>458.0847</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>474.0806</td>
<td>Sulfamethoxazole N-glucuronide</td>
<td>Metabolite of Sulfamethoxazole (antibiotic drug). Sulfamethoxazole N-glucuronide belongs to the family of phenols and derivatives. Probable drug metabolite</td>
</tr>
<tr>
<td>499.0749</td>
<td>Chitobiose</td>
<td>A dimer of β-1,4-linked glucosamine units. Found in the Amino sugar and nucleotide sugar metabolism pathway</td>
</tr>
<tr>
<td>542.0678</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>544.0641</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>559.0487</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>610.0551</td>
<td>ADP-mannose; GDP-L-fucose</td>
<td>ADP-mannose is one of the major substrates of human hydrolases. Belongs to the class of organic compounds known as purine nucleotide sugars; GDP-L-fucose is a sugar nucleotide and a readily available source of fucose. The monosaccharide plays several important metabolic roles in complex carbohydrates and in glycoproteins</td>
</tr>
<tr>
<td>632.0497</td>
<td>CDP-glucose</td>
<td>CDP-glucose is a substrate for Uridine diphosphate glucose pyrophosphatase. It is found in amino sugar and nucleotide sugar metabolism.</td>
</tr>
<tr>
<td>875.1453</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
<tr>
<td>941.1351</td>
<td>unknown</td>
<td>No sensible hit</td>
</tr>
</tbody>
</table>

Table 5.4 Putative identifications of the top 10 significant features observed in positive and negative ESI LC-MS for drug treated cell extracts.

A variety of biomolecules appear to be potentially indicated as being affected by BaP treatment, which suggests that the therapy could be targeting amino sugar, nucleotide sugar and carbohydrate metabolism and lysine biosynthesis which are all indicated by more than one metabolite tentatively assigned.

Data analysis to this point has primarily focussed on statistical significance assigned to features which were suggestive as being responsible for the observed separation of groups between control and drug treated cell extracts according to PCA from a consideration of thousands of metabolite features. Changes in each feature observed
in the top 10 significant features for each ionisation mode were consistent between both HL60 and K562 cell lines which is encouraging as this adds more evidence to the biological interpretation of drug effect, however this approach does not take into account the mass error in putative identifications. Narrowing the parts per million (ppm) mass error window within which putative id’s are reported and exploring pathway analysis for a large number of tentatively assigned metabolite with low mass error provides the ability to broader spectrum of data usefully. This approach is further explored in section 5.3.4.

5.3.4 Pathway Analysis of Significant Metabolites Found in BaP Treatment Cell Extract Study for HL60 and K562 Cells

The original data matrix (post-removal of m/z values arising directly from drug) of 2085 Taverna annotated features (assigned from the aforementioned mass database searching) in positive mode and 1697 annotated features in negative mode (which were employed in univariate and multivariate analysis previously reported), were subjected to further refinement according to:

- fold changes calculated manually
- \( p \)-value (t-test)
- mass accuracy (ppm) for putative identification

Filters applied were \( p < 0.1 \) for features observed in both cell lines, mass accuracy < 2.5 ppm and fold changes consistent between both cell lines (either increase or decrease in both HL60 and K562 cell extracts). The mass accuracy window of 2.5 ppm reflects that expected for Orbitrap data obtained for small molecule analysis according to the instrument calibration performed. The higher cut-off for \( p \)-values of < 0.1 (previously \( p < 0.05 \) for data reported earlier) was considered appropriate for this refined data set given the extra filters (2.5 ppm mass accuracy) which provide further measures of control and confidence in assignments which were not present in earlier data matrices for analysis. This introduces a degree of scientific judgement of the data, acknowledging the limitations of restricting data analysis and interpretation to \( p \)-values of less than 0.05 alone [306]. These measures resulted in 44 discrete m/z features with putative identifications in positive mode and 62 m/z features with
putative identifications in negative mode. For each feature there were multiple putative id’s, typically 8 or 9 metabolite suggestions but in some instances up to 22 id’s for each falling within the 2.5 ppm mass error window.

Putative identifications of features were combined into one list for both positive and negative modes and processed through the pathway analysis function within MetaboAnalyst. In total, 131 feature names common to both cell lines were recognised for pathway analysis.

Results can be seen in Figure 5.8 and Table 5.5 which shows the output from pathway analysis for features observed in both cell lines according to the aforementioned filters.
Figure 5.8 Pathway analysis for common features found in both HL60 and K562 cell extracts identified to within 2.5 ppm. Pyrimidine metabolism is identified as a key pathway targeted by BaP treatment with the highest significance (p-value) and pathway impact. Further pathways are also identified.
The top 10 pathways identified in Figure 5.8 are tabulated according to significance (descending -log(p)) in Table 5.5.

<table>
<thead>
<tr>
<th>Pathway Results for HL60 &amp; K562</th>
<th>Total</th>
<th>Hits</th>
<th>-log(p)</th>
<th>Holm adjust</th>
<th>FDR</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrimidine metabolism</td>
<td>60</td>
<td>9</td>
<td>9.3565</td>
<td>0.007</td>
<td>0.01</td>
<td>0.23</td>
</tr>
<tr>
<td>D-Glutamine and D-glutamate metabolism</td>
<td>11</td>
<td>3</td>
<td>5.4452</td>
<td>0.341</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Phenylalanine metabolism</td>
<td>45</td>
<td>5</td>
<td>4.3272</td>
<td>1.000</td>
<td>0.35</td>
<td>0.14</td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td>41</td>
<td>4</td>
<td>3.2178</td>
<td>1.000</td>
<td>0.80</td>
<td>0.09</td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td>48</td>
<td>4</td>
<td>2.7311</td>
<td>1.000</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>32</td>
<td>3</td>
<td>2.5205</td>
<td>1.000</td>
<td>1.00</td>
<td>0.18</td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td>56</td>
<td>4</td>
<td>2.2844</td>
<td>1.000</td>
<td>1.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Inositol phosphate metabolism</td>
<td>39</td>
<td>3</td>
<td>2.0653</td>
<td>1.000</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>88</td>
<td>5</td>
<td>1.9103</td>
<td>1.000</td>
<td>1.00</td>
<td>0.19</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine biosynthesis</td>
<td>27</td>
<td>2</td>
<td>1.5461</td>
<td>1.000</td>
<td>1.00</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 5.5 Top 10 pathways identified by pathway topology analysis of metabolite features affected by BaP treatment in HL60 and K562 cell extracts. Pyrimidine metabolism shows a significant change with highest pathway impact. See text for details.

The column headed ‘total’ represents the total number of features in that pathway and the ‘hits’ column shows how many metabolites were matched from the list of 109 features uploaded. The -log(p) value is calculated from the original p-value determined from the enrichment analysis and the Holm p is the p-value adjusted by the Holm-Bonferroni method. In statistics of multiple testing, the family-wise error rate is the probability of at least one false positive over the collection of tests and the Bonferroni method corrects for multiple comparisons by assuming independence of the different tests. The Holm adjust more strongly controls the family-wise error rate and therefore has a larger probability of rejecting false hypotheses [307]. The FDR p-value is the p-value adjusted using false discovery rate. Pathway topology analysis performed here uses two well established node centrality measures (degree centrality and betweenness centrality [308]) to estimate node importance, thereby not only considering the number of features found in a particular pathway but taking into account the metabolite.
position within a given network to assess impact upon a particular pathway [294]. This acknowledges that changes occurring at important positions within a network can have a higher impact upon the biochemistry of that pathway and a numerical indication of this impact is shown in the final column of Table 5.5, with the maximum importance total being 1.

According to putative identifications of metabolite features uploaded for pathway topology, there are a number of metabolic pathways which are significantly perturbed by BaP treatment and a consideration of them all can afford deeper insight into the mode of action of BaP. Of particular interest is pyrimidine metabolism which is identified as the most significant pathway disrupted by BaP treatment and that upon which BaP appears to have the highest impact and this shall be further investigated.

Figure 5.9 shows the pyrimidine metabolism pathway obtained from the KEGG database via MetaboAnalyst. Compounds highlighted in red indicate 9 hits found with proposed names for matched features that were identified to be significant and within 2.5 ppm mass error. Given the wide-spread coverage of this study to cellular metabolic species, it is possible that other pathway metabolites not identified here were present but filtered out by the statistical parameters imposed on the data matrix.
To further investigate these pathway perturbations at metabolite level it is possible to explore the relative fold changes for each feature identified within the pyrimidine pathway. Box plots for m/z values corresponding to each metabolite are shown in Table 5.6. Where a feature is identified in both positive and negative ESI modes with the same direction change in concentration (increase or decrease with drug treatment), more confidence can be assigned to that identification.
<table>
<thead>
<tr>
<th>Putative feature from Pyrimidine pathway identified</th>
<th>Corresponding m/z found in positive ESI</th>
<th>Corresponding m/z found in negative ESI</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Glutamine</td>
<td>147.0764</td>
<td>203.0202</td>
</tr>
<tr>
<td></td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>169.0582</td>
<td>145.0521</td>
</tr>
<tr>
<td></td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>191.0398</td>
<td></td>
</tr>
<tr>
<td>Orotidylic acid (orotidine)</td>
<td>547.0248</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image5.png" alt="Graph" /></td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Uridine diphosphate glucose</td>
<td></td>
<td>565.0468</td>
</tr>
<tr>
<td></td>
<td></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>Compound</td>
<td>347.0248</td>
<td>323.0282</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>346.0410</th>
<th>321.0488</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>369.0066</th>
<th>360.9865</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>347.0248</th>
<th>323.0282</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>346.0410</th>
<th>321.0488</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>369.0066</th>
<th>360.9865</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine 5’-monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytidine monophosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-thymidylic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine 5’-phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For metabolites observed in the pyrimidine metabolism pathway, 7 appear to be down-regulated with BaP treatment. These are, orotidylic acid, uridine diphosphate glucose, uridine 5’-monophosphate, cytidine monophosphate, 5-thymidylic acid, pseudouridin 5’-phosphate and uridine cyclic phosphate. Uridine 5’-monophosphate and pseudouridine 5’-phosphate have the same molecular weight and therefore features shown for these compounds could correspond to one or both of them. L-glutamine appears to be the only metabolite feature observed in this pathway for which a fold change increase is observed with BaP treatment. Considering the position in the pathway of L-glutamine which is that of an entry metabolite, this is plausible. Unfortunately, an interpretation of this data is complicated by the fact that L-glutamine and ureidoisobutyric acid are isomeric with the molecular formula C₅H₁₀N₂O₃ and therefore share the same molecular weight. This suggests that ureidoisobutyric acid could also therefore be up-regulated with BaP treatment, however a consideration of the position of this metabolite in the pathway renders this more unlikely than L-
glutamine being up-regulated. No firm conclusions can therefore be drawn regarding one of these two compounds, but one or both of them show a marked increase in concentration with drug treatment and could potentially be further investigated using isotopically labelled standards for metabolic flux analysis [309].

