The Development and use of Infrared Micro-Spectroscopic Techniques to Monitor the Differentiation of Pluripotent Stem Cells Induced by Retinoic Acid and Synthetic Retinoid Analogues

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

2014

Graeme Clemens

SCHOOL OF CHEMICAL ENGINEERING AND ANALYTICAL SCIENCE
Contents
List of Abbreviations ............................................................................................................ 6
List of Figures ......................................................................................................................... 9
List of Tables .......................................................................................................................... 12
Abstract ................................................................................................................................. 13
Declaration ............................................................................................................................... 15
Copyright Statement ............................................................................................................. 16
Acknowledgement and Dedication ....................................................................................... 17
Introduction ............................................................................................................................. 18
Aims and Objectives ............................................................................................................... 19
Roadmap Description of Chapters ...................................................................................... 20
Publications Including Research Performed in this Thesis: ............................................... 24
Chapter 1: Introduction .......................................................................................................... 25
  1.1 Stem cells ....................................................................................................................... 25
      1.1.1 Adult stem cells ...................................................................................................... 25
      1.1.2 Embryonic stem cells (ES cells) ........................................................................... 27
      1.1.3 Embryonic carcinoma stem cells (EC cells) ......................................................... 30
      1.1.4 TERA2.cl.SP12 cell line .................................................................................... 31
  1.2 Retinoid induced differentiation .................................................................................... 34
      1.2.1 Retinoids .............................................................................................................. 34
  1.3 Infrared Spectroscopy .................................................................................................... 41
      1.3.1 Absorption Band Assignment ............................................................................. 43
      1.4.1 Nucleic Acids ...................................................................................................... 45
      1.4.2 Proteins .............................................................................................................. 46
      1.4.3 Fatty Acids and Lipids ....................................................................................... 48
      1.4.4 Carbohydrates ................................................................................................... 49
Chapter 2: Methodology and Principles/Pre-Processing .................................................. 51
  2.1 Beer Lambert Law for Quantitative Analysis ............................................................ 51
  2.2 Fourier Transform Infrared (FT-IR) Spectroscopy ...................................................... 52
2.2.1 Advantages of FT-IR ................................................................. 55
2.2.2 FT-IR Micro-Spectroscopy (FT-IRMS) ........................................ 56
2.3 Background Spectrum ................................................................. 58
2.4 Sources of IR Radiation ............................................................... 60
  2.4.1 Synchrotron Radiation ............................................................ 60
2.5 Detectors .................................................................................... 62
  2.5.1 Focal Plane Array (FPA) Detectors and Hyperspectral Images ........ 63
2.6 The Interaction of Electromagnetic Radiation with Matter and Resonant Mie scattering ........................................................................ 65
  2.6.1 Extended Multiplicative Signal Correction algorithm (EMSC) ........ 71
  2.6.2 Resonant Mie Scattering algorithm (RMieS-EMSC) ...................... 71
2.7 Pre-Processing Steps after RMieS-EMSC Correction ......................... 73
  2.7.1 Noise Reduction ........................................................................ 73
  2.7.2 Vector Normalisation ................................................................. 73
  2.7.3 Spectral Transformation to Derivatives ....................................... 73
2.8 Multivariate Spectral Analysis of Recorded IR Spectra ....................... 74
  2.8.1 Principal Component Analysis (PCA) ......................................... 75
  2.8.2 Supervised Multivariate Analysis .............................................. 80
2.9 Cross-Validation ........................................................................ 82
  2.9.1 Bootstrapping or K-Fold Cross-Validation ................................... 84
  2.9.2 Predicted Residual Error Sum of Squares (PRESS) ....................... 85

Chapter 3: Previous Studies using FT-IR Micro-Spectroscopy (FT-IRMS) to Investigate Stem Cell Differentiation ................................................................. 87

Chapter 4: ................................................................................................. 91

The Action of all-trans-Retinoic Acid (ATRA) and Synthetic Retinoid Analogues (EC19, EC23 and AH61) on Human Pluripotent Stem Cells; Differentiation Investigated using Single Cell IRMS ......................................................................... 91

  4.1 Abstract ...................................................................................... 91

  4.2 Cell Sample Preparations ............................................................. 92
  4.2.1 Protocol for the Cell Culture of TERA2.cl.SP12 Cells .................... 92
4.2.2 Cell Fixation ........................................................................................................94
4.2.2 Cytospinning ........................................................................................................96
4.2.3 FT-IRMS Instrument used and Experimental Setup .............................................96
4.3 Flow Cytometry Analysis .........................................................................................96
4.4 Data Pre-Processing ...............................................................................................98
4.4.1 Extracting Relevant Cell Information from IR Chemical Images .......................98
4.4.2 Data Pre-Processing and Chemometrics .............................................................105
4.5 Control Sample Spectra Reproducibility ...............................................................108
4.6 Comparison of Control (DMSO and Untreated TERA2.cl.SP12 Cells) and EC23, EC19, AH61 and ATRA Treated TERA2.cl.SP12 Cells at 7 Days ..........................108
4.7 Biochemical Differences between Control Cells; DMSO Treated TERA2.cl.SP12 and Untreated TERA2.cl.SP12 Cells .................................................................110
4.8 Biochemical Differences between Retinoid and DMSO Treated (control pluripotent stem cells) TERA2.cl.SP12 Cells at 7 Days ............................................................111
   4.8.1 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 7 Cell Samples .................................................................115
   4.8.2 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 5 Cell Samples .................................................................116
   4.8.3 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 3 Cell Samples .................................................................117
4.9 Comparing the Different Actions of the Retinoids on the TERA2.cl.SP12 Cells ....119
4.10 Monitoring the Spectroscopic Signals of Differentiating Cells through Time using IRMS ....................................................................................................................123
4.11 Conclusions ........................................................................................................125

Chapter 5 ....................................................................................................................127

The Action of Retinoids on Human Pluripotent Stem Cells at Early Time Points after Retinoid Addition and the Investigation of the RMieS-EMSC Correction algorithm......127

5.1 Abstract ..............................................................................................................127
5.2 Introduction and Aims ........................................................................................128
5.3 Cell Sample Preparations ....................................................................................128
5.4 Data Pre-Processing ............................................................................................129
5.4.1 Improvements Made to the Single Cell Finding Function ........................................129
5.4.2 Data Pre-Processing and Chemometrics ...............................................................132
5.5 Comparison of Control (DMSO treated TERA2.cl.SP12 Cells) and EC23, AH61 and ATRA Treated TERA2.cl.SP12 Cell Spectra ................................................................133
5.6 Discussion of the Results Found ..............................................................................138
5.7 Evaluation of the RMieS-EMSC Correction Algorithm ............................................139
5.8 Conclusions .............................................................................................................146
Chapter 6: ......................................................................................................................148
The Recording of IR Spectra from Single Cells Immersed in Aqueous Solution ..........148
6.1 Abstract ..................................................................................................................148
6.2 Introduction and Aims ............................................................................................149
6.3 Cell Sample Preparations ......................................................................................154
6.3.1 Formalin Fixed Cells ..........................................................................................154
6.3.2 Live A498 cells ..................................................................................................154
6.3.3 IRMS Investigation of Stem Cell Differentiation using Living TERA2.cl.SP12 Cells ......................................................................................................................154
6.4 Materials and Methods .........................................................................................155
6.4.1 Micro-fluidic Device Design ..............................................................................155
6.4.2 Micro-static Device ............................................................................................158
6.4.3 Synchrotron Fourier Transform Infrared Microspectroscopy (SR-FT-IRMS or SR-IRMS) Instrument and Experimental Setup .........................................................158
6.4.4 Attenuated Total Reflectance Fourier Transform Infrared Micro-spectroscopy (ATR-FT-IRMS or ATR-IRMS) Instrumentation .........................................................159
6.4.5 Benchtop Fourier Transform Infrared Micro-spectroscopy (FT-IRMS) Instrumentation and Experimental Setup ..............................................................................159
6.4.6 Aqueous Buffer Removal Function ....................................................................159
6.5 Evaluation of the Aqueous Buffer Removal Function ............................................163
6.6 Comparison of Hydrated Formalin Fixed DMSO and EC23 Treated TERA2.cl.SP12 Cell Spectra ...........................................................................................................171
6.8 Conclusion .............................................................................................................181
Chapter 7: Overall Conclusions from the Study and Future Work.................................................................183

7.1 Conclusions ........................................................................................................................................183

7.2 Future work .........................................................................................................................................185

Chapter 8: References ..............................................................................................................................187

Chapter 9: Supplementary Material for..................................................................................................197

9.1 Chapter 4 ............................................................................................................................................197

9.1.1 Chapter 4 – Day 7 Data Analysis .....................................................................................................198

9.1.2 Chapter 4 – Day 5 Data Analysis .....................................................................................................204

9.1.3 Chapter 4 – Day 3 Data Analysis .....................................................................................................205

9.1.4 Chapter 4 - Retinoid Treated Cell Spectra Comparisons .................................................................207

9.1.5 Chapter 4 - Monitoring the Spectroscopic Signals of Differentiating Cells through Time using IRMS ..........................................................................................................................208

9.1.6 Chapter 4 - Day 14 Retinoid Treated EC Cells ...............................................................................211

9.2 Chapter 5 ............................................................................................................................................213
List of Abbreviations

ATRA – All-trans-retinoic acid

APL - Acute Promyelocytic leukaemia

BMP - Bone morphogenetic protein

bFGF - Basic fibroblast growth factors

CaF₂ – Calcium Fluoride

CNS - Central nervous system

CRABP-I and II - Cellular retinoic acid binding proteins I and II

CRBP-I and II - Cellular retinol binding proteins I and II

DMSO – Dimethylsulfoxide

DMEM – Dulbecco’s modified eagle serum

DNA - Deoxyribonucleic acid

DTGS - Deuterated Triglycerine Sulfate

EB – Embroid bodies

EC cells – Embryonic carcinoma stem cells

EMSC - Extended Multiplicative Signal Correction algorithm

EMR - Electromagnetic radiation

ES cells – Embryonic stem cells

FACS - Fluorescent activated cell sorting

FCS – Foetal calf serum

FGF2 - Fibroblast growth factor-2

FIR – Far Infrared

FT – Fourier transform

FT-IR – Fourier Transform Infrared
FT-IRMS Fourier Transform Infrared Microspectroscopy

FPA – Focal plane array (detectors)

hECCs – Human embryonic carcinoma stem cells

hESCs – Human embryonic stem cells

HCL – Hydrogen chloride

IR – Infrared

ICM – Inner cell mass

IPS - induced pluripotent stem cells

IVF - *in vitro* fertilisation

LDA- Linear discriminant analysis

LIF - Leukaemia inhibitor factor

LPB - Ligand binding pocket

MACS - Magnetic-activated cell sorting

MCT - Mercury Cadmium Telluride (detector)

MEF – Murine (mouse) embryonic fibroblasts

mESCs - Murine (mouse) embryonic stem cells

mRNA – Messenger RNA (ribonucleic acid)

NA – Numerical aperture

NIR – Near Infrared

OPD - Optical path difference

PBS – Phosphate buffer solution

PC – Principal component

PC-LDA – Principal component linear discriminant analysis

PCs – Principal components

PCA – Principal component analysis
PCR - Polymerase chain reaction

qPCR - Real time polymerase chain reaction

RALDH - Retinal dehydrogenases

RARs - Retinoid acid receptors

RMieS – Resonant Mie scattering

RMieS-EMSC - Resonant Mie scattering - Extended Multiplicative Signal Correction algorithm

RNA –Ribonucleic acid

ROLDH - Retinol dehydrogenases

RXRs - Retinoid x receptors

SNR – Signal to noise ratio

S/N – Signal to noise

SR – Synchrotron Radiation

SR-IRMS – Synchrotron Radiation – Infrared Microspectroscopy

VAD - Vitamin A deficiency

ZPD – Zero path difference
List of Figures

1 Human Embryonic stem cells (hESCs) and their potential differentiation lineages (page 28)
2 The chemical structures of all-trans-retinoic acid (ATRA) and its natural isomers 9-cis-retinoic acid (9CRA) and 13-cis-retinoic acid (13CRA) (page 35)
3 The chemical structures of the synthetic retinoids AH61, EC19 and EC23 (page 39)
4 Immunocytochemical staining for the neuronal marker Tuj1 after 21 days of culture in the presence of 10 mM of ATRA, EC19 or EC23 (page 40)
5 Electromagnetic wave (page 42)
6 The atomic displacement of the HCl molecule after a quantised transition to a higher vibrational energy state (page 42)
7 Molecular vibrations of the carbon dioxide molecule (CO₂) (page 43)
8 Typical mid-infrared absorption spectrum of a biological sample including the molecular assignment of the main absorption bands (page 44)
9 DNA structure (page 46)
10 Covalent peptide bond linking amino acids together (page 47)
11 Both α-helix and β-sheet protein secondary molecular structures (page 48)
12 IR absorption spectrum of the main macromolecules found within a biological sample (page 50)
13 Monochromatic radiation directed through a sample (page 52)
14 Schematic of the Michelson interferometer (page 53)
15 Example of an interferogram (page 55)
16 Schematic of an FT-IR micro-spectrometer incorporating a single element detector (page 57)
17 Schematic of a Schwarzschild objective incorporated within the IR microscope (page 57)
18 Single cell targeted with apertures (page 58)
19 Vibration–rotation MIR absorption spectrum of gaseous small molecules found in the atmosphere (page 59)
20 Infrared signal at the detector through different aperture sizes using synchrotron light and light emitted from a Globar source (page 61)
21 Example of the many pixel element detectors incorporated into a FPA detector (page 64)
22 Total absorbance heat map image of TERA2.cl.SP12 cells distributed onto a CaF₂ IR substrate (page 65)
23 The interaction of EMR with matter (page 67)
24 An IR spectrum recorded from a single cell containing scattering contributions in the spectrum, the same spectrum corrected using RMieS-EMSC algorithm and the EMSC algorithm (page 70)
25 Processes involved in the RMieS-EMSC algorithm (page 72)
26 PCA breakdown (page 77)
27 The effect of mean centering the data and how PCA rotates the two dimensional axis of the mean centered data to find new linear boundaries known as principal components (page 78)
28 Demonstration of PCA using simulated data (page 79)
29 LDA demonstration (page 81)
Visual interpretation of how the data set is partitioned when using 5-fold cross-validation (page 84)

Cross-validation example; an LDA scores plot from the 10-fold cross-validation of 3 different classes of IR

Recorded spectra (page 85)

TERA2.cl.SP12 cells at low confluency and at high confluency (page 93)

TERA2.cl.SP12 cells at low confluency and phase micrographs of cells exposed to ATRA and EC23 (page 94)

Flow cytometry analysis results based on the SSEA-3 (pluripotent stem cell marker) and the A2B5 (neuronal cell marker) markers for cell culture samples left for 7 days (page 95)

Flowchart explaining the steps involved in the function used to extract single cell information from IR spectral images (page 99)

Visual of the step by step processes involved for the in-house made single cell extraction function (pages 100-104)

Data processing and Chemometric analysis flow chart (page 107)

(Day 5 sample data) comparing DMSO sample biological replicates (page 108)

LDA 1 vs. LDA 2 of spectra recorded from DMSO, untreated (control), ATRA, EC23 and AH61 treated cells (page 109)

IR spectra recorded from DMSO and Control cell samples compared using PCA (page 111)

PC-LDA scores histogram from the average bootstrap result of spectra recorded from DMSO and EC23 treated cells left in sample cultures for 7 days (page 112)