5.3.4.1 Considering the Biological Significance of Pathway Topology Analysis Results

Pyrimidines are nitrogen containing bases consisting of a six-membered ring. Cytosine, thymine, uracil and orotic acid are the 4 key pyrimidine bases found in mammalian cells and are crucial in forming the backbone of DNA (cytosine and thymine) and RNA (cytosine and uracil). Nucleosides are synthesised when a sugar adds to a base to produce, for example, uridine and thymidine and if a phosphate is subsequently added, nucleotides are formed (e.g. AMP). Nucleotides participate in varied roles in cellular processes as they are essential sources of energy for driving most cellular reactions. Cytosine nucleotide derivatives are intermediates of phospholipid biosynthesis, many coenzymes are derived from nucleotides and these molecules also contribute to the regulation of cellular metabolism. Cells are able to produce these necessary species through de novo synthesis and therefore rapidly proliferating cancer cells are able to produce their own energy sources and synthesise building blocks for reproduction. The downregulation of the majority of metabolites observed in the pyrimidine metabolism pathway here could be indicative of a slowing down in cell proliferation with BaP treatment, rendering AML cells less tumour like and more healthy and suggesting one of the chemotherapeutic modes of action of BaP is through this pathway. The corresponding apparent upregulation of L-glutamine is in agreement with this. Glutamine is a major fuel source for rapidly proliferating cells and a decrease in the level of glutamine in the blood of AML patients (observed by 1H NMR) has been previously reported [310]. In a healthy state, glutamine would not typically be depleted to this extent, accounting for the relative higher glutamine concentrations in BaP treated cell extracts when compared to AML control cell extracts.
There are other pathways identified as being significantly affected by drug treatments as shown in Table 5.5. and a thorough biological interpretation of them all could potentially yield a further insight into drug modes of action. If these pathways are considered alongside all of the polar metabolomic data presented here, there is overlap between top 10 significant compounds putatively identified in Table 5.4 and the pathway topology analysis output. A number of the putatively identified metabolites are features of phenylalanine metabolism, amino sugar and nucleotide sugar metabolism and valine, leucine and isoleucine biosynthesis, for which pathway impact values of 0.14, 0.19 and 0.11 were calculated from pathway topology analysis respectively. This affords more confidence to the outcomes of this metabolomic study.

5.3.5 Pathway Analysis of Significant Metabolites Found in Cell Extract Study for Comparison of HL60 and K562 Cells Response to BaP Treatment

The aforementioned pathway topology analysis considers metabolite features found in both HL60 and K562 cell lines and therefore represents commonality of how BaP treatment affects both cell lines. Given the different stages of differentiation of both cell lines, it is hypothesised that an analysis of data derived from individual cell lines could yield information on how the cell lines respond differently to drug therapy.

Employing the same data selection parameters as described in section 5.3.4, a total of 340 metabolite feature names were matched with database identities for pathway topology analysis for HL60 cell extracts and 591 metabolite feature names for K562 cell extracts, providing a comprehensive analysis.

Figure 5.10 shows the pathway analysis results for metabolite features observed in HL60 and K562 cell extracts respectively with Table 5.7 detailing the numerical outputs for the top 10 significant pathways identified for each cell line.
Figure 5.10 Pathway topology analysis for features found in HL60 and K562 cell extracts respectively, identified to within 2.5 ppm. Amino sugar and nucleotide sugar metabolism is the most significant pathway affected by BaP treatment in K562 cells but is of much lower significance in HL60 cells.
<table>
<thead>
<tr>
<th>Pathway Results for HL60</th>
<th>Total</th>
<th>Hits</th>
<th>-log(p)</th>
<th>Holm adjust</th>
<th>FDR</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Glutamine and D-glutamate metabolism</td>
<td>11</td>
<td>4</td>
<td>5.3833</td>
<td>0.367</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine biosynthesis</td>
<td>27</td>
<td>6</td>
<td>4.8421</td>
<td>0.623</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>Pentose and glucuronate interconversions</td>
<td>53</td>
<td>9</td>
<td>4.8286</td>
<td>0.624</td>
<td>0.18534</td>
<td>0.16</td>
</tr>
<tr>
<td>Phenylalanine metabolism</td>
<td>45</td>
<td>8</td>
<td>4.6813</td>
<td>0.714</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td>20</td>
<td>4</td>
<td>3.1743</td>
<td>1.000</td>
<td>0.62</td>
<td>0.22</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>60</td>
<td>8</td>
<td>3.0698</td>
<td>1.000</td>
<td>0.62</td>
<td>0.16</td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td>41</td>
<td>6</td>
<td>2.9036</td>
<td>1.000</td>
<td>0.63</td>
<td>0.07</td>
</tr>
<tr>
<td>Alanine, aspartate and glutamate metabolism</td>
<td>24</td>
<td>4</td>
<td>2.5977</td>
<td>1.000</td>
<td>0.71</td>
<td>0.26</td>
</tr>
<tr>
<td>Ascorbate and aldarate metabolism</td>
<td>45</td>
<td>6</td>
<td>2.5298</td>
<td>1.000</td>
<td>0.71</td>
<td>0.11</td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td>48</td>
<td>6</td>
<td>2.2843</td>
<td>1.000</td>
<td>0.81</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathway Results for K562</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>88</td>
<td>19</td>
<td>6.1284</td>
<td>0.174</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine biosynthesis</td>
<td>27</td>
<td>8</td>
<td>5.0845</td>
<td>0.489</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Pentose and glucuronate interconversions</td>
<td>53</td>
<td>11</td>
<td>3.7189</td>
<td>1.000</td>
<td>0.49</td>
<td>0.17</td>
</tr>
<tr>
<td>D-Glutamine and D-glutamate metabolism</td>
<td>11</td>
<td>4</td>
<td>3.7139</td>
<td>1.000</td>
<td>0.49</td>
<td>0.31</td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>32</td>
<td>7</td>
<td>2.9600</td>
<td>1.000</td>
<td>0.62</td>
<td>0.13</td>
</tr>
<tr>
<td>Inositol phosphate metabolism</td>
<td>39</td>
<td>8</td>
<td>2.9103</td>
<td>1.000</td>
<td>0.62</td>
<td>0.29</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>60</td>
<td>11</td>
<td>2.8991</td>
<td>1.000</td>
<td>0.62</td>
<td>0.32</td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td>48</td>
<td>9</td>
<td>2.6607</td>
<td>1.000</td>
<td>0.62</td>
<td>0.23</td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td>41</td>
<td>8</td>
<td>2.6589</td>
<td>1.000</td>
<td>0.62</td>
<td>0.14</td>
</tr>
<tr>
<td>Alanine, aspartate and glutamate metabolism</td>
<td>24</td>
<td>5</td>
<td>2.1935</td>
<td>1.000</td>
<td>0.89</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 5.7 Top 10 pathways identified by pathway topology analysis of metabolite features affected by BaP treatment according to cell type. Amino sugar and nucleotide sugar metabolism is the most significant pathway affected by BaP treatment in K562 cells but does not feature in the top 10 significant pathways for HL60 cells. 6 pathways out of the top 10 are identified as being common between the two cell types (illustrated here by matched colour shading).
Pathway analysis reveals the metabolic pathways that are most significantly perturbed in both HL60 and K562 cell lines; with 24 hour BaP treatment are D-glutamine and D-glutamate metabolism, valine, leucine and isoleucine biosynthesis and pentose and glucoronate interconversions, as well as others. In addition, it is clear that BaP treatment impacted some metabolic pathways differently between the two cell lines.

The most striking difference of BaP impact upon cellular biochemical pathways occurs within amino sugar and nucleotide sugar metabolism which is significantly disrupted by BaP treatment in K562 cells but far less so in HL60 cells. 19 specific pathway metabolites were identified within K562 cells in contrast to just 5 in HL60 cells (not shown in the top 10 table - pathway is the 33rd in significance reported for HL60 cells). The position of these compounds within the pathway are illustrated in red in Figure 5.10. Feature identifications refer to KEGG compound identifiers.
Figure 5.11 Amino sugar and nucleotide sugar metabolism pathway with position of significant features found in metabolomic analysis of K562 control and BaP treated cell extracts shown in red.
Amino sugar and nucleotide sugar metabolism provides the cell with energy and building blocks for a multitude of metabolic processes and this pathway is closely linked with fructose and mannose metabolism, also a significant pathway disrupted by BaP treatment observed in K562 cells only. Fructose and mannose metabolism produces important nucleotide sugars (GDP-D-mannose and GDP-L-fucose) which are essential substrates for the synthesis of anchors for N-glycans, thereby having an important role in protein folding and cell-cell interactions. It is interesting that the perturbation of these pathways is more pronounced for K562 cell extracts. Considering the putative metabolite identifications that were uploaded for pathway topology, the larger number of features found in K562 indicates significantly more metabolites identified in these pathways for K562 cell extracts when compared to HL60 cell extracts. This can be accounted for in part by considering the amount of biomaterial submitted for analysis for this cell line. Data for the original biomass extracted according to the procedure described in Figure 3.14, are available and reveal an average biomass across replicates of 1.91 mg for HL60 cell extract and 2.70 mg for K562 cell extract, representing 29% more biomaterial of K562 cell extracts submitted for analysis. It is possible that this accounts for some differences observed between the pathway analysis of different cell lines and these growth-related differences should be taken into consideration when interpreting data presented in this section.

5.3.6 Considerations for Metabolomic Data Interpretation

The limitations of mass-based database searching need to be considered when interpreting metabolomics data. There are typically multiple m/z values for one metabolite which can arise from different adducts, and data presented here included numerous formate adducts (likely arising from the addition of 0.1% formic acid in the mobile phase) as well as sodium adducts and sodium formate adducts. Furthermore, mass-based identifications cannot discriminate between isomers which have the same elemental composition. Multiple identifications for metabolite features is therefore a common observation in untargeted metabolomics in which a molecular ion can reportedly have more than 100 putative identifications [302]. As observed here, this can make data interpretation somewhat challenging and in this case all sensible putative identifications suggested were taken forward for pathway analysis. Retention
time data from separation of cell extract components by UHPLC prior to infusion into the Orbitrap was able to assist in filtering \( m/z \) assignments. Furthermore, by studying two AML cell lines, further confidence is provided by considering mode of drug action on both cell lines together. If an observed response to drug is replicated across both cell lines, this increases the likelihood of the biological response being significant and thus reinforce the interpretation of drug mode of action.

To verify the putative identifications reported here from mass-based search results, standard (authentic) compounds matching those identified would have to be analysed by MS or MS-MS (or MS\( ^n \)) along with the biological sample of interest on the same method. This would allow for comparison of retention time data from UHPLC and mass spectra obtained for both the standard and sample to confirm identification or by comparing fragmentation data obtained for both. By extending to MS\( ^n \), a higher degree of confidence could be obtained for metabolites.

### 5.4 Conclusion

Metabolomic analysis of polar AML cell extracts reveals that pyrimidine metabolism is targeted by BaP treatment in both HL60 and K562 cell lines. The higher concentration of glutamine observed with drug treatment suggest a decrease in cell proliferation, which is further supported by the concurrent downregulation of 7 further metabolites putatively identified within this pathway. This could be further explored by flux analysis with isotopically labelled glutamine. Limitations of untargeted metabolomic data interpretation are considered and to definitively confirm the identification of these metabolites, matched standard compounds would be analysed according to the same UHPLC-MS methodology employed in this study. Notwithstanding this, multiple hits for metabolic pathways identified here have provided an extra degree of confidence to putative identifications.
Chapter 6 Exploring the Effect of BaP on the Lipidome of AML Cells Using Mass Spectrometry-Based Lipidomics

6.1 Introduction

The identification and quantitative determination of different classes of lipids and the vast array of molecular species therein is a powerful tool in understanding drug-cell interactions. Because of the physicochemical diversity of these biomolecules, no one mass spectrometric technique alone is currently able to cover the whole lipidome of a cell [311]. A range of mass spectrometric platforms offer the potential for lipid characterisation through different modalities, such as ToF-SIMS which can add another dimensionality to the analysis of cells through imaging capabilities. This chapter presents the findings of a lipidomics study employing UHPLC-MS and ToF-SIMS to investigate changes induced in the lipidome of AML cells with BaP treatment. Lipid classes observed to be significantly perturbed with drug treatment are reported and relative fold changes in lipid species between control and drug treated cells are further explored.

6.2 Materials and Methods

6.2.1 Cell Culture and Drug Treatment

Acute myeloid leukaemia cell lines (HL60 and K562) were cultured as described in section 3.2.2. Drug treatment with BaP was administered as described in section 3.3.3. Replicates were prepared for this lipidomics study. Six replicates of each cell line were cultured in T-75 flasks along with six control flasks. Each flask was seeded from the same pooled cell suspension as illustrated in Figure 3.15.

6.2.2 Preparation of Cell Extracts for Mass Spectrometric Analysis

Cells were extracted at 24 hours for lipidomic profiling according to the protocol described in Figure 3.14. Data presented in this chapter are from the analysis of non-polar (chloroform) extracts obtained from this extraction procedure.
For UHPLC-MS analysis, the cell extract solution was transferred to a reduced volume HPLC vial sealed with a septum containing screw cap and stored in the HPLC autosampler at 4°C during analysis. For ToF-SIMS analysis, cell extracts were deposited onto clean silicon wafers (cleaning protocol described in section 3.5.1) in 3 x 10 µL aliquots with each aliquot of sample on the wafer allowed to fully air dry air before applying the next one. Each prepared silicon wafer was washed in 0.15 M ammonium formate solution (30 s hold in wash solution followed by 5 s rinse in fresh cleaning solution) to remove any salt residue which could cause matrix ionisation effects. Wafers were air dried before sample transfer and analysis.

### 6.2.3 UHPLC-MS Methodology

Cell extracts were analysed using an Acquity UPLC system coupled to an electrospray (ESI) LTQ–Orbitrap XL hybrid mass spectrometer according to the conditions detailed in section 2.2.5.

### 6.2.4 ToF-SIMS Methodology

Cell extracts were analysed on the BioToF SIMS instrument (described in section 2.1.5) according to experimental parameters detailed in section 2.1.6.

### 6.2.5 Experimental Design for Obtaining Mass Spectrometric Data

Replicates (n = 6 for UHPLC-MS and n = 4 for ToF-SIMS) for each of the two cell lines for control and BaP treatment were analysed. This totalled 24 injections for each ionisation mode of UHPLC-MS, with technical replications of pooled QC cell extract performed at the beginning of the run (5 x QC injections) and after every 5 sample injections. All UHPLC-MS analysis was performed in an 80 hour run time and sample analysis order was randomised to eliminate any bias.

For ToF-SIMS analysis, 3 technical replicates across each wafer were obtained for each of the 4 replicates, totalling 48 spectral acquisitions for each of positive and
negative ionisation modes. This large SIMS data set was acquired over a period of 2 months.

6.2.6 UHPLC-MS Data Processing and Lipid Identification

Data files were processed using the XCMS algorithm according to the workflow described in section 5.2.5 and tentative lipid identification was carried out using Taverna as described in section 5.2.6. Metaboanalyst 3.0 was employed for multivariate analysis of lipidomics data.

6.2.7 ToF-SIMS Data Processing

Each raw spectrum was internally mass calibrated manually using well defined spectral peaks common to all samples analysed at $m/z$ values of 23.0 (Na$^+$), 86.1 (choline fragment) and 184.1 (PC lipid head group) for positive ion spectra and $m/z$ 17.0 (OH$^-$), 79.0 (PO$_3^-$) and 97.0 (H$_2$PO$_4^-$) for negative ion spectra.

Data pre-processing and multivariate analysis was carried out using in-house written algorithms in MATLAB.

For pre-processing of mass spectra, a number of steps were performed in the following order: a) square root (this was taken to reduce influence from features with high intensity in the spectra, thereby facilitating interpretation of lower signal peaks), b) vector-normalisation (sum-normalisation was also evaluated, however, grouping observed in PCA was not as clear as for vector-normalised data), c) mass-binning to 1 amu (this was necessary to overcome subtle differences in mass calibration which could result in peaks appearing in different mass channels. Binning ensured that all variables were correctly aligned in the data matrix, thereby reducing the number of data points in each spectrum and facilitating spectral deconvolution by statistical methods), d) mean centering (allowing variance from the mean of the data to be investigated).

Principal components analysis was employed to reduce the dimensionality of the multivariate data and explore any apparent grouping between samples analysed. Canonical variates analysis (CVA, a supervised technique similar to discriminant
function analysis (DFA) described in section 1.5.6) was then performed on the PC scores to discriminate between control and drug treated HL60 and K562 cell extracts (4 groups) on the basis of the retained principal components (PCs) and *a priori* knowledge of these 4 groups. The CVA code employed here is described in reference [312]. In order to address over-fitting, a cut-off was applied which was the number of PCs that explained over 95% of the variance in the data. Because SIMS data described in this chapter was not being used to build a predictive model, rather to identify any potential markers of drug action, cross validation of the CVA model was not required. The use of CVA for the analysis of ToF-SIMS data has been previously described [313].

### 6.3 Results and Discussion

Lipidomic profiling of the non-polar extracts of HL60 and K562 control and BaP treated cell extracts by UHPLC-MS (6 replicates) and ToF-SIMS (4 replicates) analysed in positive and negative ionisation modes for each technique, produced extremely large data sets. For UHPLC-MS, the data matrix contained 6907 features in positive mode and 4254 features in negative mode for each replicate analysis, which were further refined *via* the Taverna workflow to give an output of 3452 individual annotated features in positive and 2579 features in negative mode. Unfortunately, to date, there are no readily available database libraries to refine the large ToF-SIMS data matrix in a similar fashion, however both data sets were subjected to multivariate analysis for further deconvolution to assist in data interpretation. This facilitated the identification of lipid classes and individual lipid species from complex spectra obtained by both techniques. A targeted data analysis approach was then adopted to quantify relative changes in certain lipid species upon drug treatment.
6.3.1 Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by UHPLC-MS

Features annotated by Taverna were imported into MetaboAnalyst for principal component analysis to explore biological differences between the two cell lines with drug treatment. Data were subjected to log transformation and Pareto scaling to give all annotated features a normal distribution prior to statistical analysis as previously described in section 5.3.1.

An initial analysis of PCA scores and loadings (not shown) revealed a number of peaks purportedly arising directly from drugs administered rather than biological changes induced by drug. This resulted in the positive mode data matrix being refined down to 3421 annotated features (representing a removal of 31 specific drug features) and the negative mode data matrix refined to 2549 features (removal of 30 drug features). The scores plots from PCA for features identified in positive and negative modes (after removal of features arising directly from drug) are shown in Figure 6.1. Quality controls were included for this analysis by means of assessing instrument variation over the course of the experiment, and these are shown encircled in pink in Figure 6.1. For both positive and negative modes, QCs are tightly grouped at the centre of the PC space, indicating high instrument precision throughout the entire analytical run.
Figure 6.1 PCA scores plots obtained for HL60 and K562 control and BaP treated cells and corresponding QCs analysed in positive and negative ESI modes. Data shown for all features found in 6 replicates included in the lipidomic UHPLC-MS study after the removal of features attributed specifically to drug administered. Coloured ellipses represent 95% confidence limits calculated from scores.
It is immediately apparent that one biological replicate, the 6\textsuperscript{th} K562 control sample, appears to be an outlier in both positive and negative ionisation modes. When looking at the raw data for this sample, many missing values were observed in both positive and negative data matrices therefore providing justification for exclusion of this sample from further analyses.