Comparing the mean spectra of all 3 day cell samples spectra (amide I spectral range 1600-1700 cm\(^{-1}\)) (page 119)

(Day 7 sample data) LDA 1 vs. LDA 2 of spectra recorded from ATRA, EC19, EC23 and AH61 treated cells (page 121)

PC-LDA scores plot (LDA 1 vs. LDA 2) of spectra recorded from AH61 treated cells samples at the 3 day, 5 day, 7 day and 14 day time points (page 123)

Graphical step by step of the improvements made to in-house developed single cell finding function (pages 129-130)

(chapter 5 data) PC-LDA scores plot (LDA 1 vs. LDA 2) of all ATRA, EC23, AH61 and DMSO treated (control) cell spectra recorded from the 9 day cell sample (page 133)

(chapter 5 data) PC-LDA scores plots (LDA 1 vs. LDA 2) of DMSO, AH61, ATRA and EC23 cell spectra recorded from the day 7, 5 and 3 samples (page 136)

All un-processed cell spectra recorded for the day 7 spectral comparisons (chapter 5 data analysis) (page 141)

Two cell spectrums exhibiting a greater level of RMieS and a spectrum exhibiting a low level of RMieS. Cell spectra corrected using the RMieS-EMSC algorithm and 8 iterations, and 200 iterations (page 142)

Cell spectra from Fig. 49 (a) corrected for scattering using the RMieS-EMSC algorithm with the first ten PCs selected and 3 iterations, and corrected using 30 pcs and 3 iterations (page 145)

Spectrum of buffer solution and a RAW spectrum of buffer solution plus single cell absorptions (page 150)

Raw cell plus buffer spectrum and aqueous buffer spectrum (page 152)

Schematic of the in–house made micro-fluidic device (page 157)
Raw cell plus buffer spectrum, Matrigel reference spectrum, aqueous buffer spectrum and pure cell spectrum produced (page 161)

The influence of the reference spectrum on the in-house developed buffer removal function (page 162)

Pure cell absorption spectra produced from one single cell plus buffer spectrum using all three processing methods (page 164)

Comparison of the outputted pure cell absorption spectra produced using the same raw cell plus buffer spectrum, but different reference spectra (page 165)

Comparison of mean cell spectra recorded in air on a CaF₂ slide and recorded in aqueous buffer in the micro-fluidic device (page 167)

Spectra recorded of formalin fixed TERA2.cl.SP12 cells, dry and in air, and hydrated in aqueous buffer using ATR spectroscopy (page 169)

Spectra recorded of formalin fixed TERA2.cl.SP12 cells, dry and in air, and hydrated in aqueous buffer using ATR spectroscopy. Spectra transformed to second derivatives (page 170)

PC-LDA comparisons of cell spectra recorded at the synchrotron facility in aqueous buffer, spectra recorded from the DMSO and EC23 formalin fixed cells from the 7 day sample cultures. Buffer contribution removed using processing method 3 (page 173)

(Live cell study) Average cell spectrum and aqueous buffer spectrum recorded in close vicinity on the benchtop FT-IR Micro-spectrometer. Absorption cell spectrum after the buffer contribution was removed using processing method 3 (clear over subtraction of both the amide I and II bands) and comparison of the mean second derivative spectra, ATRA and DMSO cell spectra (page 176)

(Live cell study) PCA scores plot and mean spectral comparison of spectra recorded from ATRA and DMSO treated living cells; cells treated for 7 days (page 179)

Supplementary Figures (chapter 9)

PCA of spectra recorded from all biological replicates of untreated and DMSO treated TERA2.cl.SP12 cell samples. PCA of 5 day samples PCA of 3 day samples (chapter 4) (page 197)

PC-LDA scores histogram from the average bootstrap result of spectra recorded from DMSO and AH61 treated cells left in sample cultures for 7 days (chapter 4) (pages 198-199)

PC-LDA scores histogram from the average bootstrap result of spectra recorded from DMSO and EC19 treated cells left in sample cultures for 7 days (chapter 4) (pages 200-201)

PC-LDA scores histogram from the average bootstrap result of spectra recorded from DMSO and ATRA treated cells left in sample cultures for 7 days (chapter 4) (pages 202-203)
PC-LDA data analysis of 5 day data; both EC23 and AH61 vs. DMSO treated cells (chapter 4) (page 204)

PC-LDA data analysis of 3 day data (chapter 4) (page 205-206)

PC-LDA and mean spectra comparison of retinoid treated cell spectra; day 7 data (chapter 4) (page 207)

Data analysis of retinoid treated cell spectra at different time points after retinoid addition (chapter 4) (pages 208-209)

Data analysis of DMSO treated cell spectra at different time points (chapter 4) (page 210)

Data analysis of cell spectra recorded from day 14 samples (chapter 4) (pages 211-212)

PC-LDA spectral loadings and mean spectra of all 5 day data (chapter 5) (page 213)

PC-LDA spectral loadings and mean spectra of all 3 day data (chapter 5) (page 214)

List of Tables

1 Different levels of cellular potency (Page 26)

2 Spectral differences between retinoid treated cell spectra and DMSO control cell spectra at 7, 5 and 3 days (page 114)

3 Assigning spectral bands to molecular vibrations (page 118)

4 Spectral differences when comparing spectra recorded from retinoid treated cells and pluripotent stem cells using PC-LDA; 5, 7 and 9 day cell sample spectra (page 135)

5 Spectral differences between formalin fixed hydrated EC23 treated cells and control (DMSO treated) stem cells at 7 days. Data recorded from cells in aqueous buffer, with buffer removed using processing method 3 (page 177)

Supplementary Tables (chapter 9)

1 Spectral Differences between ATRA, AH61 and EC23 Spectra Recorded from Retinoid 14 Day samples and 7 Day Control Samples (page 216)
Abstract

- University of Manchester
- Graeme Neal Clemens
- Doctor of Philosophy in the Faculty of Engineering and Physical Sciences
- “The Development and use of Infrared Micro-Spectroscopic Techniques to Monitor the Differentiation of Pluripotent Stem Cells Induced by Retinoic Acid and Synthetic Retinoid Analogues”
- 31st October 2014

In this study, Fourier Transform Infrared Micro-Spectroscopy (FT-IRMS, or IRMS), coupled with multivariate analysis is shown to be an effective tool to further investigate the differentiation of human pluripotent stem cells and monitor the alternative affects different retinoid compounds have on the induction of differentiation. Stem cells and differentiated phenotypes being distinguished based on the spectral fingerprint recorded and not through the addition of fluorescent or magnetic biomarkers.

IRMS detected differences between cell populations as early as 3 days of retinoid compound treatment. Populations of stem cells treated with different retinoid compounds could easily be distinguished from one another during the early stages of cell differentiation, demonstrating the potential of that IRMS technology as a sensitive screening technique to monitor the status of the stem cell phenotype and progression of differentiation along alternative pathways in response to different compounds.

The study also investigates some of the problems with intrinsic resonant Mie scattering (RMieS), produced when recording IR spectra from single cells in air. Some single cell spectra recorded containing RMieS were not corrected properly when using the RMieS-EMSC scattering correction and typical correction parameter options; correction of such spectra using the RMieS-EMSC algorithm resulted in the production of cell spectra with odd absorption band structures after correction. This study demonstrates that alternative
correction parameter options for the RMieS-EMSC correction can successfully overcome this problem.

Early experimental results show that IRMS is a sensitive technique capable of capturing discrete molecular changes of pluripotent stem cells as they are induced to differentiate, without the need for expensive and potentially damaging exogenous labels. Studies in the literature predominately use IRMS to investigate chemically fixed cells, however, this is a form of cell death and once cells are fixed they become un-viable. Critically, cell phenotype should be understood from living cells with the biochemistry being maintained after information is recorded so that single cells can retain their biological function. IR spectroscopy has the potential to record spectra from living cells when coupled with the micro-devices to provide a suitable water barrier for IR spectroscopic investigations. This study demonstrates that cell spectra can be recorded from cells in an aqueous environment, with a new processing method being described to accurately remove the water buffer contributions contained within raw single cell spectra recorded in an aqueous environment. Therefore, highlighting the potential of IRMS for the screening of living cells, potentially identifying stem cells from other cell phenotypes based on their biochemical make-up and without cellular destruction.
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.

G. N. Clemens
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, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

(iv) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgement and Dedication

First of all i would like to thank Manchester University for funding two years of my PhD. It has been an honour and a pleasure to work with such a prestigious University.

In my time here at Manchester University I have been helped by many people who have helped me along on my journey and taught me a great deal. In particular I would like to thank my supervisor Professor Peter Gardner who gave me the opportunity to do a PhD, your advice and knowledge has shaped me for the future, and I am forever grateful. I would also like to thank Dr. Kevin Flower, and Professor Andy Whiting and Professor Stefan Przyborksi from Durham University for their immense help and support throughout my studies.

Throughout my three years at Manchester University I have made some great friends who have helped me along the way. I would like to acknowledge Dr. Caryn Hughes, Dr. Konrad Dorling, Dr. Francis Ball, Adriana Ortega-Zeifert and Dr. Alex Henderson for not only their friendship, but their help and support during my PhD. In particular I would like to make a special acknowledgement for Dr. Caryn Hughes who has been an inspiration.

Finally, I would like to thank all my friends and family for their continued support and advice during the PhD days. I do have to make a special thank you though to my lovely wife Katie, I apologise for the increased level of grumpiness over the time I have spent doing the PhD.

I would like to dedicate this thesis to my Grandma, Margaret Longmire, who sadly passed away during the writing of the thesis, and my mother, Karen Kent, who is battling breast cancer. You both gave me the strength to get this done.
Introduction

The use of stem cells in biomedical applications is currently an area of intense interest given the huge potential stem cells offer for producing human cells and tissue for cell based regenerative therapies, such as repairing or replacing damaged cells and tissue. Currently, tissues and organs are replaced through donors but the need far outweighs the supply \(^1\). Therefore, there is intense interest in the use of stem cells to replace cells and tissues. However, a number of challenges still remain in stem cell biology and regenerative medicine, which includes the purity of cell populations being ensured. For example, pluripotent stem cells have the potential for tumour development if transplanted into a patient during clinical regenerative therapy \(^1, 2, 3\). Thus, ensuring that a cell population contains no pluripotent stem cells is extremely important. Currently, methods used to monitor the phenotype of stem cells and their differentiated progeny include fluorescent activated cell sorting (FACS) or fluorescent imaging. Both methods require the addition of exogenous fluorescent dyes or markers, which then attach to specific molecules on the surface of cell. However, fluorescent markers have a limited lifetime, after which they undergo irreversible photobleaching and can no longer be probed. The major concern though is the potential to alter the function (attachment can modify a cells surface chemistry) or even damage live cells, compromising their clinical use. Other issues include a limited number of known biomarkers and the lack of biomarkers specificity \(^4\). In the case of cardiomyocytes, gastrointestinal and corneal stem cells, no biomarkers have actually been found which can identify these cell phenotypes \(^2\). Immunocytochemical staining and flow cytometry also provide no direct biochemical information from the biological samples thus, no insight into the biochemical processes involved with differentiation. Other single cell analysis methods, such as mass spectrometry and polymerase chain reaction (PCR) have associated problems, including the destruction of cell samples as a result of interrogation. Therefore, there is a real need an analytical method which can characterise cell phenotype rapidly, without the addition of potentially harmful biomarkers to cell surfaces, inexpensively, requiring minimum sample preparation and more importantly, without the destruction of the cell. Infrared radiation causes vibrations of bonds of molecules within the sample that absorbs it. The wavelength of the IR radiation absorbed depends on the atoms involved and the strength of any intermolecular interactions. Therefore, each molecule can have a different spectrum; in essence, the IR spectrum is a fingerprint of the sample thus, IRMS has the potential to identify and monitor cell phenotype based on the vibrational signal recorded from a cell. Importantly, this can be achieved without the use of reagents, long sample preparation steps and without the destruction of the cells being interrogated.
The natural retinoid, All-trans-retinoic acid (ATRA), is an important mediator of cellular differentiation and apoptosis, regulating many biological processes, particularly during the development of the central nervous system (CNS), where ATRA is the principal active retinoid found during early embryonic CNS development\(^5\). Because of the differentiation capabilities of ATRA the compound has been successfully used to direct the differentiation of human stem cells; ATRA being known to induce the differentiation of human pluripotent stem cells to produce neural cell phenotypes\(^5, 6, 7\). Synthetic retinoids (EC23, EC19, AH61) have also been synthesised at Durham University, with EC23\(^5, 6\) and AH61\(^8\) being shown to have differentiation capabilities similar to ATRA but with enhanced compound stability, a more desirable function for producing homogeneous cultures of specialised cell types.

**Aims and Objectives**

Firstly, single cell spectra are recorded in this study using both the high brightness achieved from synchrotron radiation (SR-IRMS) and through chemical images recorded using focal plane array (FPA) detectors. The study describes some of the problems with extracting relevant cellular information from chemical images and describes the development of a new processing method for extracting single cell spectra from such images.

Previous studies have been successful in identifying different cell types based on the biochemical spectral fingerprint recorded and multivariate classification algorithms\(^9, 10, 11\). This study evaluates whether cell spectra recording using IRMS, coupled with multivariate algorithms, can be used to identify embryonic carcinoma (EC) TERA2.cl.SP12 stem cells from their differentiated derivatives, produced through adding retinoids to cultures. Analytical methods used to distinguish cell phenotype on this embryonic differentiation model include immunofluorescence, flow cytometry and Matrix-assisted laser desorption/ionisation - Time of flight Mass spectrometry (MALDI-TOF MS); cellular differentiation not being seen earlier than 7 days of retinoid addition to cultures using these methods. Therefore, the aim of the study is to see if IRMS coupled with multivariate algorithms is comparable to previous analytical methods, distinguishing EC stem cells from their differentiated phenotypes at 7 days, and to investigate the biochemical responses leading to exit from the normal cell cycle into differentiation. The study also investigates...
whether the signals of cellular differentiation can be identified at earlier time periods after retinoid introduction to cell cultures, *in vitro*.

Maltman *et al.* showed significant protein profile differences from control (undifferentiated TERA2.cl.SP12 cells), EC19 and EC23/ATRA treated EC cells, and subtle protein differences between ATRA and EC23 treated EC cells, which go on to produce neuronal functioning cells through differentiation. With IR spectroscopy being sensitive to protein structure, the study investigates whether ATRA and EC23 treated EC cells can be discriminated, even though cells treated with such retinoids produce neuronal phenotypes.

Finally, through the use of two micro-devices (home-made microfluidic device and a commercial micro-static device), which provide appropriate optical water path-lengths, and the development of a new buffer removal processing method, this study shows that recording IR spectra from living cells is possible. The new buffer removal function described in the study is assessed against other processing methods typically used in IRMS living cell studies.

**Roadmap Description of Chapters**

**Chapter 1: Introduction**

**Stem Cells:** A short overview of stem cells is given. The chapter starts by explaining the unique properties of stem cells and why they are considered to have great potential for tissue engineering, regenerative medicine and the development of new drugs. A brief description of both adult and embryonic stem cells follows, highlighting the potency differences between the two.

In this study, the TERA2.cl.SP12 embryonic carcinoma stem cell line will be used as a model of embryonic cellular differentiation. Embryonic carcinoma stem cells (EC cells) are a type of embryonic stem cell but they represent a malignant caricature to embryonic stem cells found the inner cell mass during the blastocyst stage of embryogenesis (ES cells). Chapter 1 will describe the different properties between EC and ES cells and why
the use of EC cells for experiment can be a more favourable embryonic cellular model for
differentiation studies, when compared to the use of ES cells. A brief description of the
TERA2.cl.SP12 cell line follows, explaining its previous use as a cell line to model early
embryonic neuronal differentiation, being induced to differentiate to form neuronal cells
after retinoid acid addition to cell cultures.

Undifferentiated pluripotent stem cells can cause tumour formation through teratomas
when transplanted into human hosts. Therefore, if stem cells are ever going to fulfil their
potential for regenerative therapies, a robust, reliable and accurate analytical method is
needed to distinguish cell phenotype before transplantation. The chapter will discuss
current analytical methods used to distinguish stem cells and their differentiated progeny
and their inadequacies for cell sorting and identification. Ideal methods for screening
cellular phenotype will have to be rapid, non-destructive, and without the need for
potentially harmful/non-specific biomarkers being added to cells.

**Retinoic Acid and Retinoids:** Retinoic acid (all-trans-retinoic acid (ATRA)) is the major
natural metabolite of vitamin A, is the principle active retinoid found during early
embryogenesis, a powerful regulator of neuronal differentiation, adult brain maintenance
and functioning and has been used to differentiate stem cells to produce specific
specialised cell types, *in vitro*. Chapter 1 will briefly describe the intrinsic biological
properties of ATRA and how it has been successfully used to differentiate TERA2.cl.SP12,
*in vitro*, to produce neuronal cell types, representing a good cellular model of neuronal
embryonic development. However, ATRA also suffers from isomerisation and oxidation,
which may lead to heterogeneous cell populations produced through cellular
differentiation. The chapter will explain why synthetic retinoids based on the chemical
structure of ATRA can be used as alternative differentiation agents.

Infrared Spectroscopy: A short overview of infrared (IR) radiation and IR spectroscopy
will follow, briefly describing the background and scientific principles behind IR
spectroscopy and why the properties of the analytical technique are favourable for the non-
destructive identification of single cells. IR spectroscopy being able to record direct
information from a sample, with samples being identified based on their IR signal
recorded, the signal being a fingerprint of the chemical structure of the sample. To finish, the molecular vibrations contributing to the main absorption present in a mid-infrared (MIR) absorption spectrum of a typical biological sample are assigned, followed by a brief description of the main macromolecules which makeup a biological cell, and the functional group vibrations which can be tentatively assigned to each main macromolecule.

Chapter 2: Methodology Principles

This chapter will start by describing the principles and instrumentation involved in Fourier Transform IR (FT-IR) spectroscopy, followed by the advantages of coupling FT-IR spectrometer with an optical microscope (FT-IRMS). The chapter will then follow on by providing detail on the inherent problem of IR light scattering when wavelengths of incident IR radiation are similar in size, or substantially less than the biological sample or components which make up the sample. This is then followed by an explanation of the principles behind the correction algorithm used in this study to correct for resonant Mie scattering (RMieS), and descriptions of the data processing and multivariate analysis methods used in this study.

Chapter 3: Sample Preparation and Instrumental Setup

Chapter 3 will describe previous studies in the literature which have investigated stem cell differentiation using IR spectroscopy techniques. The chapter will provide an overview of their studies and their findings.

Chapter 4: The Action of all-trans-Retinoic Acid (ATRA) and Synthetic Retinoid Analogues (EC19 and EC23) on Human Pluripotent Stem Cells Differentiation Investigated using Single Cell IRMS

This section describes the experiment where IRMS is used to record single cell spectra from EC TERA2.cl.SP12 stem cells and retinoid (ATRA, EC23, EC19 and AH61) treated TERA2.cl.SP12 cells; formalin fixed cells. Due to previous studies showing that the different retinoid compounds have different biochemical actions on the TERA2.cl.SP12 cell line, multivariate analysis is then used to see whether the spectra recorded from the cells undergoing different retinoid treatments could be distinguished based on the different biochemical structure of the cells. In this chapter, IR spectra are recorded from cells after
3, 5, 7 and 14 days after retinoid treatment. As well as describing this experiment and its findings, there is also a description of a new processing method for extracting relevant single cell information from recorded chemical images. This processing method was used in this experiment to extract the single cell spectra information from the chemical images recorded from the single cell samples.

Cell sample preparation was kindly provided by Andrew Henderson at Durham University. After formalin fixation, cell samples were then delivered to Manchester University where they were then put on to IR substrates and spectra were recorded using a Varian 670 FT-IR spectrometer, interfaced with a Varian-620 imaging infrared microscope. The microscope was equipped with a 128 × 128 liquid nitrogen cooled MCT focal plane array detector. Cell culture methodology and cell sample preparation are also described in this chapter, as well as flow cytometry methodology and results.

**Chapter 5: The Action of Retinoids on Human Pluripotent Stem Cells at Early Time Points after Retinoid Addition and the Investigation of the RMieS-EMSC Correction algorithm**

Chapter 4 shows that IRMS coupled with multivariate analysis can distinguish differentiating/differentiated retinoid treated TERA2.cl.SP12 cell spectra from control TERA2.cl.SP12 cell spectra. Therefore, this chapter investigates whether the biochemical change brought about by retinoid introduction to cell samples can be seen earlier than 3 days using IR spectroscopy, and if by providing a greater number of sample time points after retinoid introduction, can the spectral changes be mapped through time helping one to understand the spectral biomarkers of differentiation from this cell line? The chapter also further investigates the RMieS-EMSC correction algorithm.

Cell sample preparation was kindly provided by B. A. Murray at Durham University. After formalin fixation, cell samples were then delivered to Manchester University where cell samples were put on to IR substrates and spectra were recorded using a Varian 670 FT-IR spectrometer, interfaced with a Varian-620 imaging infrared microscope. The microscope was equipped with a 128 × 128 liquid nitrogen cooled MCT focal plane array detector.
Chapter 6: The Recording of IR Spectra from Single Cells Immersed in Aqueous Solution

If IRMS is going to live up to its potential as an analytical method for screening cell phenotypes, it must be able to record cell spectra from living cells in an environment close to in vivo conditions of the cells. This chapter investigates the practical problems with recording IR spectra from cells in an aqueous environment. The section will also propose a novel method for removing the water contributions from a raw cell plus buffer spectrum to produce pure cell spectra against typically used processing methods. This chapter will use both formalin fixed and live cells for experimental investigation.

The formalin fixed TERA2.cl.SP12 cell samples used in Chapter 4 were used for the experimental investigation of hydrated cells using an in house developed microfluidic device. Spectra were recorded using the infrared beamline (MIRIAM) at the Diamond Light source. For the TERA2.cl.SP12 live cell investigation, cells were kindly provided by Prof. Stefan Przyborski at Durham University. Cell culture was performed by myself under the supervision of Dr. Caryn Hughes at Manchester University, and the recording of cell spectra from living cells was performed on the Varian benchtop FT-IRMS.

Chapter 7: Conclusion and Future

A review of the findings of the study is given followed by future proposals for using IR spectroscopy as a tool to probe living cells for cell type characterisation.

Publications Including Research Performed in this Thesis:


Chapter 1: Introduction

1.1 Stem cells

Stem cells have a huge amount of potential in both tissue engineering and regenerative medicine. This is due to the fact that when needed, they can become specialised through differentiation, creating many different cells needed for life. As well as tissue engineering and regenerative medicine, stem cells also have a large volume of potential for the research and development of new drugs and to model and understand in vivo cellular systems.\textsuperscript{12}

Stem cells have some unique properties that make them different from other cells found within living organisms. Firstly, under certain lab conditions stem cells are capable of replicating themselves many times over, and for long periods, thus, potentially producing millions of stem cells, \textit{in vitro} (self-renewal)\textsuperscript{1}. Secondly, while being a stem cell they have no capability of performing specialised functions, such as carrying oxygen (blood cell) or transporting signals around the body (nerve cells). Thirdly, stem cells have the unique potential to create specialised cells through differentiation. Usually, differentiating cells undergo several phases of differentiation, becoming more specialised through each phase.\textsuperscript{1} Specialised cells can be produced through asymmetric cell division, a stem cell dividing to produce a daughter stem cell (genetically identical to the parent) and a one non-stem cell, which differentiates to produce specialised cells. This asymmetric cellular model allows the stem cell pool to be maintained, while producing cells that will differentiate to produce functional cell types for replacement. However, the asymmetric cellular division model is not capable of expanding the number of stem cells, a disadvantage when there is a need to rapidly replenish the number of stem cells after injury or disease. Therefore, many stem cells are also capable of symmetric cellular differentiation, producing either two identical stem cells, or two identical cells capable of differentiation.\textsuperscript{1,13,14}

1.1.1 Adult stem cells

Typically, stem cells can be split into two categories, embryonic and adult stem cells. Adult stem cells are needed throughout a human’s body in order to sustain the upkeep of cells and tissues throughout a life time. Adult stem cells are thought to be mainly multipotent in nature, which means that they are believed to only have the capacity to produce cells types related to the type of tissue that they come from (Table 1). Therefore,
the role of adult stem cells in the body is primarily repair and maintenance of the tissue or organ where they are located. They are thought to be undifferentiated cells, which live in niches among differentiated specialised cells in a specific tissue or organ. While in their niches adult stem cells remain quiescent (non-proliferating) until needed. When triggered, either through disease or maintenance, they can then renew themselves (divide creating genetically identical daughter cells), and also differentiate to produce specialised cells needed for the repair/maintenance of the tissue from where they are found. For example, we as humans lose skin epithelium daily, therefore, a high turnover of skin epithelial cells is needed for the longevity of human life, and adult stem cells provide this. Another example would be haematopoietic adult stem cells, which are found in bone marrow. These adult stem cells are solely responsible for the maintenance and production of all types of blood cells needed.

Table. 1 Different levels of cellular potency.

<table>
<thead>
<tr>
<th>Cell potency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Totipotent cells</td>
<td>The potential to differentiate to form any cell found in the human body. Zygote cell, formed through the fertilisation of a female egg.</td>
</tr>
<tr>
<td>(Greatest potency)</td>
<td></td>
</tr>
<tr>
<td>Pluripotent</td>
<td>Like totipotent cells, pluripotent cells can differentiate to produce any cell found in the human body. However, they cannot create a full human organism</td>
</tr>
<tr>
<td>Multipotent</td>
<td>The capacity to produce cells types from the tissue they originate from. Adult stem cells found in blood, brain, liver, intestine, skin etc.</td>
</tr>
<tr>
<td>Nullipotent</td>
<td>Incapable of differentiation</td>
</tr>
<tr>
<td>(Least potency)</td>
<td></td>
</tr>
</tbody>
</table>
Other than bone marrow and the skin, adult stem cells have been found in a number of tissues and organs including the brain, heart, peripheral blood, blood vessels, skeletal blood, teeth, gut, liver, ovarian epithelium and testis. Over the last 60 years there has been a real excitement about adult stem cells, with adult stem cells being found in tissues previously thought not to exist. There have also been incredible clinical success stories involving adult stem cells. None more famous than the treatment of Leukaemia through the transplant of haematopoietic stem cells from healthy matched donors (sister, brother etc.) to replace the cancerous haematopoietic cells, which over produce white blood cells 1. However, not all tissue types have their own stem cell niche.

1.1.2 Embryonic stem cells (ES cells)

Embryonic stem cells (ES) are seen as the “Holy Grail cell” due to the potential they have in creating any cell needed in the human body through differentiation (Table 1). ES cells therefore have unprecedented potential for tissue donor shortage issues and research into the embryogenesis. The long term self-renewal of ES cells is said be due to high levels of telomerase activity. Telomerase being a ribonucleoprotein, which adds telomere repeats to chromosome ends to maintain the telomere length. Telomerase expression can be correlated with immortality in human cell lines. Human adult cells show shortened telomeres with age and eventually enter senescence after a finite proliferative life cycle 15.

ES cells are found in the inner cell mass (ICM) during the blastocyst stage of embryogenesis. They are pluripotent in nature (Table 1) as they can differentiate to form any cell found in the human body, differentiating to produce cells from all three germ layers (ectoderm, mesoderm and endoderm) 16. Potential ES cells can be tested for pluripotency through the reconstruction of embryos and the generation of chimeric mice (gold standard for murine ES cells), or the demonstration that the cells can differentiate down the three germ layers through teratomas formation when implanted in immune deficient mice (gold standard for human embryonic stem cells (hESCs)), in vivo 17. A teratoma represents a tumour which contains specialised derivatives from all three germ layers, formed from the spontaneous differentiation of pluripotent stem cells 18.

ES cells can be removed from the ICM during embryogenesis and cultured in vitro (Fig. 1). If not removed from the ICM, embryogenesis will continue to produce the embryo
proper. Typically hESCs are produced from embryos created for *in vitro* fertilisation (IVF). Embryos which are not implanted can be used as a source for hESCs as they would normally be thrown away. This method is seen as an ethically superior way to source hESCs, embryos being donated with approval of the parents.

![Diagram of human embryonic stem cells (hESCs) and their potential differentiation lineages.](image)

**Fig. 1** Human Embryonic stem cells (hESCs) and their potential differentiation lineages.

In order for ES cells to keep their stemness (pluripotency and long term self-renewal), certain culture conditions are needed in order to stop ES cells from spontaneously differentiating. Typically ES cells need feeder layers which help to maintain their pluripotency, but also allow them to grow (replicate themselves many times) in culture. Murine ES cells (mESCs) have been successfully cultured using mitotically inactivated murine embryonic fibroblasts (MEF) and feeder layers with exogenously Leukaemia
inhibitor factor (LIF) added to successfully prevent differentiation. Human ES cells can be maintained in a similar manner with feeder layers/serum, or if cultured without, basic fibroblast growth factors are needed (bFGF)\textsuperscript{16}. However, they do not respond to exogenously added LIF\textsuperscript{19}.

Since the first mES lines were established, strides have been made in mESCs research over the last 20+ years. In particular, mESCs can be used to reconstitute early murine embryos. This has then led to genome manipulation technology, which has produced hundreds of “knock in” or “knock out” transgenic animals. This research has resulted in a better understanding of gene expression and regulation \textit{in vivo}. This type of research obviously cannot be performed on human systems because of the extreme ethical opposition\textsuperscript{17}. However, there is only so much one can learn about human models from mESCs. The two cell types (mESCs and hESCs cells) not only differ in their physical properties, but also from the antigens which they express. Murine ES cells express the SSEA1 embryonic antigen but not SSEA-3 and 4, whereas hESCs express the opposite. This has ultimately led to a greater amount of research into hESCs with more cell lines being found\textsuperscript{17}.

For both mESCs and hESCs, their differentiated progeny is a product of the surroundings of the ES cells. As already mentioned, pluripotency is maintained when certain conditions are maintained in culture. If these conditions are lost, for example, the removal of hESCs from feeder layers, spontaneous differentiation will occur. Spontaneous differentiation will result in a heterogeneous spread of somatic tissue representative of the three germ cell lineages (resembling early post transplantation embryos)\textsuperscript{16}. Also, if hESCs are removed from their feeder layers and transferred to a suspension culture, hESCs produce embryoid bodies (EB). These EB’s are 3D spherical structures in a suspension culture with cells representing derivatives from all three germ layers. However, although spontaneous differentiation is a satisfactory method to test the cells for pluripotency, it is not a sufficient method for the production of certain cell types. As a result, many studies have researched methods for directed differentiation of hESCs to produce high yields of specific phenotypes\textsuperscript{17}. For example, cells from mesodermal lineages have been created through using irradiated mouse bone marrow stromal cells (S17 cells) to differentiate hESCs to produce a variety of hematopoietic cells (produce blood cells). This could mean that bone marrow stromal cells are necessary for hematopoietic differentiation\textsuperscript{20}. Cells from ectodermic lineages, such as neuron progenitor cells have been created from the differentiation of hESCs through culturing in a basel medium supplemented by the fibroblast growth factor-2 (FGF2), in the presence of the bone morphogenetic protein
Neuron progenitor cells can then differentiate to form neurons and astrocytes. Retinoic Acid is also a good inducer of neuronal differentiation of ESCs and has been used in a study by Barharvand et al. to differentiate hESCs to produce neuronal cells representative of early neurogenesis, in vivo. The differentiation of hESCs to produce cells from endodermal lineages has proven to be more difficult though. However, one study has shown that adherent hESCs can be differentiated to form cells from an endodermal lineage through treatment with dimethylsulfoxide (DMSO) and then sodium butyrate.