PCA scores plots were re-plotted to exclude this outlier (and without QCs which are not included in statistical analysis of the data) and are shown in Figure 6.2.
Figure 6.2 PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts following the removal from the data matrix of K562 control outlier.
Principal component 1 clearly shows discrimination between the two cell lines, with 36.9% of the variance assigned to inherent biological differences between HL60 and K562 cells for features observed in positive mode and 38.3% for features found in negative mode. This correlates with polar metabolomics data reported in section 5.3.1.

The trend between control and BaP treatment is similar for both cell lines, with drug treated cells observed towards the positive (top half) of the PCA space, and accounting for 19.1% of the discrimination along PC 2 for data obtained in positive and 25.1% for negative mode. However, in both ionisation modes, the observed grouping is tighter for HL60 cells, with there being discrete clustering of replicates from control and drug treated HL60 cells but overlap of these groups (class membership inferred from 95% confidence ellipses calculated from scores) for K562 cells. In both modes, tight grouping is observed within the principal component space for K562 control cells, therefore the overlap can be attributed to the larger spread of K562 BaP treated replicates, which could arise from differing response to BaP within the 24 hour treatment regime. Data could also suggest that BaP may have a more pronounced effect on HL60 cells within this timeframe. To facilitate visualisation of this trend, principal components were projected into a 3-dimensional space and are illustrated in Figure 6.3.
Figure 6.3 3-dimensional visualisation of PCA scores plots obtained for HL60 and K562 control and BaP treated non polar cell extracts.

The 3-D PCA plot exemplifies the greater distance between HL60 control and BaP groups along PC2 when compared with the closer proximity of the same K562 sub classes.

6.3.2 PCA loadings analysis for UHPLC-MS non polar feature selection

Loadings plots were constructed to determine variables that were strongly contributing to the class separation between control and drug treated cells observed by PCA. In the representation of loadings (shown in Figure 6.4 and previously described in section 5.3.2), variables with loadings in a given position in the loadings plot contribute heavily to observations whose scores are found in a similar position in the complementary scores plot from Figure 6.2.
Features identified at the extremes of the plots along the loadings 2 axis for both positive and negative ionisation mode loadings (top 80 features from the positive region of each loadings plot corresponding to features contributing to grouping arising from BaP treatment and bottom 80 features from the negative region of the loadings plot representing those contributing to observed grouping of control samples) were tabulated and cross-referenced with the output from Taverna workflow.

Of these 320 features in total, 22 features in positive mode and 9 features in negative mode corresponded with significant changes consistent across both cell lines with $p$-values < 0.05 and within 2.5 ppm mass error. These statistical parameters provide a high degree of confidence in putative identification of each feature identified via database searching carried out in the Taverna workflow. Of these 32 features across both ionisation modes, 24 were identified as being lipid related species and are shown in Table 6.1, with box and whisker plots to show relative fold changes of the identified feature between control and BaP treated cell extracts for HL60 and K562 cell lines respectively.

Figure 6.4 Loadings plots obtained for HL60 and K562 non-polar cell extracts analysed by UHPLC-MS, showing $m/z$ features contributing to class separation between control and drug treated extracts along PC2 shown in Figure 6.2.
<table>
<thead>
<tr>
<th>m/z</th>
<th>Putative ID from KEGG/HMDB/LipidMaps search</th>
<th>Mass error ppm</th>
<th>Relative fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>286.3102</td>
<td>octadecenol; Stearaldehyde</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>334.2950</td>
<td>Dihydroxyoctadecanoic acid; dihydroxy stearic acid; dihydroxy-stearic acid; MG[15:0]</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>441.3348</td>
<td>ethynyl-dihydroxycholecalciferol; dihydroxycholesterol; dihydroxycholestan-one; dihydroxy-cholestan-al; Cholest-ene-triol-; cholesstan-oxo-diol; Cholesterol Dormatinol; norcholesterol;</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>494.3602</td>
<td>PC[17:0]; PE[20:0]; Arachidyl carnitine; hydroxy-cholesterol(d); epoxy-cholesterol(d); hydroxycholesterol(d)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Mass (amu)</td>
<td>Compound</td>
<td>Conc. (M)</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>668.4262</td>
<td>PE[29:2]</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>698.6657</td>
<td>DG[40:0]</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>726.6967</td>
<td>DG[42:0]</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>736.5836</td>
<td>PA[37:0]</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>776.5431</td>
<td>PS[35:1]; PG[35:3]</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>Compound</td>
<td>Ratio</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>832.5682</td>
<td>PI[36:0]; PS[35:0]</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>860.7465</td>
<td>PC[42:0]</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>902.7573</td>
<td>PC[44:0]</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>311.2214</td>
<td>Epoxy-hydroxy-octadecadienoic acid;</td>
<td>2.32</td>
<td></td>
</tr>
<tr>
<td>538.5203</td>
<td>Ceramide [34:1]; N-Palmitoylsphingosine</td>
<td>1.77</td>
<td></td>
</tr>
<tr>
<td>m/z</td>
<td>Putative ID from KEGG/HMDB/LipidMaps search</td>
<td>Mass error ppm</td>
<td>Relative fold change</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>612.3253</td>
<td>PS[21:0]; LysoPC[20:4]; LysoPC[18:1]; PC[18:1]; PC[18:0]</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>800.5193</td>
<td>PC[38:9]; PC[36:6]; PE[39:6]; PS[35:1]; PS[35:0]; PC[34:3]; PE[37:3]; Glucosylceramide [36:2]; PC[32:1]; PC[32:2]; PE[35:2]; PE[35:1]</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>301.1813</td>
<td>Oxotestosterone; Hydroxyestradiol-methyl ether</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>618.4491</td>
<td>PE[28:0]; PA[30:1]; Cer[34:2]; Cer[36:2]</td>
<td>2.07</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>Description</td>
<td>Conc.</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>642.4492</td>
<td>PE[28:0]</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>762.5411</td>
<td>PC[36:5]; PC[34:2]; PC[34:3]; PE[37:2]; PC[32:0]; PE[35:0]; SM[34:0]; Cer[40:1]</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>876.7066</td>
<td>TG[53:8]; PC[40:0]</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>966.7149</td>
<td>PC[44:1]; PC[46:4]</td>
<td>0.43</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.1 Putative identifications and relative fold changes of significant (p < 0.05) lipid related species observed in positive and negative ESI UHPLC-MS identified by multivariate analysis as contributing to differences between control and BaP treated cell extracts. m/z values listed in red indicate a decrease with BaP treatment and those in green indicate an increase with BaP treatment. Red and green box plots indicate relative amounts of each feature in HL60 BaP treated and control cell extracts respectively, whilst blue and turquoise show the same for K562 cell extracts. All species identified are within 2.5 ppm mass error.

Data strongly suggest that cellular lipid pathways are a shared target of BaP therapy in these cell lines. There could be some correlation with general saturation state of lipid species tentatively identified, as fully saturated species observed in positive ionisation mode generally appear to decrease with BaP treatment and polyunsaturated species with increasing number of double bonds are associated with a fold change increase with BaP. Features observed at m/z 776.5431 and m/z 668.4262 are, however exceptions to this, being unsaturated species that appear to decrease with BaP and this same trend is not so apparent for negative mode ionisation data. A detailed biological interpretation of this data is limited by the fact that multiple species can be detected for any given molecular formula, and this is further complicated by adduct formation during electrospray ionisation. Features identified in Table 6.1 represent a variety of adduct species including potassium, sodium and sodium formate adducts. To improve the identification of lipid species by UHPLC-MS, the scanning modes of tandem mass spectrometry could be employed to detect fragments of different lipid classes which will have different characteristic fragment ions or neutral losses for the building blocks of the molecule. This approach is well utilised and widely reported [314] and would provide further confidence to data reported here. Furthermore, lipid standards would need to be analysed for confident assignment of putative identifications reported here.

ToF-SIMS, however, typically suffers less from complex adduct formation of individual biomolecular species and the possibility to sputter and detect intact lipids
from the surface of a sample is highly desirable. Whilst the technique does not afford the high degree of mass accuracy achieved by UHPLC-MS and can contribute interferences of another kind (for example, isobaric interference and matrix effects), if SIMS data obtained is interpreted in conjunction with the UHPLC-MS data presented here, a further mass spectrometric dimension can be added to this lipidomic study and this is explored in 6.3.3.

6.3.3 Unsupervised Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by ToF-SIMS

The large SIMS data matrices of calibrated spectra obtained for positive and negative ionisation modes between $m/z$ 0-1000, were initially subjected to PCA to explore any obvious trend between classes of data. Figure 6.5 shows the scores plots for PCs 1 and 2 for both positive and negative ionisation modes.
Figure 6.5 PCA scores plots obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive and negative ionisation modes. Each point represents a replicate analysis from experimental replicates (n=4) and technical replicates (n=3).

Principal components 1 and 2 do not capture any obvious trend in grouping according to either cell type or drug treatment, however there may be evidence of some grouping arising from the 2 month time frame over which replicates were analysed. In total 9 principal components were examined for data obtained from each ionisation mode with no clear indication of variance arising from BaP treatment. Another possible cause for lack of grouping arising from inherent cellular biology across the population.
of cell extracts, could be due to the entire mass range being analysed, which is dominated by low mass peaks. Figure 6.6 shows a representative positive ion SIMS spectra of the mass range \( m/z \) 50 – 1000, which is dominated by the phosphocholine head group at \( m/z \) 184 peak. In-lays of enlarged views of the spectra (\( m/z \) 300 – 900 and \( m/z \) 572 - 587) for control HL60 cell extract reveal an abundance of biological species, including diacylglycerols (DAGs).

![Figure 6.6 Positive ion characteristic ToF-SIMS spectra (\( m/z \) 0-1000) and in-lay of enlarged spectra (\( m/z \) 300-900) showing an abundance of biological species for BaP treated K562 cell extract.](image)

The positive ion data matrix was thus further refined to eliminate high abundance, low molecular weight ions and PCA was subsequently performed on a reduced region of the spectra for each cell type. Figure 6.7 shows principal components 3 and 4, for K562 control and BaP treated data obtained between \( m/z \) 400-800 where many lipid related species reside. Data were scaled by fourth-rooting rather than square-rooting here for optimum presentation of results.
Figure 6.7 PCA scores and loading 3 plot obtained for K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive ionisation modes for lipid related features found in the region m/z 400-800. Each point represents a replicate analysis from experimental replicates (n=4) and technical replicates (n=3), extreme outliers removed.