1.1.3 Embryonic carcinoma stem cells (EC cells)

Another type of embryonic stem cell is the embryonic carcinoma stem cells (EC cells). EC cells are found in malignant teratomas known as malignant teratocarcinomas. Teratocarcinomas are tumours composed of various types of tissue foreign to their site of origin, but with undifferentiated malignant stem cells contained within the tumour. These EC cells found in teratocarcinomas have unlimited proliferation (long term self-renewal), pluripotent in nature and can metastasize (produce tumours elsewhere). EC stem cells are considered to be the stem cells which create the malignant carcinomas thus, providing the concept of cancer stem cells. The malignancy of EC cells does not come from the differentiated products they produce as the majority of cells produced are non-malignant. Their malignancy comes from their ability to solely produce a whole teratocarcinoma.

This was demonstrated in 1964 by Kleinsmith and Peirce who produced a study to test whether EC cells were actually the stem cells producing teratocarcinomas by injecting a single EC cell into a new mouse host. The transplanted single cell had the capacity to produce a whole teratocarcinoma in its new host, thus confirming the cancer stem cell theory. After the discovery of EC cells, markers were found for the EC cells (e.g. glycolipid antigens SSEA-3 and 4) enabling the identification of pluripotent cells. Experimental conditions were also discovered which enabled cells to be cultured in vitro.

These factors, coupled with a connection between murine EC stem cells and murine cells found in the ICM of murine embryos, led to both mESCs and hESCs being derived in 1981 and 1998-2000 respectively.

When comparing ES cells to EC cells, EC cells have a much reduced pluripotency and also an abnormal karyotype (number of chromosomes in the nucleus of the cells). However, with knowledge into the control and manipulation of hESCs being in its infancy, the reduced pluripotency of EC cells can be seen as a favourable characteristic. For example, ES stem cells represent a greater risk of spontaneous differentiation due to slight changes.
in culture conditions. This is not favourable when performing experiments looking into the differentiation of ES cells. Ideally, ES cells should remain pluripotent until induced to differentiate down a certain lineage, spontaneous differentiation will result in spurious results. Because of this reduced pluripotency, human embryonic carcinoma stem cells (hECCs) are often much easier to look after in culture with no feeder layers being required for most cell lines. This is not the case for hESCs however. Therefore, even though hECCs are said to be a caricature of normal hESCs, have a reduced pluripotency and are genetically abnormal, some cell lines have a good capacity to differentiate down certain germ line lineages. These cell lines represent good models of early embryonic differentiation, providing a simpler/more robust working environment when compared to hESCs. As a result, hECCs will be used in this study to investigate whether infrared spectroscopy is a capable tool to characterise stem cell from differentiated progeny.

1.1.4 TERA2.cl.SP12 cell line
One such hECC line is the TERA2 cell line derived from a teratocarcinoma (germ cell tumour). The TERA2 is one of the oldest hECC lines established. It was established from a lung metastasis of a testicular teratocarcinoma in 1975. The TERA2 showed great propensity for differentiation down an ectodermal route (neuronal differentiation) while displaying pluripotency through creating a teratocarcinoma in a nude mouse host. A derivative of the TERA2 cell line found through using SSEA-4 immuno-reactive magnetic particles is the TERA2.cl.SP12 cell line. The TERA2.cl.SP12 cell line showed enhanced neuronal differentiation when compared with the TERA2 cell line, forming both neurons and glia when induced by retinoic acid. The TERA2.cl.SP12 cell line therefore represents a good model of early embryonic neuronal differentiation and will be the cell line used in this study. It can be characterised using both immunofluorescence and flow cytometry to detect the cell surface markers SSEA-3, SSEA-4 and TRA-160. These cell surface antigens are also characteristic of undifferentiated pluripotent hESCs. The down regulation of such markers signals cellular differentiation. However, there are still many hurdles involved in stem cell research which need to be overcome before stem cell therapy can become commonplace in clinical therapy. One of which is the problem with immune rejection of transplanted tissue. For adult stem cells, there has been success with differentiated progeny being transplanted from suitable donors, such as brothers and sisters who have inherited the same genetics. Foreign material with dissimilar genetics will activate an immune response, rejecting the foreign tissue and summoning a “seek and destroy” mission. To be able to produce new tissue derived from embryonic stem cells which has the same genetic information will be extremely difficult. Therefore, the
transplantation of any tissue produced using hESCs has to be accompanied by use of immunosuppressive drugs in order for it to be successful. Even with immunosuppressive drugs, the success rate may be low with possibilities of adverse side effects from the drugs administered\(^\text{27}\). This is not a particularly ideal situation to promote the cause of hESCs for regenerative medicine. However, the generation of a new type of pluripotent cell, known as the induced pluripotent (IPS) stem cell, may hold the answer to the problem of immune system rejection. IPS cells are somatic cells that have been reprogrammed to have pluripotency. This is achieved through re-introducing certain transcription factor genes, such as OCT-4, to the somatic cells can alter their function giving them pluripotency. One such study by Kim \textit{et al.} showed that the introduction of OCT-4 gene to neural adult stem cells was enough to regain pluripotency\(^\text{28-29}\). This would potentially mean that cells representing embryonic stem cells could be created from the patients thus, solving the problem of immune system rejection.

Another insufficiency in stem cell research is the identification of stem cells and their differentiated derivatives in cell populations. According to Rippon \textit{et al.}\(^\text{17}\) in 2003 there had been no method for the directed differentiation of hESCs which had achieved a 100% yield of pure differentiated specialised progeny; the problem being that undifferentiated (pluripotent) stem cells have the potential to cause tumour formation, known as teratomas, when transplanted into a human host\(^\text{17}\). It is therefore extremely important that sufficient quality control methods are in place to characterise undifferentiated cells and purify differentiated progeny before implantation. Currently, the quality control methods used for characterisation and cell sorting of stem cells and their differentiated derivatives include immunocytochemistry, fluorescence-activated cell sorting (FACS) and magnetic-activated cell sorting (MACS). These methods rely on cells having a surface marker that can be recognised by a fluorescent/magnetic micro-bead tagged antibody (biomarker), which can be used to monitor certain stage specific antigens. Pluripotent hESCs can express certain stage specific embryonic antigens (SSEA-3, 4, and TRA-1-60) and genes (OCT-4) when undifferentiated\(^\text{17, 23}\). When directed towards differentiation, the stage specific biomarkers of pluripotency are down-regulated with new stage specific biomarkers being up-regulated. The new up-regulated biomarkers are specific to the phenotype produced through differentiation. For example, in a study by Draper \textit{et al.} the H9 hESC line was induced to differentiate by the addition of retinoic acid. During differentiation, the characteristic surface antigens of pluripotency were down-regulated and several ganglioside and glycolipid antigens appeared. These were detected by the VIN-IS-56 (GD3 and GD2), VIN-2PB-22 (GD2), A2B5 (GT3) and ME311 (9-O-acetyl-GD3) antibodies, representative
of neural ectodermal differentiation. However, these methods are reliant on biomarkers being completely cell type specific, with the assumption that each cell type (stem cell and differentiated phenotype) has a specific biomarker that will bind to cell type with high specificity. In terms of the latter, in many cases this is not true. The additions of exogenous fluorescent dyes or markers also come with problems; fluorescent biomarkers having a limited lifetime, after which they undergo irreversible photobleaching and can no longer be probed and the addition of such markers could modify a cell's surface chemistry, or even potentially damage a cell, compromising their clinical use. Other negative factors with the use of such markers include their use requiring specific specialist knowledge about the cellular systems that they are used with, analysis results saying little about the molecular changes due to differentiation, methods are time consuming/labour intensive and the agents needed for the analytical methods are expensive (biomarkers can be expensive). Therefore, in their current formats, immunocytochemistry, FACS and MACS are impractical for future clinical characterisation/cell sorting.

Other analytical methods used to characterise stem cells and their differentiated progeny include reverse transcription polymerase chain reaction (RT-PCR), real time polymerase chain reaction (qPCR) and both western, northern blotting and mass spectroscopy. All techniques have been successful in being able to monitor the molecular changes of a stem cell through cellular differentiation. However, all techniques rely on the destruction of the cell during cellular analysis. Analytical methods which involve the destruction of a cell during analysis are not appropriate for identifying/sorting stem cells and their differentiated progeny. Ideal methods for screening cells will have to be rapid, non-destructive, and without the need for exogenous markers/label. This opens up the door for optical vibrational spectroscopy, which has the potential to have all of these properties (no need for incredible expensive biomarkers or solvents) and with spatial resolutions high enough to capture spectral signals from single cells. In using vibrational spectroscopy, stem or differentiated cells can then be identified through their spectral signals. This may also give insight to the average macromolecular changes of a stem cell when proliferating and differentiating, therefore, offering more than just the quantification of added markers. One such technique is known as infrared spectroscopy.
1.2 Retinoid induced differentiation

Suboptimal culture conditions of embryonic stem cells can induce spontaneous differentiation. However, spontaneous differentiation produces a heterogeneous mix of differentiated progeny, representative of all three germ cell lineages, at different stages of differentiation, in vitro. When investigating the differentiation of embryonic stem cells, spontaneous differentiation is an unwanted event where the control of the experiment is lost. Ideally, when investigating differentiation, embryonic stem cells should stay pluripotent in culture until directed to differentiate down a certain lineage. In this experimental design the differentiated progeny will be known. Therefore, the chemical events which lead to the loss of pluripotency and gain of specialised function can be examined. One such compound which can induce the differentiation of embryonic stem cells is the retinoic acid, all-trans-retinoic acid (ATRA). In many studies, certain embryonic stem cell lines treated with ATRA are induced to differentiate down the ectodermal lineages producing neuronal cells.

1.2.1 Retinoids

There are over 4000 natural and synthetic retinoids, which are structurally and/or biochemically linked to retinol (vitamin A), a fat-soluble vitamin essential for many processes during life. ATRA is the major natural metabolite of vitamin A. ATRA and its two isomers, 9-cis-retinoic acid (9CRA) and 13-cis-retinoic acid (13CRA) (Fig. 2 (a)), have essential roles in both embryogenesis and adult homeostasis. These include cellular differentiation, proliferation, apoptosis and embryonic development. In particular, ATRA is seen to play a major role in the embryonic development and maintenance of the central nervous system (CNS) 5, 7, 39. This is exemplified from problems seen from humans with vitamin A deficiency (VAD); VAD causing night blindness, immune system impairment (can be potentially fatal) 6 and weight loss 40. However, not only deficiency in vitamin A can cause a problem, excessive uptake of vitamin A can also cause toxicity to the CNS, liver, bone, skin 40 and mutations of the foetus during embryogenesis 6. Therefore, it is critically important that optimum levels of vitamin A are maintained for normal cellular development and biological processes. Mammals cannot synthesise vitamin A so it must be ingested as β-carotene from plants, or from preformed vitamin A that has been digested by animals. Once digested, retinol is then stored in the liver as the retinyl ester, retinyl palmate, through reversible enzymatic activity. When needed, retinol is secreted from the liver and transported by a retinol binding protein (RBP) to the circulatory system 39.
Retinol can then pass through the cell membrane bound to the RBP receptor to either be stored for future use, or enzymatically converted to its active metabolite form, ATRA. Retinol is converted to ATRA in a two-step process. Firstly, the reversible oxidising of retinol by retinol dehydrogenases (ROLDH) and then the irreversibly oxidising of retinal by retinal dehydrogenases (RALDH) to produce ATRA. Gene ablation studies have shown that RALDH is a key enzyme for the post-implantation mammalian embryo. The removal of RALDH leads to early embryonic death. ATRA is the principle active retinoid found during early embryogenesis, a powerful regulator of neuronal differentiation and adult brain maintenance and functioning.

Fig. 2 (a) The chemical structures of all-trans-retinoic acid (ATRA) and its natural isomers 9-cis-retinoic acid (9CRA) and 13-cis-retinoic acid (13CRA). (b)Generic retinoid structure.
Cellular binding proteins, such as the cellular retinol binding protein I (CRBP-I) and the cellular retinoic acid binding proteins I and II (CRABP-I and II) are carrier proteins which have a high binding affinity to both retinol and retinoic acid respectively. CRBP-I is thought to be involved in the metabolism of retinol into either its storage form (conversion to retinyl esters), and in conjunction with the p450 cytochrome enzyme, the oxidisation of ATRA into its polar metabolites, 4-hydroxy-retinoic acid and 18-hydroxy-retinoic acid. The primary function of CRABP-II is thought to be the transfer of ATRA to the cell nucleus where it can then bind to its nuclear receptors.\(^7,39\)

The biological effects of retinoids are mediated by their interaction and binding with their specific retinoid nuclear receptors proteins belonging to the steroid/thyroid family.\(^39\) There are two classes of retinoid nuclear receptors, retinoid acid receptors (RARs) and retinoid x receptors (RXRs). Both act as ligand-inducible transcriptional regulation factors, regulating certain target genes that have multiple biological processes. The RAR nuclear receptors can be divided into three sub groups (α, β and γ), which in turn can then be subdivided into further receptor isotherms (RAR β\(_1\), RAR β\(_2\), RAR β\(_3\))\(^39, 41-42\). The RXR receptor can also be divided into three groups (RXR α, β and γ) adding to the complexity of genes regulated by retinoids throughout the body\(^39, 41-42\). ATRA regulates gene expression by binding to RAR nuclear receptors, which then forms heterodimers with RXR nuclear receptors (RAR-RXR), said to increase the transcription efficiency. RXR nuclear receptors have an affinity for the structural isomer of ATRA, 9CRA, and not ATRA. RXR nuclear receptors can homodimerise with themselves to form RXR-RXR dimers as well as with RARs and other nuclear receptors\(^39, 41-42\). However, 9CRA can also bind to RAR nuclear receptors. The reason for this is thought to be due to the shape of the 9CRA compound. Each nuclear receptor has a ligand binding pocket (LPB) which enables the retinoid ligands to bind. RAR receptors have a linear “I” shaped LPBs, whereas RXR receptors have shorter “L” shaped LBPs. This then result’s in the linear compound, ATRA, only having an affinity with RARs and the more flexible, 9CRA, being able to bind to both\(^39\). Abnormalities in the binding of these retinoid nuclear receptors have been shown to cause serious abnormalities in embryonic development and also been linked to various adult neurological diseases\(^6\).

Stem cell differentiation involves a series of epigenetic changes to the genome, but do not involve changes to the DNA sequence.\(^43\) The binding of ATRA to RAR nuclear receptors alters their interactions with proteins components of the transcription complex at numerous genes in stem cells. Through the binding of ATRA, some of these proteins components of
the transcription complex then either alter epigenetic marks on histones or on DNA thus, altering chromatin structure, which then leads to an exit from pluripotency and the differentiation process is started. Such epigenetic changes include DNA methylation and acetylation histone modifications of specific amino acid histones that surround DNA in the nucleus. DNA methylation involves the addition of methyl groups on cytosines of CpG di-nucleotides and is one of the best known epigenetic processes involved in gene expression. In cancer cells, DNA methylation was the first epigenetic alteration to be observed with hypermethylation of CpG islands near tumour suppressor genes being observed, which resulted in the switching off these genes. For embryonic stem cell differentiation, hESCs have been shown to have a unique DNA methylation signature when compared to differentiated cells. Pluripotent hESCs expressing unmethylated pluripotent genes, such as OCT-4 and Nanog, these genes are then found in a methylated state in differentiated cell phenotypes. DNA methylation has been shown through experiment to be essential for differentiation but not for maintenance of the pluripotent state. Therefore, in some hESCs, the addition of retinoids to cell populations initiates a sequence of interactions which modify the epigenetic state of the stem cells through methylation, altering gene expression and starting the differentiation process.

As well as inducing stem cell differentiation, due to the fact that ATRA and other retinoids have been shown to regulate genes that control cell apoptosis and differentiation, interest has been shown in the use of retinoids as a therapeutic agent to treat cancerous and diseased cells; the addition of retinoids differentiation or killing such cells. An example of this is the use of ATRA to treat Acute Promyelocytic leukaemia (APL), ATRA driving malignant cells into differentiation and death. However, ATRA has been shown to isomerise, both in the cell through established biological processes, and under normal lab conditions during tissue culture etc.

Isomerism of ATRA in the cell is seen as an important biological pathway due to the different isomers of ATRA regulating different genes. The isomers of ATRA, 9CRA and 13CRA, have been thought to the cause of unwanted side effects reported from patients treated with ATRA. For example, 13CRA is used in the treatment of acne but has been linked to increases in suicide and depression among some patients; a serious problem when treating less severe ailments. Therefore, care must be taken when administering ATRA. Otherwise the biological side effects 9CRA and 13CRA will be felt by patients.

ATRA also isomerises under normal tissue culture conditions due to the five conjugated
double bonds in its chemical structure which act as chromophores; ATRA absorbing light in the 300-400 nm wavelength range \(^6\)-\(^7\). Therefore, if ATRA is to be used to induce cellular differentiation and appropriate care is not taken to stop isomerisation, a heterogeneous mix of differentiated derivatives will be produced in, \textit{in vitro}, much like spontaneous differentiation. This is because of the isomers of ATRA having alternative abilities in inducing cellular differentiation, possibly as a result of their different binding abilities to retinoid nuclear receptors \(^6\). Isomerisation also results in the reduced concentration of ATRA in culture. Cellular differentiation has been seen to be activated in a concentration dependant manner. Therefore, a reduced concentration may also reduce the potency of ATRA for cellular differentiation \(^6\). However, ATRA can also be broken down into its polar metabolites in cell by the p450 cytochrome enzyme. This will also result in the reduced concentration of ATRA in culture, but the polar metabolites, 4-hydroxy-retinoic acid and 18-hydroxy-retinoic acid \(^39\), may also induce different cellular responses in culture. The increased metabolism of ATRA in the cell has been linked with patient resistance to ATRA during long term during clinical treatment \(^49\). Therefore, even if measures are put in place to reduce the isomerisation of ATRA, the concentration of ATRA can still be reduced through biological pathways within cells. This results in ATRA not being an ideal compound to induce cellular differentiation. To reduce the variability in the differentiated progeny produced and to improve the reproducibility in directed cellular differentiation, a compound which can induce differentiation but will stay in the same form and concentration every time it used, is needed. Currently this is not possible with ATRA \(^5\), \(^48\).

Initially, attempts were made to stop the isomerisation of ATRA by adding additives to the ATRA compound. However, this did not completely solve the problem and the addition of such additives had the potential for adverse side effects on stem cells. A new solution was sought, and this came from the design and production of synthetic retinoids \(^5\). Synthetic retinoids can be designed so that they have the same or similar binding to that of ATRA but are stable under typical laboratory conditions during cell culture. They also have the potential to be designed so that they have greater retinoid binding receptor selectivity. Research into these synthetic retinoids can help in the understanding of each RAR and RXR biological roles in both tissue and during embryogenesis, as well as reducing toxicity seen from treatments using ATRA.
Recently, some synthetic retinoids have been synthesised at Durham University. These synthetic retinoids are based on the hydrophobic and polar terminus of the ATRA structure, but without the light sensitive linker unit (Fig. 2 (b)). The synthesised synthetic retinoids are known as EC23, EC19 and AH61 (Fig. 3). EC23 and EC19 have had the light sensitive polyyne removed and replaced with two aromatic rings linked together with a carbon triple bond. This new structure increased the physical stability of the compounds resulting in them not being sensitive to light. EC19 and EC23 are structural isomers; the only difference in structures is the position of the carboxylic acid functional group on the polar terminus of the retinoid. EC19 has a carboxylic acid functional group in the *meta* position, while EC23 has a carboxylic acid in the *para* position on the aromatic ring. Both have had their biological effects on the TERA2.cl.SP12 cell line tested. Christie *et al.* showed through both immunofluorescence and flow cytometry, that EC23 could induce the differentiation of the TERA2.cl.SP12 cells in a similar manner to that of ATRA, producing neuronal cell types. EC19 however, did not have the same biological action. Instead, EC19 produced mainly epithelial and very few neuronal cells after 14-21 days (Fig. 4). This showed that even slight changes in the chemical structure can have different biochemical effects. It is thought that the differing biological response induced by EC19 is due to its structural shape. EC23 mimicked the linear “I” shape of ATRA, thus, similar binding to the RAR nuclear retinoid receptor proteins. However, having the carboxylic acid in the *meta* position on the aromatic ring gives the chemical structure of EC19 a different shape. This could alter the binding to the retinoic binding proteins and may explain the differing biological action on the TERA2.cl.SP12 cells. Recently, another synthetic retinoid has been synthesised at Durham University, known as AH61. Like ATRA and EC23, AH61 has been shown to induce the differentiation of the TERA2.cl.SP12 cell line to produce neuronal phenotypes, again similarly to that of ATRA and EC23. AH61 chemical structure showed only limited signs of isomerisation and no chemical degradation, unlike ATRA.

![Fig. 3 The chemical structures of the synthetic retinoids AH61, EC19 and EC23.](image)
Christie et al. showed that even though both EC23 and ATRA induced the differentiation of the TERA2.cl.SP12 cells to form neuronal cells (Fig. 4), EC23 showed an ability to produce neural tissue at a similar or even better degree than that observed for ATRA. EC23 was also shown to induce differentiation at lower concentrations to ATRA. This could be a direct result of EC23’s greater stability (no isomerisation) and the possibility that EC23 is resistant to metabolism through the same ATRA biological pathways in a cell, resulting in a greater concentration of EC23 in the cells/culture media. However, the possibility of EC23 having a greater binding affinity with the nuclear retinoid binding proteins cannot be ruled out as a factor either.

Fig. 4 Immunocytochemical staining for the neuronal marker Tuj1 after 21 days of culture in the presence of 10 mM of ATRA, EC19 or EC23. Control cells (undifferentiated) displayed no expression of Tuj1. ATRA and EC23 treated cultures contained areas of high Tuj1 expression with cultures containing EC19 restricted to areas containing highly proliferative cells. No expression was detected in the large flat cells which dominated the EC19 cultures. Nuclear counter-staining with Hoescht 33342 is shown for the EC19 and control cultures to show the position of the cells.

A study by Maltman et al. looked into the different protein profiles (using MALDI-TOF MS) of TERA2.cl.SP12 cells after treatment with EC19, EC23 and ATRA, after 7 days. Results showed that the stable retinoid, EC19, had a different protein profile to that of the pluripotent TERA2.cl.SP12 cells treated with both ATRA and EC23. This was not surprising considering the epithelial plaques were produced during the cultures of EC19 treated cells (Fig. 4). Results also showed that both ATRA and EC23 treated cells had different protein profiles from the pluripotent EC cells; both ATRA and EC23 treated cells
showing similar up-regulation of cellular retinoid response proteins (CRBP-I, CRABP-I and II), indicating a similar biochemical response for EC23 and ATRA. However, results also showed some slight, but significant differences in protein profiles; EC23 treated EC cells greater Profilin-1 content when compared to ATRA treated EC cells. Profilin-1 is known to play a key role in neurite outgrowth and elevated levels have been seen during ATRA treatment of cancerous cells, indicating a role in the inhibition of cellular proliferation. These slight protein profile differences between ATRA and EC23 treated cells, found using mass spectrometry, would typically go unnoticed using flow cytometry and immunofluorescence and could be beneficial in understanding the molecular pathways induced by ATRA and EC23, and also their ability to induce cellular differentiation. However, mass spectrometry can be extremely time consuming and involves the destruction of the cells during analysis. Therefore, through use of IR-MS, can spectral differences be seen between the pluripotent TERA2.cl.SP12 cells and the retinoid treated (AH61, EC19, EC23 and ATRA treated) TERA2.cl.SP12 cells left in culture for 7 days? If so, can the same also be said for cells left in cultures for 3 and 5 days? In addition to these points, similarly to previous studies, can spectral differences between neuronal differentiating cells (ATRA, EC23 and AH61) be seen?

1.3 Infrared Spectroscopy

Infrared radiation (IR) is found in the infrared region of the electromagnetic radiation (EMR) spectrum. The infrared spectral region covers a wide range of frequencies, typically segmented into the near (NIR), mid (MID) and far (FIR) infrared ranges. Each wave of IR propagates in a certain direction with a sine wave motion and is defined by its constant wavelength ($\lambda$) or, frequency ($\nu$); frequency being the number of complete wave cycles that occur in a second (Fig. 5). A wave of radiation also has energy, which is directly proportional to the frequency and can be understood using the Max Planck’s quantum formula

\[
E = (h\nu) = \left(\frac{hc}{\lambda}\right) \text{ J mol}^{-1}
\]  

(Eq. 1)
Fig. 5 Electromagnetic wave; wavelength ($\lambda$) – the distance travelled by the wave to complete one wave cycle \textsuperscript{51-52}.

For infrared spectroscopy, IR radiation is typically passed through a sample of interest (transmission spectroscopy) and collected by a detector. When IR radiation is passed through a sample frequencies of IR radiation are absorbed when the frequency of IR radiation is equal to the frequency of a specific vibration of molecule, resulting in the molecule becoming quantised and raising the molecule to a higher vibrational energy state (Fig. 6). The frequency of the IR radiation absorbed depends on the atoms involved in a molecule and the strength of any intermolecular interactions. Therefore each molecule can have a different spectrum; in essence the IR spectrum is a fingerprint of the biochemical structure of the sample.

Fig. 6 The atomic displacement of the HCl molecule after a quantised transition to a higher vibrational energy state (harmonic displacement of the hydrogen atom); raising the HCl molecule to the higher vibrational state results in increased displacement of the hydrogen atom \textsuperscript{52}.

All molecules have $3N$ ($N =$ number of atoms in the molecule) degrees of a freedom. For a simple molecule, such as water ($\text{H}_2\text{O}$), three atoms are involved thus, nine degrees of
freedom and three of representing vibrational modes of freedom. For linear molecules, such as carbon dioxide (CO₂), four vibrational modes are allowed. However, not all molecular vibrations absorb IR radiation, for a molecule to absorb IR radiation the displacement of the molecular atoms must cause a change in dipole moment. For example, the CO₂ molecule is linear in shape and as a result has four vibrational modes, the symmetric stretching, the anti-symmetric and two bending vibrational modes; the two bending vibrational modes of CO₂ have the same energy and are degenerate (Fig. 7).

During the symmetric stretch of the molecule the oxygen atoms of the molecules are both stretched and compressed simultaneously causing no change in net dipole. As a result, the symmetric stretch vibration of the molecule does not absorb IR radiation, whereas both the anti-symmetric and bending vibrations of the CO₂ molecule do.

As well as being defined by frequency, energy and wavelength, a wave of radiation can also be defined by its wavenumber, which is the measure of many complete wave cycles occur in one unit of length (normally a centimetre) and has the advantage of being directly proportional to energy (υ which has units of cm⁻¹). IR frequencies in the MIR spectral range (400 – 4000 cm⁻¹) have energies which correspond with the vibrational energies of organic molecules. Therefore, recording spectra of biological samples using MIR radiation allows measurement of the molecules that make up the sample.

1.3.1 Absorption Band Assignment
With the vibrational frequency of a molecule being directly related to the atoms involved and the strength of any intermolecular interactions, absorptions of IR frequencies can be
assigned to their particular molecular vibrations. These molecular vibrations may then be assigned to certain macromolecules present in the sample. Fig. 8 shows a typical absorption spectrum of a biological sample. In Fig. 8 the main absorption bands are assigned to their representative molecules. As a general rule, stronger molecular bonds tend to have greater vibrational frequencies than weaker ones. Therefore, triple bonds have greater frequencies than double and single bonds. Bending a bond is also easier than stretching one leading bending vibrations to have lower vibrational frequencies than stretching vibrations. Hydrogen bonding also affects the frequencies at which a bond vibrates, with hydrogen bonding increasing the bond strength resulting in greater frequencies of absorption. Overtone and combination absorption bands can also be found in a spectrum; overtone absorption bands being multiples of the fundamental vibration and combinations absorption bands occurring when two or more fundamental absorptions excited simultaneously, these absorption bands have weaker intensities than fundamental vibrations. The combination of all the absorption bands results in a unique IR spectrum for the biological sample under investigation.

![Image of typical mid-infrared absorption spectrum](image)

**Fig. 8** Typical mid-infrared absorption spectrum of a biological sample including the molecular assignment of the main absorption bands found in the MIR spectrum and their macromolecular assignment. 

- Amide I: C=O
- Amide II: N-H, C-N
- C-H: Lipids
- C=O ester: Phospholipids
- N-H: Proteins
- C-N: Lipids
- PO$_2^-$: Carbohydrates
- N-H: Nucleic acids
- C-O: Lipids
The determination of the molecular vibrational frequencies has been made easier by the introduction of vibrational frequency correlation tables, often referred to as "Colthup charts", which were developed to allow chemists to assign absorption bands in a spectrum to a vibrational mode of a specific molecule.

1.4 The Macromolecules that make up a Biological Sample

Biological samples, such as cells and tissue, are made up of four main macromolecules (Nucleic acids, carbohydrates, proteins and fatty acid (lipid) structures). These main macromolecules provide the basis for life and play critical roles in biological functions, such as death, proliferation and differentiation. The vibrational frequencies of the functional groups of these macromolecules are well defined and have been assigned in Fig. 9. As certain IR absorption bands in the IR spectrum can be assigned to certain macromolecules, the monitoring of the size, shape, position and intensity of each of these absorption bands can help to explain the net biochemistry changes of the main macromolecules contained within a biological sample.