Principal component 3 appears to discriminate between control and drug treatment for K562 cell extracts included in this analysis. Out of 12 possible data points for each group projected into the PC space (4 reps, triplicate analyses across wafer), 8 were included for BaP treated extracts and 11 for control cell extracts, with the remaining samples excluded as extreme outliers in the PC space. An investigation into the corresponding loadings for PC 3, reveals peaks at m/z 552 and 580 in the positive loadings, suggesting higher contribution of these species to the biochemistry of control rather than BaP treated cell extracts, and m/z 604 in the negative loadings, indicating an increased contribution from this species with BaP treatment. Analysis of lipid species within biological systems using ToF-SIMS is well established and lipid peak assignment is widely reported [112]. This allows putative assignment of the lipid species observed here as diacylglycerols. Specifically, m/z 552 and 580 representing DAG(32:0) and DAG(34:0) respectively (commonly reported m/z values for these species are 551.5 and 579.5, which correlate given nominal mass binning performed here which would round up these values to the next integer) and m/z 603 representing DAG(36:2) (commonly reported as m/z 602.5). This data thereby suggests that fully saturated DAG could be more prevalent in control cell extracts and maybe susceptible to depletion by BaP, with the possibility of polyunsaturated species (such as
DAG(36:2) increasing with BaP treatment. Whilst no firm conclusions can be drawn from this relatively small data set alone, this effect is further explored using supervised multivariate analysis described in section 6.3.4.

6.3.4 Supervised Multivariate Analysis of HL60 and K562 Cellular Lipid Profiles by ToF-SIMS

Principal component-canonical variates analysis (PC-CVA) was performed on the data matrices for both positive and negative ionisation modes with spectral pre-processing according to parameters described in section 6.2.7. Biological and technical replicates were included for all samples except four spectra obtained from K562 BaP samples and one from HL60 BaP which were excluded as outliers. A variety of mass ranges were selected for this analysis with the aim of excluding dominating peaks in the spectrum which masked the potential for discrimination according to interesting biology. Of the mass ranges explored, \( m/z \) 300 − 900 was chosen as the largest available data matrix to describe the data within stable discriminations. Figure 6.8 shows CVA scores and loadings obtained for cell extracts analysed in positive mode SIMS.
Figure 6.8 PC-CVA scores and loading for CV 2 obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in positive ionisation mode, for features observed in the region m/z 300-900.

The PC-CVA scores plot shows discrimination between cell type along CV 1, which describes 94.2% of the variance within the data set and between control and drug treated cell extracts along CV 2, which describes 5.08% of the variance contributing to the observed grouping of control and drug treated classes. There is some noticeable inter-class overlap for K562 cell extracts however, HL60 cell extracts classify into
their respective groups well. This trend correlates with data obtained by UHPLC-MS, described in the PCA scores plot in Figure 6.2 and further suggests a more marked effect of BaP upon the lipidome of HL60 cells when compared to K562 cells.

Data are further interrogated in the loadings plot, in which features observed in positive loadings are more characteristic of drug treated cells than control cells, and features observed in the negative loadings are more correlated with control cells and represent compounds which are likely to be down regulated with BaP treatment.

Positive loadings between the $m/z$ range 500-700 contain a large number of lipid features which are highly correlated with BaP treatment. This region contains many DAG species which appear to be up-regulated with BaP such as $m/z$ 604 corresponding to DAG (36:2). This phenomenon can be further explored by extracting DAG peak areas from the original data matrix and data for this are reported in section 6.3.5.2. Interesting peaks that are present in the negative loadings which are suggestive of high correlation with control cells are $m/z$ 349, 370, 473, 489, 733, 756 and 772. Many of these features do not directly correspond with $m/z$ values previously reported in literature and therefore tentative assignment of them remains a challenge. To overcome this, the ability to acquire data of higher mass accuracy or to obtain fragmentation information with tandem MS would be desirable.

Figure 6.9 shows CVA scores and corresponding loadings obtained for cell extracts analysed in negative mode SIMS.
Figure 6.9 PC-CVA scores and loading for CV 2 obtained for HL60 and K562 control and BaP treated cell extracts analysed by ToF-SIMS in negative ionisation mode, for features observed in the region m/z 300-900.
Again, distinction between the 2 cell types is the main discriminatory factor within the data, contributing to 98.9% of the variance within the data matrix and showing tight grouping for these two classes along CV 1. CV 2 describes 0.99% of the remaining variance within the data and is suggestive of classification according to drug treatment, however grouping is more dispersive than observed with positive ionisation data and there is overlap between control and BaP treated cell extracts for both HL60 and K562 cell types. This might suggest that negative ionisation data obtained throughout the duration of the ToF-SIMS experiment is less robust, which could be due to negative SIMS ionisation at an instrument level being more variable across the 2 month experimental time-frame than positive mode SIMS. An interpretation of the loadings for CV 2 reveals a larger number of features correlating with the negative loadings, which are more representative of control cells. Positive loadings which are more likely arising from drug treatment contain far fewer features which suggests that a number of biological species could be down-regulated within both cell lines during BaP treatment. Interesting features showing correlation with control cells are $m/z$ 305, 325, 377, 403, 473, 616, 642 and 735 to name a few. Many of these features do not directly correspond with $m/z$ values reported in literature and therefore tentative assignment of them is challenging, however, $m/z$ 385 has a noticeable contribution to the negative loading on CV2 and can be tentatively assigned to cholesterol based upon literature citations [112]. This is suggestive of BaP treatment decreasing cholesterol in HL60 and K562 cells.

Multivariate analysis of a large ToF-SIMS data matrix presented here certainly indicates that cellular lipid biosynthesis is a target for the mode of action of BaP therapy. The more apparent discrimination of BaP treatment in HL60 cells as observed by tighter grouping (when compared to K562 cells) in UHPLC-MS PCA scores and ToF-SIMS PC-CVA scores suggests that BaP has a greater impact on HL60 cells. This can be explained in terms of a consideration of the growth rate of respective AML cell lines. HL60 cells have a doubling rate of 22 – 24 hours, whereas K562 cells double in approximately 24 – 26 hours. This means that depending upon at which stage in the cell cycle a K562 cell is upon dosage, not all cells will have gone through a complete cell cycle during the 24 hour treatment regime and therefore the effect of BaP upon these cells will be to a lesser extent, thus giving rise to a wider spread in mass spectrometric data obtained for lipid profiles. As HL60 cells proliferate faster, there is
an increased demand for the synthesis lipids to form new cell membranes which will occur for all HL60 cells within the treatment period, regardless of cell cycle stage. To further investigate BaP effects on each cell line, CVA could be performed on data derived from HL60 alone and K562 alone which may afford some insight into treatment impact according to cell differentiation.

### 6.3.5 A Targeted Approach to Determine Lipid Saturation Status in Control and BaP Treated Cell Extracts Analysed by ToF-SIMS

A further understanding into the mechanistic action of BaP on AML cellular lipids can be derived from adopting a targeted analytical approach to determine lipid saturation of different classes of lipid species. UHPLC-MS data presented in section 6.3.2 was suggestive of saturated lipid species being more correlated with control cells and polyunsaturated lipids more correlated with cells after BaP treatment. Furthermore, results from spectroscopic studies applied to this system (presented in Chapter 4) indicate a significant change to the methylene:methyl ratio derived from lipid bands in the spectra. It is therefore of interest to target lipid saturation changes observed within the ToF-SIMS data which can provide further confidence in biological interpretation by employing this complementary analytical platform.

Peak areas were extracted from the original data matrices and tabulated for known lipid species present in cell extracts of both AML cell lines studied. For all peak areas reported here, average values were calculated across technical replicates (triplicates) for each biological replicate analysed and reported as bar charts to aid visualisation of lipid change with BaP treatment. Data are reported for fatty acid acyl species observed in negative ionisation mode and DAG lipid species observed in positive ionisation mode.
6.3.5.1 Probing BaP Induced Saturation Changes of Fatty Acid Acyl Chains

Fatty acid acyl chains are key building blocks for the majority of cellular lipid species and contribute to the huge diversity in these biomolecules. For example, the glycerophospholipid phosphatidylcholine (PC), can present with different fatty acyl compositions which may differ in carbon chain length, the number of double bonds present in the fatty acyl chains and the position of the double bonds within the hydrocarbon chain [311].

Fatty acids ionise readily in negative ToF-SIMS and have been widely reported in a variety of biological matrices. Here, fatty acid acyl chains for C16 and C18 chain lengths were identified in control and BaP treated AML cell extracts. Figure 6.10 shows a characteristic negative ion ToF-SIMS spectrum for a BaP treated cell extract.

![Figure 6.10 Negative ion characteristic ToF-SIMS spectrum of BaP treated cell extract showing m/z range 100 – 500 with fatty acid acyl chains identified.](image)
The two regions of the negative ion spectra highlighted in Figure 6.10 show fatty acid acyl chains for C16 and C18 chain lengths. The relative intensities of these different fatty acids can be compared because the secondary ion yield of the compounds are very similar as a result of them having similar physicochemical properties [315]. This allows specific information regarding the saturation states of these fatty acid acyl lipids to be determined.

Peak areas were extracted for fatty acyl chains observed in all four replicates for control and BaP treated cell extracts. These corresponded to fatty acids C16:0 (palmitic acid), C16:1 (palmitoleic acid) and C16:2 (hexadecadienoic acid) at m/z values 255, 253 and 251 respectively and fatty acids C18:0 (stearic acid), C18:1 (oleic acid) and C18:2 (linoleic acid) at m/z values 283, 281 and 279 respectively.

Data were tabulated for each individual fatty acid and normalised to the total sum of fatty acid peak areas within each region (eg. for C16, the sum of C16:0, C16:1 and C16:2). The normalised values were averaged across the three replicate analysis for each wafer and results are shown in Figure 6.11 and Figure 6.12 for HL60 cell extracts and K562 cell extracts respectively. Standard deviation error bars are included, calculated from the actual data for each bar shown to indicate typical reproducibility of the measurements.
Figure 6.11 Plot to show average normalised peak areas for fatty acid acyl chains of different degrees of saturation, identified in HL60 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in negative ionisation mode. No data was determined to be significant, p-values > 0.05 obtained for all data in this series.