1.4.1 Nucleic Acids

Nucleic acids can be categorised into two groups, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is made up of 4 different nucleotide bases. (Fig. 9) These bases are paired together as adenine (A) to thymine (T) and guanine (G) to cytosine (C). In RNA, T is replaced by uracil (U). The bases are paired together by hydrogen bonding. Complementary base pairs are connected to other base pairs though a sugar-phosphate backbone (Fig. 14). DNA is predominately found in the cell nucleus with some DNA being also found in mitochondria. The nucleotides which make up the DNA structure can act as short term carriers of energy but are mainly known for the storage and retrieval of genetic information. RNA is synthesised from DNA, when needed, and carries the genetic information needed to synthesize proteins.
For both DNA and RNA, phosphate bonds make up the sugar-phosphate backbone holding the DNA and RNA strands together. These phosphate groups can be directly associated with nucleic acids as no other macromolecule found in the cell has a phosphate group as part of the molecular structure. The P=O stretching vibration, frequency around 1250 cm$^{-1}$ can therefore be a discriminate absorption for nucleic acids, as well as the symmetric phosphate stretching vibration and the triplet of nucleic acid bands around 1055 cm$^{-1}$ (~1040, 1060 and 1080 cm$^{-1}$)\textsuperscript{54-55}.

### 1.4.2 Proteins

Proteins are made up of many amino acids. Amino acids have both amine and carboxylic acid functional groups and are linked together by a covalent polypeptide bond (Fig. 10). Each protein can be characterised by its unique sequence of amino acids. Proteins make up the majority of a cell’s mass and are considered as the building blocks from which cells are built. The genetic information contained within DNA gives out instructions for organism functioning and development. In order for the instructions to be passed on, genetic information contained in DNA is transcribed into messenger RNA (mRNA), which then synthesise specific proteins. These proteins execute almost all functions in the human cell. For example, some proteins are enzymes promoting many chemical reactions from a cell’s surface and other proteins can transport messages from cell to cell. They can also be found in the plasma membrane of cells, forming channels which allow small molecules in and out of the cell; these small molecules carry messages to the cell nucleus\textsuperscript{56}.
As proteins are made up of primarily amide groups which link amino acids together, the main vibrational absorptions from proteins will be due to a carboxyl (C=O) and an amine bond (N-H) (amide absorption bands). Typically, the two greatest absorption bands from a biological sample are the amide I and II absorption bands. The amide I absorption, located at ~1653-58 cm$^{-1}$, is made up of vibrations from the C=O stretching (80%), C-N stretching (10%) and N-H bending (10%) vibrational modes. The amide II absorption, located at ~1550-60 cm$^{-1}$, is made up of vibrations from the N-H bending (60%) and C-N bending stretching (40%) vibrational modes. Fig. 12 shows that the amide I and II absorption bands also contain overlapped absorptions with nucleic acids and lipids chemical structures. However, the main absorption contributions making up the two bands come from protein molecules. Because proteins are the most abundant macromolecule in a biological sample, not surprisingly, both the amide I and II absorptions are the most intense bands in an IR spectrum of a biological sample. The amide I absorption band gives the most information on protein structure and has been used the most to look at small changes in the molecular geometry and hydrogen bonding of the peptide bond, synonymous with proteins.

The amide I band contains C=O molecular information from the peptide backbone of proteins, which also reveals protein secondary structure. The secondary structure describes the folding pattern of the protein; the two main folding patterns are known as α-helix and β-sheet molecular structures. Both structures differ through their hydrogen bonding, which therefore defines their structural shape. Proteins with a α-helix configuration have a single
polypeptide chain which twists around itself to form cylinder (Fig. 11). Proteins with α-helix conformations have a hydrogen bond between the C=O molecule from one peptide to the N-H of another peptide after every fourth peptide bond; this maintains the helix structure. Regions of α-helix are extremely abundant in proteins, such as transport proteins and receptors found in the cell membranes. The spectral band representative of α-helix structural protein can be found in the IR spectrum between 1650-1660 cm\(^{-1}\).

Proteins with β-sheet conformations can either form through parallel polypeptide chains, or with polypeptide chains which run in opposite directions (anti-parallel). The structure is held together through many hydrogen bonds between neighbouring polypeptide chains (Fig. 11). Absorption bands characteristic of β-sheet proteins can be found in the IR spectrum between 1620-1640 and 1690-1695 cm\(^{-1}\). By investigating IR bands associated with both α-helix and β-sheet proteins the total α-helix and β-sheet protein content of a sample can be quantitatively understood.

![Fig. 11 Both α-helix and β-sheet protein secondary molecular structures](image)

1.4.3 Fatty Acids and Lipids
A fatty acid is made up of a chemically unreactive, long hydrophobic, hydrocarbon chain, which has a highly reactive carboxylic acid located at one end of the molecule. Certain fatty acids, such as triacylglycerols, also act as an energy source for the cell. Like glucose,
when they are broken down they release energy to the cell. They are predominately found in the cytoplasm of the cell.

Lipids are another type of fatty acids. Lipids are main the major molecular constituent of the cell membrane, with the cell membrane being made up of mainly phospholipids. Lipids are therefore critical in all biological systems as they keep all the macromolecules in their necessary compartments. Other types of lipids regulate the flow of molecules through the lipid cell membrane barrier and have also been seen to be involved in cross talk with other molecules (sphingolipids).

Lipids are mainly made up of C-C and C-H bonds that have vibrations primarily found in the 2830-3000 cm$^{-1}$ spectral range. Fig. 12 shows that all four major macromolecules absorb in the 2830-3000 cm$^{-1}$ spectral range as well as lipid molecular structures. This is not particularly surprising considering that C-C and C-H bonds are the building blocks of organic molecules. However, the problem of absorption band overlap is not considered that much of a problem in this spectral range due to lipid molecule vibrations having a greater signal than the other three main macromolecular structures. Therefore, generally vibrations between 2830-3000 cm$^{-1}$ are characteristic to lipid molecular structures when investigating biological samples with IR.

1.4.4 Carbohydrates

Carbohydrates, such as glucose, have hydroxyl and aldehyde functional groups contained within their chemical structure. Cells will obviously need energy in order to do work and perform functions. Sugars (saccharides), also known as carbohydrates provide energy for cells but a monosaccharide known as glucose is considered as the main source of fuel. Glucose provides energy when it is broken down by glycolysis, breaking the molecule into two pyruvate molecules. Each pyruvate molecule is then broken down further to water and some nitrogen atoms producing energy for the cell.
The main functional group of carbohydrates is the C-O molecular bond. Typically absorption bands with peak maxima’s at 1020, 1030 and 1040 cm\(^{-1}\) can be associated carbohydrates\(^{54}\). DNA and RNA structures contain a sugar-phosphate backbone which also contains C-O molecules meaning DNA/RNA will also contribute to absorptions associated with the C-O molecular bond. However, the absorption band at \(\sim 1024\)-30 cm\(^{-1}\) is predominately associated with C-O vibrations from carbohydrates\(^{54,\ 61}\).
Chapter 2: Methodology and Principles/Pre-Processing

For an IR spectrum to be produced, IR light is produced from a source, the light emitted then interacts with a sample, and the light interactions are measured at a detector. This thesis focuses on the use of IR absorption spectroscopy (MIR light frequencies 900-4000 cm\(^{-1}\)) to record IR spectra from biological in transmission mode; transmission mode meaning that the light from a source is passed through a biological sample and the loss of light intensities due to frequencies of IR light being absorbed by the sample are measured at a detector. In the previous chapter it was explained that the vibrational frequency of a molecule is directly related to the atoms involved and the strength of any intermolecular interactions, therefore, absorptions of IR frequencies can be assigned to particular molecules, which in turn can, give information of the certain macromolecules present in the sample. However, IR spectroscopy can also be used as a method to measure how much of something is present in the sample. Therefore, the analytical technique can be used for both qualitative and quantitative measurements.

2.1 Beer Lambert Law for Quantitative Analysis

According to the Beer-Lambert law, the amount of IR EM radiation absorbed and transmitted through a sample is directly related to the concentration and thickness of the sample (Eq. 2)\(^51\). From Eq. 1 it can be seen that absorbance is related to the multiplication of molar absorptivity of the sample (\(\varepsilon\)), the concentration (\(c\)) and the pathlength of the sample (\(l\)).

\[
A = \varepsilon c l
\]

(Eq. 2)

The Beer-Lambert law can also be used to describe the relationship between the intensity of IR light entering the sample (\(I_0\)) and light leaving (\(I\)); the absorbance (\(A\)) being the difference between the two (Fig. 13). Taking the log of initial light intensity divided by the light intensity leaving the sample will also give the amount of light absorbed by the sample (Eq. 3)\(^62\).
Fig. 13 Monochromatic radiation directed through a sample. Intensity of the transmitted frequency of IR EMR \((I)\) reduced due to absorption.

\[
A = (\log \left( \frac{I_0}{I} \right))
\]  

(Eq. 3)

The Beer-Lambert law also states that the absorbance should be linear with increasing concentrations and a gradient of \((\epsilon l)^2\). In using the Beer-Lambert law and applying it to all frequencies in the mid-infrared spectrum, a transmission or absorption spectrum can be plotted against wavenumber frequency. With each absorption band in a spectrum being directly related to concentration, calibration curves can be produced for quantitative analysis. This can only be achieved if standards of known concentration can be made from a known sample. With this type of quantitative analysis unknown concentrations of the sample can be understood.

2.2 Fourier Transform Infrared (FT-IR) Spectroscopy

Fourier transform infrared (FT-IR) spectroscopy incorporates an interferometer based on the Michelson interferometer (Fig. 14). The Michelson interferometer has a beamsplitter which splits the IR beam of radiation emitted from the source into two paths and then recombines the beams after a path difference is introduced; path difference introduced through shifting the moveable mirror. The path difference causes interference between the two beams when they recombine at the beamsplitter. The interference beam recombined at the beamsplitter is then sent on to be measured as a function of the path difference at the detector.
A Michelson interferometer is made up of two mutually perpendicular mirrors, one which can be moved and one that is fixed in position (Fig. 14). An interferometer also needs a beam splitter to partially reflect a collimated beam of radiation to the fixed mirror and partially transmit the radiation to the moveable mirror. For an ideal beamsplitter, 50% of the radiation will be transmitted and 50% reflected; however, in practical terms this is seldom the case. When the two beams reflect back of the mirrors and return to the beamsplitter they interfere and are again partially transmitted and reflected\(^{51, 53}\).

For a rapid continuous scan interferometer the moveable mirror is moved constantly at a velocity greater than 0.1 cm s\(^{-1}\). The moveable mirror is controlled by a Helium Neon (He-Ne) laser and provides an optical path difference (OPD) of \((n + \frac{1}{2})\lambda\)\(^{51, 53}\). If one considers a collimated monochromatic beam of IR radiation being outputted from the source, when the two mirrors are mutually perpendicular, all wavelengths of the IR beam interfere constructively at the beamsplitter. This is also known as the zero path distance ((ZPD) also known as the centreburst for polychromatic radiation) (Fig. 15). When the mirrors are in this arrangement the detector receives a sum of all of the frequencies of IR radiation intensities, with all light travelling to the detector and none to the source. When
the moveable mirror is moved by a small displacement, for example, $\frac{1}{4}$ of a wavelength, the pathlengths of the two split beams of monochromatic radiation are now $\frac{1}{2}$ of a wavelength of out of phase. In this arrangement, the two beams of IR radiation now interfere destructively at the beamsplitter. When interfering destructively, instead of the complete sum of the two beam intensities travelling to the detector, now, the complete sum of the two beam intensities travel to the source. Moving the mirror by again another $\frac{1}{4}$ wavelength will put the two beams out of phase by 1 wavelength. When displaced by a one whole wavelength, the beams will again be in phase, interfering constructively, with the sum of the two beam intensities reaching the detector $^{52-53}$.

As the mirror moves away from the ZPD constructive and destructive interferences occur when the two beams are recombined. As a result, the detector receives a signal which is the intensity of the light beam (signal intensity) vs. the optical path difference (or time) (Fig. 15). The interferogram in Fig. 15 can be though as a “coded” representation of all the frequencies of light being emitted by the source. Therefore, through placing a sample in the pathway of the interference signal travelling to the detector, lost signal intensities from each light frequency can be measured, creating a transmission spectrum. An absorption spectrum is produced by transforming the time domain interferogram (intensity of light at the detector vs. OPD, or time) to the frequency domain. This is done using the two halves of the cosine Fourier transform pair, Eq. 4. Eq. 4 (a) is used to produce an interferogram, like in Fig. 4, and (b) is used to produce an absorption spectrum from the interferogram (absorption vs. frequency) $^{53}$.

\[
I(\delta) = \int_{-\infty}^{+\infty} B(\tilde{\nu}) \cos 2\pi \tilde{\nu} d\tilde{\nu} \quad \text{(Eq. 4 (a))}
\]

\[
B(\tilde{\nu}) = \int_{-\infty}^{+\infty} I(\tilde{\nu}) \cos 2\pi \delta d\delta \quad \text{(Eq. 4 (b))}
\]
2.2.1 Advantages of FT-IR

Before FT-IR spectrometers dispersive IR spectrometers were solely used for IR spectrometer analysis. Dispersive spectrometers incorporate a prism/grating to separate the light being emitted from the source into individual frequencies with each individual frequency then directed through a small slit to the sample and detector; each frequency of light being measured one at a time. As a consequence, to record an IR spectrum over a large frequency range requires the sequential measurement of individual frequencies.

For FT-IR spectroscopy, all frequencies of light reach the detector and are measured simultaneously resulting in a time advantage (equal to the number of resolution elements in the spectrum (M)) and secondly, a sensitivity advantage (SNR of a spectrum measured with a FT-IR spectrometer being greater by $\sqrt{M}$ over dispersive instruments (multiplex advantage)) $^{51}$. As a result of the time advantage, the SNR can be increased further by a process known as signal averaging, which gives a SNR advantage of $\sqrt{N}$ (with N being equal to the number of co-added scans) $^{51}$. A further advantage being known as the Jacquinot or throughout advantage (Without slits needed to let through only certain frequencies of light more light reaches the detector, which again improves the SNR).
The spectral resolution of a FT-IR spectrometer is governed by the moving mirror in the interferometer, the limiting spectral resolution being the reciprocal of the pathlength difference (cm). As a result of this, the mirror only has to move small increments in distance to achieve very good spectral resolution i.e. 0.25 cm to achieve 4 cm\(^{-1}\) spectral resolution \(^{51}\). Therefore, the movement of the mirror in an FT-IR spectrometer is obviously vital, requiring great precision. To achieve the high precision needed, FT-IR spectrometers employ a He-Ne laser that acts as an internal reference for each scan, internally calibrating the wavenumber locations and giving excellent wavenumber precision. This is known as the Connes advantage, which gives FT-IR spectrometers better reproducibility than dispersive instruments \(^{53}\).

### 2.2.2 FT-IR Micro-Spectroscopy (FT-IRMS)

The speed and sensitivity of FT-IR spectrometers make them ideal for recording spectra from biological samples. However, biological samples tend to be very small in size, so small that they are difficult to see with human eyes. Biological material also tends to be extremely heterogeneous (large array of different types of tissue, i.e. nerve, muscle, connective and epithelial which make up the composition of human tissue). Early FT-IR spectrometers were designed to analyse homogeneous and large scale samples, recording an average spectrum of large spatial area. This meant that early work using FT-IR spectrometers to investigate biological samples were not particularly useful. Therefore, in order for FT-IR spectroscopy to be used for the analysis of biological samples the instrument design had to change \(^{63}\). The change came through combining optical light microscopy with infrared spectroscopy. An optical light microscope was fitted to an FT-IR spectrometer and FT-IR micro-spectroscopy was born. The optical microscope enabled users to visualise and target areas of a biological sample to interrogate with the IR beam. The IR beam is produced from the source, passed through the interferometer, through the infrared microscope to the sample and to the single element detector where an average spectrum is recorded \(^{64}\) (Fig. 16). The FT-IR microscope employed in a FT-IR micro-spectrometer is very similar to conventional microscopes except that they contain reflective mirrors instead of lenses \(^{65}\). This allows the radiation to pass through the microscope without losses in light intensity through absorption at the mirror surfaces.
Fig. 16 Schematic of an FT-IR micro-spectrometer incorporating a single element detector. Confocal operation can be used when both the upper and lower apertures employed.

Fig. 17 Schematic of a Schwarzschild objective incorporated within the IR microscope and IR light being directed through a single cell.
Typically, to focus the light onto a sample a Schwarzschild objective with a numerical aperture (NA) of \(~0.6\) is used \(^{64}\) (Fig. 16 and 17). Fig. 17 reveals how the upper Schwarzschild objective focuses IR radiation delivered from the source onto extremely small samples. The attenuated beam (loss of light intensity through absorption) leaving the sample is then re-focused using the bottom Schwarzschild condenser to a single point detector where the interferogram is measured.

Visible light and IR radiation follow the same trajectory through the microscope. Therefore, not only can visible light passed through the IR microscope be used to find biological cells on IR substrate samples, but it is also used to adjust the focus of the light. When operating the microscope in single point mode, both the upper and lower apertures can be used (if both are used then operating in confocal mode) to improve the spatial resolution (Fig. 6) \(^{53, 64}\). Confocal mode being used to reduce out of focus background signals (neighbouring biological matter that are not being targeted for spectral acquisition) while passing through focused signals to the detector, improving the SNR \(^{66}\).

![Single cell deposited onto an infrared transparent substrate. One cell is targeted with upper apertures, roughly 20 x 20 µm in size. Apertures are used so that only the targeted sample is illuminated by the source. An average spectrum of a single cell will be recorded.](image)

Fig. 18 Single cell targeted with apertures (visualised using visible light).

### 2.3 Background Spectrum

Both carbon dioxide (CO\(_2\)) and gaseous water vapour are present in the atmosphere. Small gaseous molecules, such as CO\(_2\) and H\(_2\)O vapour, show rotational fine structure in their recorded IR spectra. This is due to the fact that quantised transitions for both rotations and vibrations occur at the same time. This also applies to absorption spectra recorded from larger molecules in the gas phase. However, the individual fine structures of the transitions are too close together to be resolved in the spectrum when recording spectra using typical spectral resolutions. The vibrational-rotational fine structure of molecules in a liquid state, are also not seen due to the fact that collisions occur at a greater frequency than the rotations of the molecule \(^{53}\).
As both CO$_2$ and H$_2$O vapour are present in the atmosphere then absorptions from both CO$_2$ and H$_2$O vapour will also be found in every IR absorption spectrum recorded. Fig 19 shows the absorption contributions from CO$_2$ and H$_2$O vapour in the MIR spectral range; only the asymmetric stretching vibrational mode of CO$_2$ absorbing in the MIR spectral range (~2349 cm$^{-1}$), whereas two fundamental vibrational modes of H$_2$O vapour absorb in the MIR spectral range (the bending vibrational mode of H$_2$O (spectral range: 1350-2000 cm$^{-1}$)) and combination band (combination of absorptions from the symmetric and asymmetric stretching modes of CO$_2$ and the overtone of the bending vibrational mode) which has a spectral range of 3500-4000 cm$^{-1}$. Fortunately, biochemical matter do not absorb in the spectral range covered by the CO$_2$ asymmetric stretching absorption band. However, the absorption contributions of both H$_2$O vapour absorption bands are present over large frequency ranges, overlapping the absorption bands found in spectra recorded from typical biological samples (see Chapter 1.3-1.4).

![Fig. 19 Vibration–rotation MIR absorption spectrum of gaseous small molecules found in the atmosphere (CO$_2$ and H$_2$O vapour).](image)

To remove absorption contributions from the CO$_2$ and H$_2$O vapour found in the atmosphere, a background spectrum containing absorptions solely from the atmosphere is subtracted from a biological sample spectrum. However, if the sample spectrum and the background spectrum do not contain the same concentration of both CO$_2$ and water vapour, then absorptions due to the atmosphere will not be subtracted out properly, having a negative effect on the sample absorption spectrum; under or over subtracted water vapour producing spikey absorption bands, which can be a problem when comparing recorded
spectra using multivariate analysis, vide infra. Therefore, as a way to combat the constantly changing atmospheric concentrations of CO₂ and H₂O vapour, the sample compartment is constantly purged with dry nitrogen gas (nitrogen purging), which has been passed through a desiccator in order to remove moisture from the atmosphere. Another safe guard is to record a background spectrum regularly in order to capture the changing environment.

2.4 Sources of IR Radiation

To produce a continuous source of infrared radiation with frequencies in the mid-infrared spectral range, high temperature Blackbody sources can be used in FT-IR spectrometers. Blackbody sources have been used to emit IR EMR with frequencies in the MIR spectral range. They absorb nearly all incident wavelengths of electromagnetic radiation that come into contact, reflecting very little. When a Blackbody source is heated wavelengths of the EMR absorbed are emitted. The most common blackbody source used for FT-IR spectroscopy is the Globar source 53.

2.4.1 Synchrotron Radiation

A typical biological cell size is typically between 10 x 10 to 20 x 20 µm in size. Apertures can be used to confine IR light from the source to the particular areas of interest, i.e. single cells. However, this reduction in aperture sizes ultimately reduces the infrared photon flux reaching the detector thus, reducing the SNR 67. Conventional thermal sources, such as the Globar source, perform admirably when large apertures are employed as the majority of light reaches the detector. However, when apertures are closed down to 30 x 30 µm or smaller, Globar sources struggle to achieve an effective SNR. In order to achieve the small spatial resolutions required to spectroscopically investigate single cells while achieving acceptable S/N, a brighter IR light source is needed.
IR light produced from a synchrotron source is brighter than a conventional Globar source, producing excellent S/N at apertures that are at, or below, the diffraction limit (Fig. 20). Synchrotron radiation is produced when electrons from an electron gun are accelerated to a very high speed round a circular shaped storage ring at a synchrotron facility. Synchrotron radiation is the light emitted from relativistic electrons when they are either accelerated or decelerated along a circular synchrotron storage ring; a storage ring is made up of bending magnets that deflect the charged electrons around a circular orbit. Light is generated when electrons are traversely accelerated through the curved orbit inside bending magnets (bending magnetic radiation), or though accelerating/decelerating electrons between the straight and curved trajectories when leaving/entering bending magnets (edge radiation). The light radiation generated by a storage ring is known as synchrotron radiation. This synchrotron radiation emitted from the storage ring is then channelled into analytical instruments. Laboratories are then built adjacent to the where the synchrotron radiation is emitted. Synchrotron radiation is effectively a white light source with many frequencies of radiation spanning a large spectral range. The radiation can then be focused and tuned to let through only wavelengths of particular interest for the type of instrumentation being used. For FT-IR micro-spectroscopy, the radiation is directed into an interferometer in the same manner as the light produced from a Globar source.

The main advantages of synchrotron light are its high flux (when compared with many thermal broadband sources at many wavelengths), highly collimated emitted radiation and
small source size \(^{71}\). These properties give synchrotron radiation a high intrinsic brightness. However, the increase in brightness is not due an increase in power but rather due to light being emitted from a small sized source and collimated into small spatial regions without large losses in light intensity \(^{64, 70}\). As a result, FT-IR micro-spectrometers incorporating synchrotron light can achieve excellent spatial resolutions while retaining good S/N. This enables single cells and small spatial regions of biological tissues to be investigated. It has even been shown that spectra can be recorded from spatial areas that are smaller in size than the wavelengths of light being used to investigate. However, under these circumstances the spatial resolution will be diffraction limited \(^{69}\). For example, the mid infrared spectral range typically has frequencies that are between 2.5 \(\mu\)m (4000 cm\(^{-1}\)) and 12.5 \(\mu\)m (800 cm\(^{-1}\)) in wavelength. Therefore, if apertures of 10 x 10 \(\mu\)m or smaller are used, diffraction will occur for the longer wavelengths causing the light to be lost.

2.5 Detectors

For FT-IR micro-spectrometry, detectors which have high speed, good sensitivity and linearity are needed. Detectors used in FT-IR spectrometers operating in the MIR spectral range can be categorised as either thermal or quantum. Thermal detectors work through sensing the heating effects of the IR radiation across the whole of the infrared electromagnetic range. One such thermal detector is known as the Deuterated Triglycerine Sulfate (DTGS) detector. Quantum detectors are photoconductive and are a class of semiconductor. On interaction with emitted infrared radiation, valence electrons contained within the detector material are excited from a valence band and promoted to a higher energy state. If electrons can be promoted from a valence band to a conduction band then the material will then pass a current. One such quantum detector is the Mercury Cadmium Telluride (MCT) detector. MCT detectors have greater sensitivity than DTGS detectors (specific detectivity (D*))). Typically MCT detectors are cooled to 75-77k using liquid nitrogen (low temperatures stop random thermal valence electrons being excited by random thermal agitation, which gives rise to noise in the detector output) and because of their sensitivity are favoured of DTGS detectors for FT-IR micro-spectroscopy \(^{52-53, 72}\). An MCT detector will be used in this study.
2.5.1 Focal Plane Array (FPA) Detectors and Hyperspectral Images

Good spatial resolution can be achieved through reducing the light that reaches the single element detector present in a FT-IR micro-spectrometer. This method of recording data allows an average spectrum to be recorded from a single cell. However, many single cell spectra may need to be recorded for adequate statistical analysis. With a single element detector, to record IR absorption spectra of single cells requires sequential mapping; recording single cell spectra sequentially can be time consuming. A quicker method is to record a chemical image with a wide field of view, capturing data from many single cells in one image frame. This allows spectra to be recorded from hundreds of individual cells simultaneously. To record a chemical image when using FT-IR micro-spectroscopy a Focal Plane Array (FPA) detector is used instead of a single element detector.

A FPA detector contains many individual pixel detector elements, resulting in thousands of spectra being recorded simultaneously (Fig. 21). These recorded images are considered as hyperspectral. The best FPA detectors on the market are square MCT arrays containing 128 x 128 pixel detectors; each pixel being considered as a single detector element recording a single infrared spectrum. For a 128 x 128 MCT FPA detector combined with a 15x optical IR objective, a chemical image can be recorded containing 16384 spectra, covering a 704 µm² spatial field of view and achieving a spatial resolution of 5.5 µm; spatial resolution being defined by the pixel size and the optical magnification. As all spectra can be recorded simultaneously, multiple co-scans can be recorded and averaged together to increase the SNR of the spectra recorded in the image.

It is clear than FPA spectral imaging has time advantages over single element detectors, however, FPA detectors also produce chemical images with subcellular spatial resolutions covering a large spatial area which has enabled chemical images to be compared with stained images. Typically, disease state is identified and studied from human tissues by highly trained pathologists and chemical staining is used to help see the cellular components when using optical microscopy. IR chemical imaging could therefore be used as an aid to investigate suspicious areas of tissues that are not clearly defined as diseased or healthy after chemical staining/optical microscopy examination. As a result, IR chemical imaging has had a lot of interest in the past ten years 73.
Fig. 21 (a) Example of the many pixel element detectors incorporated into a FPA detector. For a 128 x 128 pixel FPA detector coupled with a 15x IR lens objective, each pixel covers a 5.5 x 5.5 µm sampling area thus, covering an overall area of 704 µm$^2$. (b) Each recorded spectrum can then be classified using multivariate classification algorithms and classification maps can be compared to visible images of tissues, or chemically stained tissues.

Fig. 22 shows a total absorbance heat map from a spectral image of single cells. The total absorbance heat map represents the total area under each absorption spectrum. A total absorbance heat map makes it easy to see where the cells are located in the chemical image; blue spatial regions in the image represent areas of the sample with no cellular material thus, no absorption, whereas yellow/red spatial regions in the image represent areas with the greatest absorption intensity. This type of image may display good spatial information. However, the chemical information described from the image is limited. To display chemical information in the image, absorbance peak ratios, heights and widths of interesting absorption bands can be displayed as an image. The output results from
multivariate classification algorithms can even be displayed as the output of a recorded image.

![Heat map image](image.png)

**Fig. 22** Total absorbance heat map image of TERA2.cl.SP12 cells distributed onto a CaF$_2$ IR substrate. Chemical or spectral image recorded on a 128 x 128 FPA MCT detector using an x15 optical microscope lens. This results in a 704 µm$^2$ image frame field of view from 16384 pixels and 5.5 µm$^2$ pixel size.

### 2.6 The Interaction of Electromagnetic Radiation with Matter and Resonant Mie scattering

As IR radiation comes into contact with a biological sample, light will be absorbed from the sample ($I_A$), transmitted through ($I_T$), reflected from the sample surface ($I_R$) and scattered ($I_S$) (Fig. 11 (a)). Scattered and reflected light results in some frequencies never reaching the detector.
The Beer-Lambert law is the fundamental law of quantitative spectroscopy. For a monochromatic frequency of IR radiation the transmittance is given by the power of the IR light emerging from the sample \( I(\tilde{\nu}) \) divided by the power of the IR light at the front face of the sample \( I_0(\tilde{\nu}) \) (found through recording a background spectrum) (Eq. 5); where \( \alpha(\tilde{\nu}) \) is the linear coefficient at wavenumber and \( b \) is the thickness of a pure sample at wavenumber (cm) \(^53\).

\[
T(\tilde{\nu}) = \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})} = \exp[-\alpha(\tilde{\nu})b] \tag{Eq. 5}
\]

A transmittance spectrum is created by taking into account all the frequencies passed through the sample when polychromatic light is emitted from the source. A transmission spectrum can then be turned into an absorbance spectrum through Eq. 6. The absorption of each frequency of IR light is given by the base 10 logarithm of \( \frac{1}{T(\tilde{\nu})} \). However, Eq. 6 does not consider the losses in light intensity through both reflection and scattering \(^53\) (Fig. 23 (b)).

\[
\log_{10} \frac{1}{T(\tilde{\nu})} = A(\tilde{\nu}) \tag{Eq. 6}
\]

Scattering of IR light occurs when IR light emitted by the source is redirected in all directions as it comes into contact with non-homogeneous scattering particles. The majority of light scattered by a scattering particle emits at the same incident frequency, known as elastic scattering. There are two types of elastic scattering, Rayleigh and Mie scattering. Rayleigh typically occurs when the scattering particle is small and spherical, much smaller in diameter than the wavelength of the incoming frequency of light; Rayleigh scattering is usually exhibited from gaseous compounds in the nanometre range \(^74\).
\[ I_0 = I_{\text{Absorbed}} + I_{\text{Reflected}} + I_{\text{Scattered}} + I_{\text{Transmitted}} \]

\[ I_0 = \text{Intensity of incident light from the source} \]

**Fig. 23** (a) The interaction of EMR with matter.  
(b) IR spectrum of a single cell with scattering contributions.  
(c) the interaction of IR radiation with a flat biological sample.  
(d) the interaction of IR radiation with a single eukaryotic cell.
Fig. 23 (b) shows a single cell IR spectrum recorded over the MIR spectral range containing scattering contributions; light scattering causing a broad oscillating baseline in the recorded spectrum. This type of scattering is known as Mie type scattering, which occurs when the wavelengths of incident IR radiation are similar in size, or substantially less than the biological sample or components which make up the sample. Infrared light in the MIR spectral range have wavelengths between 2.5 to 25µm in size. With the sizes of single cells typically ranging between 10 – 30µm then it is clear that Mie type scattering will be seen from some frequencies when they interact with single cells. Severely distorted baselines, like that seen in Fig. 23 (b), make discerning spectral differences from recorded IR cell spectra difficult. Fig. 23 (b) also displays a derivative like peak shape on the high wavenumber side of the amide I absorption band. This type of scattering was known as anomalous dispersion. This derivative dispersion has an effect the carbonyl absorption band, located ~1730 cm⁻¹, and the amide I band positions. This is not ideal as the amide I absorption band provides the greatest amount of information on protein secondary structure and has also been seen as a critical diagnostic marker. As proteins account for a large percentage of a single cell’s composition, adequate corrections need be made to recorded spectra in order to remove the scattering contributions.