Figure 6.12 Plot to show average normalised peak areas for fatty acid acyl chains of different degrees of saturation, identified in K562 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in negative ionisation mode. Significant data indicated with * (p < 0.05).

The observed saturation of C18 fatty acid chains appears to be consistent between HL60 and K562 cell extracts, with BaP treatment causing a small increase in the saturated C18:0 and polyunsaturated C18:2 fatty acids, with a more noticeable decrease in the monounsaturated C18:1 fatty acid across both cell lines. Significance for these changes was determined using a student t-value t-test and p-values suggest
that the BaP induced observed decrease in C18:1 from K562 cell extract is significant ($p = 0.02$), but no further C18 fatty acid data returned a significant value. This is not surprising given the overlap of error bars for some lipid species.

For C16 fatty acid chains in HL60 cell extracts, the trend is also consistent, with relative amounts of C16:0 and C16:2 increasing with BaP and a corresponding noticeable decrease in C16:1. However, despite the trend between C18 and C16 fatty acids appearing consistent, no data derived here for HL60 cells was determined to be significant. Error bars (standard deviation) are suggestive of reproducibility for this negative ion HL60 data, however low level of signal may have contributed to high $p$-values. A modified experimental strategy may overcome this issue as discussed in section 7.4. Data are not consistent with this trend for K562 cell extracts, for which C16:0 decreases and C16:1 increases with BaP although $p$-values were significant at $< 0.05$ for this data series.

This data set is challenging to interpret biologically and could be suggestive of fatty acyl lipids being under a high state of flux within the cell. Since these species are the building blocks of many other lipid classes, as well as appearing as free fatty acids utilised in a variety of cellular pathways, there may not be a consistent trend of drug treatment within this lipid class. For one monounsaturated species observed here, C18:1, there is significant evidence for there being a decrease with BaP treatment in K562 cells. It is reported that cancer cells have a relatively high content of saturated and monounsaturated fatty acid due to elevated de novo lipogenesis and high levels of the enzyme stearoyl-CoA desaturase (SCD1) which is responsible for producing oleic acid (C18:1) through the action of desaturation of stearic acid (C18:0) [316]. Data here suggest that this particular enzyme could be deactivated with BaP treatment, given the apparent decrease in monounsaturated fatty acid acyl species observed. This correlates with recently reported data which suggests that BaP therapy could be disrupting the SCD1 gene coding for this biochemical process [218]. Another significant observation is the increase in K562 C16:2 with BaP suggestive of an increase in polyunsaturation with BaP which is consistent with changes reported in Chapter 4. The lack of significant findings from HL60 cells might also have something to do with high turnover rate of these species in this faster growing cell line, however no certain conclusions can be drawn.
The concept of ‘bioactive lipids’ or lipids that activate specific signalling pathways, has been proposed, as eyes that ‘see’ the nutritional or multicellular signalling status of the environment.[317, 318]. These lipids, which are thought to be synthesized from membrane lipids, fatty acids, cholesterol and amino acids [319], act as a sensing system that enables the cell or tissue to make decisions whether to proliferate, differentiate or undergo apoptosis. These bioactive pathways can be activated by toxic or noxious substances including drugs and could also potentially explain the fatty acid lipid changes observed here.

### 6.3.5.2 Probing Saturation Changes of Diacylglycerols in AML Cells Treated with BaP

The enlarged region of the positive SIMS spectra shown in Figure 6.6 shows a variety of intact molecular species and clearly identifies peaks attributed to the DAG(34:2), DAG(34:1) and DAG(34:0) series, corresponding to \( m/z \) values of 575.5, 577.5 and 579.5 respectively. Furthermore, peaks representing the DAG(36:3), DAG(36:2), DAG(36:1) and DAG (36:0) series at \( m/z \) 601.5, 603.5, 605.5 and 607.5 respectively, are also present in all positive SIMS spectra and can be further investigated. An interrogation of the loadings vector for CVA of cell extracts shown in Figure 6.8 was suggestive of the presence of DAG species being more correlated with BaP treated cell extracts.

Peak areas were extracted for the aforementioned DAG peaks for control and BaP treated cell extracts for 4 replicates for both cell types analysed for the triplicate areas analysed on each wafer. The peak areas for each lipid within a DAG series (C34 and C36) were normalised to the sum of DAG peak areas for that series and then averaged across technical replicates for each wafer. Finally, average peak areas for replicates were averaged and plotted for each DAG lipid observed. Results can be seen for HL60 and K562 cell extracts in Figure 6.13 and Figure 6.14 respectively. Standard deviation error bars are included for each bar shown to indicate typical reproducibility of the measurements.
Figure 6.13 Plot to show average normalised peak areas for diacylglycerols of different degrees of saturation, identified in HL60 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in positive ionisation mode. No data was determined to be significant, p-values > 0.05 obtained for all data in this series.

Figure 6.14 Plot to show average normalised peak areas for diacylglycerols of different degrees of saturation, identified in K562 control (blue) and BaP treated (red) cell extracts by ToF-SIMS analysis in positive ionisation mode. Significant data indicated with *(p < 0.05).
A consistent trend for DAG saturation is observed, both intracellular for the two discrete DAG(34:x) and DAG(36:x) series analysed, but also across both HL60 and K562 cell lines. Relative amounts of saturated and monounsaturated DAGs appear to decrease with BaP treatment in both cell lines, with a concurrent increase observed for polyunsaturated DAG species containing 2 or 3 double bonds. This suggests an overall increase in the level of acyl chain unsaturation of DAGs with BaP treatment.

To determine the significance of these changes, a student’s 2-value t-test was performed across control and BaP treated cell extracts for each biological replicate. Figure 6.15 shows individual replicate peak areas for DAG(36:0) and DAG(36:2) lipids observed in K562 cells. These data series presented were determined to be statistically significant with p-values of 0.026 and 0.018 respectively. All other data series were determined to be of lesser significance, all with p-values > 0.05. This does not mean that some DAG saturation data presented here are not biologically valid, just that a degree of caution should be exercised if interpreting this data alone. This is where the added value of a multi-disciplinary approach to the lipidomic analysis of drug-cell interactions becomes apparent. This trend in saturation state data correlates well with the previous suggestions that unsaturated species increase with BaP treatment which was based upon observations of relative fold changes of lipid species putatively identified by UHPLC-MS shown in section 6.3.2.
Figure 6.15 Plot to show saturated and unsaturated DAG normalised peak areas for 4 individual replicates of K562 control (blue) and BaP treated (red) cell extracts identified by positive ionisation ToF-SIMS. Both DAG(36:0) and DAG(36:2) data series were calculated to be statistically significant (p = 0.026 and 0.018 respectively).

DAGs play an important role in the cell as signalling molecules [320]. During DAG synthesis, fatty acids are incorporated into DAGs via lysophosphatidic acid and phosphatidic acid such that DAG (36:2) is typically synthesised from two C18:1 fatty acid chains. DAGs not only exist as synthesis products from fatty acids but can also be found in cells as a breakdown product from phospholipids and as a result of these two processes, DAGs are transient species within a cell.
In section 6.3.5.1, a significant decrease in C18:1 was observed for K562 cells. As a result of this, a concurrent decrease in DAG(36:2) might also be expected if DAGs characterised here were being synthesised from fatty acids observed. This was not the case, with DAG(32:2) apparently increasing with BaP treatment. It can therefore be speculated that intracellular DAGs here are arising from the remodelling of phospholipids. Phospholipids are the most abundant lipid species found in cell membranes and during PL remodelling (breakdown), DAG species are released and can subsequently activate specific signalling pathways within the cell. It should be remembered, that data captured here (at one 24 hour time point post-drug treatment) provides a snapshot to cell phenotype at that moment in time and given the transient nature of both fatty acids and DAGs, no firm conclusions can be drawn as to whether DAGs here arise from FA synthesis or PL breakdown.

### 6.4 Discussion and Conclusions

Employing UHPLC-MS and ToF-SIMS to probe the effect of BaP therapy on cellular lipid composition has provided a useful insight into the action of this therapy. The complementary nature of these two mass spectrometric methodologies adds an extra degree of confidence to the data when a perceived biological response is replicated in both techniques. A general increase in polyunsaturated species for BaP treated cell extracts was observed, along with suggestion of a specific decrease in C18:1 monounsaturated species in K562 cells. Given reports that enhanced lipogenesis of cancer cells renders them less susceptible to lipid peroxidation and apoptosis by limiting the degree of membrane polyunsaturation, the observed effect of BaP treatment suggests anti-leukaemic activity through altering cellular lipid biochemistry, thereby rendering AML cells phenotypically less cancer like. Significant lipids tentatively identified here could be confirmed through the analysis of standards to add a further degree of confidence to these findings. Mass spectrometric data correlates well with outcomes from the spectroscopic study reported in Chapter 4. Employing the imaging capabilities of ToF-SIMS in a continuation of this study would add a further dimension to this data set by providing spatial information of cellular lipids on a single cell level.
Chapter 7 Project Conclusions and Future Work

This project has demonstrated that adopting a multi-disciplinary mass spectrometric and spectroscopic approach to cell analysis can afford a powerful insight into understanding drug mode of action at a cellular level. HL60 and K562 acute myeloid leukaemia cells were treated with a combination drug therapy (BaP) of known clinical potential and the effect of this treatment was investigated after 24 hours using complementary analytical platforms. The observation of consistent biological response to drug treatment in both AML cell lines, provided further confidence to data interpretation of the anti-leukaemic properties of BaP. This approach could be extended to the study of drug-cell interactions for other oncological systems.

Live cells were probed using Synchrotron FTIR, cell extracts were analysed with UHPLC-MS, ToF-SIMS, ATR and Raman spectroscopy, and fixed cells were imaged with Raman microspectroscopy, all of which yielded a large data set covering a diverse range of biomolecular species for further interpretation. A comparison of findings from multiple techniques revealed consistent disruption to cellular lipid species with BaP treatment, suggesting that disruption to lipogenesis is a key target of the anti-leukaemic action of BaP. Furthermore, metabolomics analysis highlighted pyrimidine metabolism, D-glutamine and D-glutamate metabolism, phenylalanine metabolism and galactose metabolism as significant pathways upon which BaP appears to have a pronounced effect.

7.1 A consideration of sample preparation and optimisation of instrument parameters facilitated the acquisition of high quality data covering a range of biomolecular compounds.

Due to the diverse nature of analytical platforms employed throughout this project, no one universal sample preparation methodology was possible. For all techniques employed, the quality of resulting data is to a large degree limited by the quality of sample being studied and the manner in which it is prepared for that particular technique. In Chapter 3, an investigation into cell fixation was reported and extraction protocols were established for efficient extraction of intracellular metabolite and lipid species. Substrates employed for deposition of cells and cell extract ranged from a
microfluidics cell device for S-FTIR analysis of fully hydrated cells, to silicon wafers for ToF-SIMS analysis of cell extracts. Method development work fell into the following 3 categories:

1. Ensuring single cells and extracts were of high integrity for analysis.

   The goal of fixation is to preserve the sample and its biochemical components in as close a state to that in vitro. Here methanol and formalin fixation were evaluated using Raman microspectroscopy. No noticeable difference of any significance was observed with lipid bands in the high wavenumber region of resulting Raman spectra (2850-3100 cm\(^{-1}\)) between methanol and formalin fixed cells. Formalin was chosen for the fixative of choice for single cells analysed in this project given previous reports of methanol causing lipids to move and potentially leach from cells. A number of recent spectroscopic studies of cells have also employed formalin fixation, one study in particular reports formalin fixed cells displaying the closest Raman spectral resemblance to that of live cells [83]. This supports the decision to employ formalin as a fixative in this study.

   A cell extraction protocol was developed and evaluated for the efficient extraction of intracellular polar and non polar biomolecules to ensure adequate biomass for analysis. UHPLC-MS was employed to evaluate different extraction protocols, with the final method selected for this project being an adaption of the Folch and Bligh and Dyer extraction methods [73]. The final composition of extraction solvent used was methanol:water:chloroform 1:1:1. This more ‘one-pot’ approach offers an alternative to the widely employed Nature protocol extraction methodology [279] and differs in there being reduced sample transfer steps between microcentrifuge tubes which facilitates a more rapid extraction procedure with potentially reduced sample loss between steps.

2. Establishing sample preparation protocols fit for the analytical technique employed.

   Where relevant, for each analytical technique, sample loading and preparation options were investigated to ensure good signal to noise for biological species of interest thus yielding high data quality. Sample loading for ToF-SIMS analysis as a function of ion yield was investigated for ToF-SIMS, with a sample loading volume of 30 µL (loaded in 3 x 10 µL aliquots) being selected as most desirable.
Matrix effects are a common interference observed in the SIMS ionisation process for molecular species. Here, the largest source of such interference was with salts found in the cell extract biological matrix. A wash protocol was developed to de-salt cell extracts and whole cells deposited on silicon wafers prior to SIMS analysis, which consisted of a 30 second rinse of sample in 0.15 M solution of ammonium formate solution for 30 seconds and a second rinse in fresh ammonium formate solution for a further 5 seconds. This methodology is in agreement with a number of biological SIMS studies which employ ammonium formate [92] or ammonium acetate [321] to remove excess salt which can crystallise and interfere with SIMS sputtering and ionisation processes.

3. Optimisation of Instrument parameters for each analytical platform.
This was an essential feature of each experiment as parameters were carefully adjusted to achieve the required sensitivity for the simultaneous determination of a wide range of biomolecular species. For Synchrotron-FTIR, Raman microspectroscopy and ToF-SIMS, data acquisition time was limited by other parameters such as life span of live cells in microfluidics device before being dislodged by air (synchrotron-FTIR), cells burning with prolonged exposure to high intensity laser power or long scan rates (Raman microspectroscopy) and sputtered ion yield within the static limit (ToF-SIMS). Defining instrument parameters for each typically required many repeat analyses with often small, sensitive adjustments to methodologies. Considerations on a macro scale involved comparing data quality obtained from two lasers (785 nm and 532 nm for Raman) and two ion guns (Au$_3^+$ and C$_{60}^+$ primary ion beams for SIMS), with the 532 nm laser and Au$_3^+$ being selected as fit for purpose for Raman cell imaging and ToF-SIMS cell extract analyses respectively.
7.2 A multi-platform spectroscopic investigation was employed to study the action of BaP therapy in AML and revealed biochemical perturbations on a single cell level.

Chapter 4 reports the findings from a large scale, multi-platform spectroscopic study of BaP treatment in AML cells which were analysed in 3 states: fully hydrated living cells, formalin fixed cells and cell extracts and which yielded a rich data set for interpretation. The high photon flux of a synchrotron source was exploited for the analysis of a large population of live AML cells. Spectra were acquired for more than 800 single cells and data revealed significant changes to vibrations associated with cellular lipids species. In particular, the ratio of CH\textsubscript{2}:CH\textsubscript{3} for both symmetric (2850 cm\textsuperscript{-1}:2870 cm\textsuperscript{-1}) and asymmetric (2920 cm\textsuperscript{-1}:2955 cm\textsuperscript{-1}) stretches was observed to increase with BaP treatment and this was determined to be a significant change in both cell lines ($p < 0.05$). This observation was further corroborated with ATR analyses of cell pellets for both cell lines for which the same significant changes were observed. Furthermore, an examination of the $\nu_{\text{asym}}$(PO\textsubscript{2}\textsuperscript{-}) band of DNA which reportedly shifts from 1220 cm\textsuperscript{-1} to 1240 cm\textsuperscript{-1} with a conformation change from B-DNA to A-DNA (when cells dehydrate or begin to die), revealed no significant change in the spectra of hydrated cells ($p = 0.198$ for K562 cells and $p = 0.774$ for HL60 cells), confirming that lipid changes observed are not simply a result of cells entering drug-induced apoptosis.

These findings show that BaP treatment targets cellular lipid species and could be rendering lipids more unsaturated, from which membranes would have increased fluidity and therefore have a phenotype more characteristic of non-cancerous cells.

Preliminary work on individual cell nuclei which were extracted from K562 cells and probed with S-FTIR, suggested that BaP treated nuclei ($n = 43$) can be discriminated from control cell nuclei ($n = 50$) from a consideration of the nuclei spectra. This is interesting, since S-FTIR of living cells indicated that the drug was not intercalating with DNA and therefore suggests other pathways within the nucleus being perturbed. Future work to expand the data set to hundreds of individual nuclei across a statistically relevant number of replicates and across both AML cell lines would provide a strong basis for interpreting the nuclear targets of BaP, for which lipids are strongly implicated. At present there are a limited number of nuclei spectroscopic studies reported in the literature and few that address drug induced changes across both whole cells and cell nuclei, which denotes one novel aspect of work undertaken here.
Raman microspectroscopy was employed to map intra-cellular lipid changes across single cells. In K562 cell images (n = 11), BaP was shown to give rise to an increase in the CH$_2$ symmetric stretch from lipid acyl chains (observed at 2850 cm$^{-1}$). These results supported the findings of S-FTIR and ATR-FTIR and furthermore indicate that the observed increase in CH$_2$:CH$_3$ is a result of lipid unsaturation with BaP treatment. Raman microspectroscopy added a further dimension to the study by providing spatial information of lipid distribution across single cells which suggested that BaP-induced lipid saturation change is uniform across a single cell. Future work would apply the same methodology to a larger population of AML cells (to include HL60 and increased number of K562 cells) to further confirm this response.

The wide scope of this spectroscopic study contributes an extra dimension to similar studies previously reported in the literature. Typically, a study investigating drug-cell interaction has focussed on one or two spectroscopic techniques and has been limited to probing cells of one morphology. Here, employing three distinct spectroscopic techniques, in conjunction will live, fixed and extracted cells has contributed a novel approach to the existing literature and exemplified the benefits of corroborating data across complementary spectroscopic platforms.

7.3 A metabolomics investigation into the action of BaP therapy in vitro by UHPLC-MS revealed a number of metabolic pathways that are disrupted by BaP therapy.

Metabolomic analysis of polar AML cell extracts reported in Chapter 5 tentatively identified in excess of 1500 significant features (ANOVA p < 0.05) across positive and negative UHPLC-MS ionisation modes. Multivariate analysis revealed clear discrimination of sample classes, with HL60 and K562 cell lines showing the greatest variance along PC1 of the scores plots and BaP drug treated cells separating from control cells along PC2. Loadings were interrogated and the top 80 loadings from the extremes of the loading vector (positive and negative loadings corresponding to BaP treated and control extracts respectively) were taken and tabulated against the top 20 significant features determined by ANOVA for each ionisation mode. All but 2 features matched findings from both strategies and tentative identifications for these was reported.
Confidence was further assigned to significant features by assessing the false discovery rate for this data set using the SAM technique. FDR’s were reported as 0.004 for positive mode features and <0.0005 for negative mode features. Of the top 10 significant features identified by ANOVA and univariate analysis, all were identified to be within the top significant features reported after false discovery rate cut offs were applied; and putative identifications were suggested for as many of these 20 significant species as possible. A variety of biomolecules were indicated as being affected by BaP treatment, and were suggestive of amino sugar, nucleotide sugar and carbohydrate metabolism and lysine biosynthesis, all being pathway targets of BaP. It was also interesting to observe that all features were found to be up-regulated in drug treated cells for both cell lines analysed as indicated by the higher relative concentrations of each feature in box and whisker plots plotted to assess relative fold change of species observed.

Pathway topology analysis was performed to probe specific pathway targets of BaP in more detail. Data revealed a number of biochemical pathways that were indicated as targets of BaP therapy consistent across both HL60 and K562 cell lines. Out of these, pyrimidine metabolism was the most significant for changes observed in both HL60 and K562 cell lines.

The higher concentration of glutamine observed with drug treatment suggests a possible decrease in cell proliferation, which is further supported by the concurrent downregulation of 7 further metabolites putatively identified within this pathway: orotidylc acid, uridine diphosphate glucose, uridine 5’-monophosphate, cytidine monophosphate, 5’-thymidylic acid, pseudouridine 5’-phosphate and uridine cyclic phosphate. Specific fold changes for each of these metabolites tentatively identified, were tabulated to further quantify this apparent downregulation.

Metabolomics UHPLC-MS studies of drug effect at a cellular level similar to that employed here are widely reported in the literature but none have addressed the novel combination therapy BaP by this technique. It is encouraging, however, that there is some overlap of data findings with a previously reported metabolomics 1H NMR study carried out by project collaborators, which characterised the metabolic features associated with AML from the serum of patients with de novo AML against age and gender matched healthy controls. Significant metabolomics differences between the
two classes were observed, for which there were differences in multiple metabolic pathways including glycolysis/glucogenesis, TCA cycle, biosynthesis of protein and lipoproteins and metabolism of fatty acids. Whilst this NMR study did not address drug treatment, of particular interest is that AML patients were found to have lower levels of glutamine when compared to healthy controls. This strongly supports the findings of the metabolomics study presented here in which glutamine was observed to be upregulated with BaP treatment and therefore suggests a change towards a healthier cell phenotype with drug treatment.

Limitations of untargeted metabolomic data interpretation were considered and to definitively confirm the identification of these metabolites, matched standard compounds would be analysed according to the same UHPLC-MS methodology employed in this study.

Metabolic flux analysis would provide a deeper insight into perturbations to metabolic pathways observed here [309]. Future work might employ $^{13}$C labelled metabolites specific to the pathways identified here (for example $^{13}$C glutamine) to determine flux distributions in metabolism of BaP treated cells.

### 7.4 ToF-SIMS and UHPLC-MS proved to be powerful tools for probing the lipidome of AML cells and revealed saturation changes in cellular lipids post-BaP treatment.

Employing complementary mass spectrometric techniques to probe the effect of BaP treatment on cellular lipid composition in Chapter 6, provided large coverage of the lipidome of AML cells through untargeted and targeted approaches and provided further insight into the action of this therapy.

For data derived by UHPLC-MS and ToF-SIMS, a general increase in polyunsaturated species for BaP treated cell extracts was observed which correlated well with findings from spectroscopic investigations reported in Chapter 4. Furthermore, a targeted approach to interpretation of ToF-SIMS lipidomic data revealed a significant ($p < 0.05$) decrease in C18:1 monounsaturated species in K562 cells. Significance for fatty acids observed in HL60 cells was low and a modified experimental procedure to improve signal intensity through increased primary ion beam shots would be expected.
to yield data of higher significance. The polyunsaturated DAG(32:0) was found to increase with BaP treatment and given the concurrent decrease in fatty acid C18:1, it is hypothesised that DAG(32:0) observed here originates from the breakdown of cellular phospholipids. Future work to confirm the origin of DAG species by mass spectrometry would be to employ isotopically labelled precursors for incorporation into fatty acids thereby providing a way of tracking cellular uptake and ultimate fate of the labelled biomolecules within the cell. This methodology was recently reported by project collaborators [218] in a study which suggests that a reduction in the lipogenic enzyme SCD1 is implicated with BaP treatment. This enzyme is responsible for producing the monounsaturated fatty acid C18:1 from C18:0 and a downregulation of it would be therefore be expected to decrease C18:1 lipid within the cell. Data presented in Chapter 6 correlates well with these findings. No further studies in the literature have probed the lipidome of AML cells with BaP treatment; furthermore, the author is not aware of any SIMS studies of both K562 and HL60 AML cell lines in any context.

For reasons previously discussed, these observed effects of drug treatment suggest anti-leukaemic activity of BaP through disruption of lipogenesis. Significant lipids tentatively identified here could be further investigated using tandem mass spectrometry to determine fragmentation pathways and therefore the building blocks from which the lipids are comprised [314]. Confirmation of lipid identifications through the analysis of standards would add a further degree of confidence to these findings.

Employing the imaging capabilities of ToF-SIMS in a continuation of this study would add a powerful dimension to this data set by providing spatial information of cellular lipids on a single cell level. Preliminary experiments have been undertaken for this and future work is planned.
7.5 Adopting a multi-disciplinary spectroscopic and mass spectrometric approach increased confidence in interpreting metabolomic and lipidomic changes with BaP treatment at the cellular level

It is widely acknowledged that not one analytical platform can provide global coverage of the cellular metabolome and lipidome. The scope of this project provided access to both polar and non-polar data from intact cells and cell extracts across a range of techniques, a scale which has not previously been reported in one single study. Interpretation of such large data sets can be challenging. Here, findings from one analytical platform that are corroborated by data obtained with a complementary or orthogonal platform, carry a further degree of confidence than if observed with a single technique alone. A significant finding is the apparent increase in lipid unsaturation with BaP treatment, a result which was identified by S-FTIR, ATR-FTIR and Raman microspectroscopy and further supported by data obtained from UHPLC-MS and ToF-SIMS. Whilst spectroscopic data was not able to provide information on specific lipid classes, the high specificity of mass spectrometry revealed a decrease in C18:1 monounsaturated species with an overall increase in many polyunsaturated species. Had this study been limited to spectroscopic data alone, one might have drawn the conclusion that monounsaturated species as well as polyunsaturated species were upregulated with BaP treatment. This exemplifies the power of adopting such a multidisciplinary approach.

7.6 Future Prospects

Future work to add further degrees of confidence in assignments (metabolite and lipid ID confirmations) and an extra dimension (ToF-SIMS cell imaging) to the outcomes of this project have been discussed. One limitation of the experimental design within this study results from the replicate samples being derived from one pooled flask of cells as detailed in Figure 3.15. Whilst replicates (n = 6) for BaP treated and control cells were cultured in individual flasks and subsequently extracted and analysed separately to each other, the project would benefit from the analysis of biological replicates derived from a different pooled source of cells at a different point in time (e.g. 3 replicates per experiment, with the experiment performed twice).
A direct comparison of data derived from all analytical platforms employed in this multi-dimensional project is highly desirable. Over-lying data from orthogonal techniques is not a trivial task and would be performed in collaboration with statisticians to assist with the manipulation of the various data matrices available. However, there is potential to reveal new information regarding the complementarity of spectroscopic and mass spectrometric platforms, especially since data derived for AML cell extracts was from one source.

There are also possibilities to mine the vast data sets for other markers of BaP treatment or to explore inter-cellular differences between cell lines.

An investigation into ROS species would be interesting. The overall increase in polyunsaturated lipid species observed here with BaP treatment implies that lipids post-BaP treatment would be more susceptible to lipid peroxidation as discussed in section 4.3.8. An increase in oxidised lipid species would therefore be expected with BaP and could be probed via observing smaller reactive breakdown products such as malondialdehyde and 4-hydroxalkenals, the presence of which are undesirable at high concentrations, since they can reportedly cause cellular damage [30]. Standards for these small molecules would be analysed by UHPLC-MS and ToF-SIMS methodologies employed here and resulting mass spectral signatures could then be used to mine existing data for presence of ROS species. Information on the relative quantitation between control and BaP treated cells could be derived.

A comprehensive treatment of the data for individual HL60 and K562 cell lines was beyond the scope of this project. However, it is likely that data derived here contains powerful information pertaining to markers of cell differentiation which could discriminate between immature HL60 cells and the more differentiated K562 cell line. Furthermore, there is the possibility to mine the data for cell-specific drug interactions which might offer a more specific insight into the mechanistic action of BaP at a cellular level.
Appendix I

Adaptation of confirmation of project collaboration email

To: Andrew Southam, Farhat Khanim and Chris Bunce @ University of Birmingham

Cc: Nick Lockyer and Roy Goodacre @ University of Manchester

From: Joanna Denbigh @ University of Manchester Date: 25th July 2013

Following a successful trip to Birmingham yesterday I just wanted to briefly summarise the aims of our collaborative work to all involved and highlight some action steps at this stage for moving forward.

Joanna Denbigh current PhD project title:
‘Lipidomic and Metabolomic Imaging of Biological Response Mechanisms in Tissues and Single Cells’

Collaborative Work:
Joanna Denbigh, under the supervision of Nick Lockyer and Roy Goodacre from the Manchester Institute of Biotechnology, University of Manchester in collaboration with Andy Southam, Farhat Khanim and Chris Bunce from the School of Biosciences, University of Birmingham.

Time-lines:
It is foreseen that work within this collaboration could continue for the duration of PhD.

Experimental:
Preliminary work with BaP treated and control cell extracts provided by Andy has been presented in the attached poster at Metabolomics 2013. There was much interest in this which was very pleasing.
Aim in moving forwards is to progress from analysis of extracts to analysis of whole cells to see if changes in lipids can be observed within the cell and at a later stage localised within cells.
The next presentation will be at SciX 2013 in Milwaukee in October 2013 at which Jo plans to present a poster with the cell extract data and if applicable some data from whole cells included.
Analysis will be using ToF-SIMS with the BioToF instrument and Raman spectroscopy with the Renishaw Invia Raman microscope with a 785 nm laser. This also has the potential for ‘live’ cell analysis. An ideal outcome would be to obtain cell images with lipid regions clearly visible.
Current work will focus on sample prep/quenching/fixing to ensure that cells are transported and analysed in most appropriate form.
Jo and Andy are very keen to publish any interesting outcomes of this work.
Appendix II

Mycoplasma Test

Mycoplasma testing was carried out on a routine basis for all cells in culture according to recommended guidelines [272]. Results presented here from 27/11/14 for K562 and HL60 cell lines (results 5 and 6 respectively) confirm the absence of mycoplasma infection.

### Mycoplasma test 27\textsuperscript{th} Nov. 2014

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6-ve</th>
<th>+ve</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8 is only positive result.

**Key** 1 = SHSY5Y (Prisc.) 2 = RBLSX38 (Sayers) 3 = K562 ADH (Just.) 4 = K562 All (Just.) 5 = K562 (Jo) 6 = HL60 (Jo) 7 = KGLA (Jo) 8 = HaCat (Kat) 9 = PANC1 (Kat) 10 = JURKAT (Steph.) 11 = LNCAP (Steph.) 12 = PC-3 (Steph.)
References


13. Lanekoff, I, Sjovall, P, and Ewing, AG, *Relative Quantification of Phospholipid Accumulation in the PC12 Cell Plasma Membrane Following*


279


120. Kotze, HL, Armitage, EG, Fletcher, JS, Henderson, A, Williams, KJ, Lockyer, NP, and Vickerman, JC, *ToF-SIMS as a tool for metabolic profiling small