In 1957, Van de Hulst published an equation that modelled Mie type scattering efficiency (Q) per wavelength (λ) (Eq. 7).

\[ Q = 2 - \left( \frac{4}{\rho} \right) \sin(\rho) + \left( \frac{4}{\rho^2} \right) [1 - \cos(\rho)] \]  \hspace{1cm} (Eq. 7 (a))

\[ \rho = 4\pi d(n - 1)\lambda \]  \hspace{1cm} (Eq. 7 (b))

\[ n = n_1 / n_2 \]  \hspace{1cm} (Eq. 7 (c))

In Eq. 7 Q represents the loss of light per wavelength due to the scattering; n being the ratio between the real refractive index of the scattering particle (n₁) and the surrounding medium (n₂) and d is the diameter of the scattering particle. Eq. 7 shows that the scattering efficiency is dependent on both the frequency (λ) and the diameter of the
scattering particle. However, Eq. 7 only describes the scattering efficiency of a non-absorbing dielectric sphere, which assumes a constant refractive index ratio for all frequencies of EMR radiation passing through \(^{75-77}\). As previously explained, biological samples are actually highly absorbing in the MIR frequency range.

The parameters which govern the absorption of radiation are the real and imaginary components of the complex refractive index; \(\tilde{n}(\tilde{\nu})\) being the real refractive index, \(k(\tilde{\nu})\) the imaginary and \(n(\tilde{\nu})\) refractive index \(^3\) (Eq. 8).

\[
\tilde{n}(\tilde{\nu}) = n(\tilde{\nu}) + i k(\tilde{\nu})
\]

(Eq. 8)

The Kramers-Kronig transform algorithm relates both the real and imaginary parts of the complex refractive index. The outputs from the Kramers-Kronig transform being the real refractive index spectrum, minus the average real refractive index; known as the \(n_{KK}\). Through plotting the \(n_{KK}\) spectrum of a compound (\(n_{KK}\) vs. wavenumber (cm\(^{-1}\))) it can be seen that for absorbing particles the real refractive index changes per wavenumber \(^{75}\). The non-resonant Mie type scattering equation (Eq. 7) previously assumed a constant value for the real refractive index. Therefore, rather than a constant value for \(n\), Bassan et al. \(^{75-77, 79}\) adapted Van de Hulst’s equation to incorporate the Kramers-Kronig transform spectrum (\(n_{KK}\)) of a non-scattering reference spectrum \(^{79}\).

Plotting \(Q\) vs. wavenumber for a constant value for \(n\) (original Mie theory for a non-absorbing particle) results in a \(Q\) curve which looks like the broad Mie type scattering. Therefore, Eq. 7 is sufficient enough to explain Mie type scattering. However, \(Q\) curves produced from the adaptation by Bassan and colleagues \(^{75-77, 79}\) showed that this new absorption resonance theory could explain both Mie type scattering and what was the previously described as the anomalous dispersion artefact. As a result of this research the term \textit{resonant} Mie scattering (RMieS) was created.
Fig. 24 (a) An IR spectrum recorded from a single cell containing scattering contributions in the spectrum (b) scattered spectrum corrected using the RMieS-EMSC algorithm (20 iterations used) and scattered spectrum corrected using the EMSC algorithm (c).
2.6.1 Extended Multiplicative Signal Correction algorithm (EMSC)
Kohler and colleagues were the first to attempt to try and remove the scattering contributions from absorption spectra containing scattering contributions and through their work the Extended Multiplicative Signal Correction (EMSC) algorithm was created. The EMSC correction algorithm uses a least squares fitting algorithm to remove the broad sinusoidal baselines produced as a result of Mie scattering. However, the algorithm does not attempt to solve the spectral distortions previously thought of as the anomalous dispersion artefact thus, they still exists after EMSC correction (Fig. 24 (c)).

2.6.2 Resonant Mie Scattering algorithm (RMieS-EMSC)
The RMieS-EMSC algorithm, created though a collaboration between Kohler, Gardner, Byrne and Bassan incorporated the RMieS scattering theory and the EMSC algorithm (Fig. 25). For the RMieS-EMSC correction algorithm, a non-scattering absorption spectrum (Z_{Ref}) is used as the reference spectrum. An ideal reference would be a scatter free spectrum of the same sample, however, as previously explained, due to the size and shape of a single cell, a scatter free spectrum of a single cell is difficult to obtain. One can use a scatter free spectrum recorded from another material and this has been effectively shown in a number of studies by Bassan et al. who have used a scatter free Matrigel spectrum to successfully correct RMieS scatter contributions in IR recorded spectra from cell and tissue samples; Matrigel being an artificial basement membrane consisting mainly of protein. A scatter free spectrum is acquired through recording a spectrum from a thin layered film Matrigel film.

For the algorithm, the Z_{Ref} is transformed using the Kramers-Kronig to produce a n_{KK} spectrum (real refractive index spectrum). Through varying the average real refractive index (typically 1.1 to 1.5) and the scattering particle radius, many potential scatter curves (Q) are produced, which are then decomposed into Principal Components (PCs) using PCA. In transforming Q curve information into PCs, the scattering curve information is summarised in a low number of PC variables, which explain the majority of possible Q curve scattering variance. Least squares regression is then performed to find the curve information which best fits the scatter contributions contained in the Z_{Raw} spectrum. The RMieS-EMSC algorithm then subtracts a curve that is the sum of a constant value offset (c), a sloping baseline (m\bar{v}) (see Fig. 25), and the new RMieS Q curve which best fits the
scattering contributions contained within the $Z_{\text{Raw}}$ spectrum recorded; this should leave a scatter free absorption spectrum as the output\textsuperscript{76-77} (Fig. 24 (b)).

If a non-ideal reference spectrum is used, i.e. Matrigel, then the corrected spectrum will have some components of the reference spectrum contained within the outputted spectrum. To improve this, the RMieS-EMSC was made iterative. Now, after the first iteration of the algorithm, the new corrected spectrum ($Z_{\text{corrected}}$) then becomes the new reference spectrum for the next iteration (Fig. 23); this process is repeated for all iteration loops set at the start. The theory behind this being that with more iterations of the algorithm the less it is likely that components of the Matrigel, or another non-ideal reference spectrum, will be found in the corrected pure absorption spectrum thus, producing a scatter free absorption spectrum more like the real pure absorption spectrum\textsuperscript{75}.

![Diagram](image.png)

Fig. 25 Processes involved in the RMieS-EMSC algorithm\textsuperscript{79}.
2.7 Pre-Processing Steps after RMieS-EMSC Correction

2.7.1 Noise Reduction
IR recorded spectra from single biological cells tend to have low absorption signals due to their small size and thickness. Therefore, an in house function was made in MATLAB to reduce the noise and improve the SNR of recorded cell spectra. The function firstly performs PCA on data set, transforming the data to PCs. The user can then retain a number of PCs explaining the relevant chemical information, while removing PCs with low eigenvalues and containing spectral noise. The retained PCs are then reconstructed to the original data size; this is done by matrix multiplication with the remaining PC scores vectors being multiplied by the remaining PC loadings vectors, which have been transposed. For this study, to err on the side of caution, the first 60-70 PCs were retained so as to make sure chemical information is lost, while the spectral noise contained in the remaining is removed from the data set thus, reduction in noise contributions in the outputted spectra. This function is beneficial when transforming the recorded cell spectra to derivatives as transforming spectral data to derivatives increases the noise influence.

2.7.2 Vector Normalisation
Spectra recorded from biological samples often show differences due to sample thickness. This is because the sample thickness of a sample is hard to control. According to the Beer Lambert law, the pathlength of a sample is proportional to the absorption. Therefore, there is a need to process each spectrum so that there is a commonality between each spectrum recorded. This can be done by adjusting each absorption peak in a spectrum to its own internal standard, which is known as normalisation. Through normalising recorded spectra, spectral differences can now be thought of as biochemical differences rather than greater absorptions due to sample thickness. One type of normalisation is known as vector normalisation. For vector normalisation, each wavenumber absorption variable of the spectrum is squared and each squared variable then summed to give one total overall spectrum result (T). Each wavenumber absorption variable is then divided by the square root of T, normalising each spectrum independently.

2.7.3 Spectral Transformation to Derivatives
An IR absorption spectrum of a biological sample has many overlapped absorption bands in the MIR spectrum, especially absorption bands found in the in the fingerprint spectral
range. Through transforming the original absorption spectrum to spectral derivatives, overlapped absorption bands can be resolved; giving the illusion of enhanced the spectral resolution. Resolving overlapped absorption bands can be very advantageous when trying to find small biochemical changes from recorded cell spectra. Another advantage when transforming absorption spectra to derivatives is the elimination of baseline shifts often caused by scattering, or by the subtraction of poor background spectrum.

Derivatives are ultimately the measure of the slope of the absorption bands in the original spectrum. When transforming an IR absorption spectrum to a first derivative spectrum, the absorption bands of the first derivative spectrum cross the x axis (wavenumber / cm\(^{-1}\)) at the same point as the original absorption bands peak maxima, and has both positive and negative maxima bands at the points of the maximum slopes from the original zero order absorption bands. In contrast, the absorption bands of the second derivative spectrum have a negative peak maximum at the same position (wavenumber / cm\(^{-1}\)) as the original absorption bands peak maxima. However, converting spectra to derivatives has some drawbacks. Firstly, when transforming to derivatives, spectral absorption information can be lost. Secondly, the band narrowing process comes at the cost of reducing S/N of the transformed derivative spectra, when compared to that of the original. Thirdly, when transforming to spectral derivatives, broader original absorption bands having poor S/N when compared to narrower absorption bands contained in the original spectrum. A consequence of this being that sharp features of spectral noise are enhanced with higher order derivatives \(^{84-85}\). A way of reducing the increases in spectral noise influence is to employ Savitzky-Golay smoothing. However, care also needs to be taken with the number of smoothing points used, too many can distort the spectra resulting in spectral information being lost.

2.8 Multivariate Spectral Analysis of Recorded IR Spectra

An MIR absorption spectrum of a sample has many wavenumber variable absorption values. Comparing spectra recorded from different biological samples involves looking for differences, or similarities, in the absorption intensities at each variable contained within a recorded spectrum. For biological samples, the differences in absorption intensities can be considered as biochemical differences because the absorption spectrum is directly related the sample structure. However, when one visually compares spectra recorded from
different biological samples, the differences between the recorded spectra may not be obvious. For example, one would expect spectra recorded from stem cells and their differentiated derivatives to be different due to differentiated cells becoming functionalised. However, on visual inspection of recorded spectra, the spectral differences between the different groups may not be obvious, with biochemical changes being subtle. An infrared absorption spectrum of a single cell may well be characteristic of the molecular structure, however, unlike mass spectroscopy, the technique is unable to explain information on all of the proteins involved in a cell; an IR absorption spectrum only being able to explain spectral information on the net biochemical structure of the cell. Spectral differences between a spectrum recorded from a stem cell and a differentiated cell spectrum may well constitute small spectral changes, i.e. changes in absorption band shape, frequency position and size, which explain the net biochemistry changes of the large macromolecules that formulate the cellular structure. These spectral variations may also be small when compared to the biochemical changes of a cell undergoing life cycle changes. All described factors make deciphering biochemical differences between cell spectra difficult. A further difficulty being the number of cell spectra recorded; typical spectral investigations of single cells involving the comparison of hundreds of single cells spectra, with some studies comparing thousands. For a human to do this would be incredible difficult and time consuming. Therefore, multivariate analysis can be used as a way to quickly sieve out the relevant spectral variations from the recorded cell spectra. One such method is known as Principal Component Analysis (PCA)

2.8.1 Principal Component Analysis (PCA)

PCA is an unsupervised multivariate algorithm which can be used to find spectral differences between IR spectra recorded from different biological samples. Being unsupervised means that the algorithm has no information on the group classification of the IR spectra recorded, therefore, treating each IR absorption spectrum separately.

Before PCA can be performed the original data matrix should firstly be mean centred. By mean centring the data, the new data matrix will have absorption intensities at each variable spread about zero. This can be useful if some intensities variables have very different means and can be used as a way of normalising the variance. Mean centring is achieved by subtracting each variable by the overall variable mean from the original data set. The new mean centred data matrix \(X\) has the dimensions \(l \times j\) \((l\) being the number
of samples and $J$ the number of wavenumber variables in the data matrix) (Fig. 26). When $X$ is inputted into PCA the data is transformed into new matrices known as the scores ($T$) (has the same number of rows and columns as samples and variables from the original data set), loadings ($P$) (square matrix which has the same number of rows and columns as there are variables in the original data set) and the error matrix ($E$) (Fig. 26). This transformation constructs new variables contained within the new modelled matrices $T$ and $P$ called principal components (PCs). These PCs are linear boundaries capturing as much variability from the original data matrix as possible (each column in both $T$ and $P$ being a different PC). Using the scores matrix ($T$), each PC boundary can be visualised graphically (2d or 3d) with the distance between each spectrum and the PC boundary centre being the score for each spectrum. By plotting the scores from each PC, one can see how much each spectrum relates to the variance explained within the PCs, and in doing so, whether the spectral variance explained in one of the PCs is able to discriminate between the spectra recorded from the different samples. If the spectral variance contained in a PC does discriminate between different groups of recorded absorption spectra, then the question is why? Plotting the loadings of the discriminating PCs shows how much weight each wavenumber variable has on the spectral variance contained within the PC. Therefore, enabling the understanding of the biochemical differences between different groups of IR spectra recorded. The error matrix ($E$) represents the error between the original mean centred data ($X$) and the modelled ($T \times P$). Errors can be due to instrumental noise, or measurement error$^{82-83}$.

In Fig. 27, a simple data set with many IR spectra (1881) but just two wavenumber variables are described. In Fig. 27 (a), the data has been mean centred and then plotted. However, it can be seen that if a line was plotted through the zero positions of both the 1550 and 1650 cm$^{-1}$ axis, the most variability within the data set will not be captured. By performing PCA on the data, the variable axes are rotated. Now, in one direction the maximum amount of variance or the largest spread will be captured (PC1). The next rotation (PC2) will then be orthogonal (at 90° to the first PC) to the last, capturing the second largest amount variance from the original mean centred data set, Fig. 27 (b and c). For data which has greater dimensions than two, the process will be the same with each new PC being at right angles to all of the preceding, until n PCs have been identified ($n$ being the number of wavenumber variables).
Principal Component Analysis (PCA)

\[ X = TP^T + E \]

Samples \((I)\)  
Variables \((J)\)  

Loadings \((P)\)  
Principal Components \((A)\)  
Scores \((T)\)  
Principal Components \((A)\)  

\((A) = J\)

Represents the scores and loading vectors for the first principal component (PC1)

Fig. 26 PCA breakdown.
Fig. 27 The effect of mean centring the data and how PCA rotates the two dimensional axis of the mean centred data to find new linear boundaries known as principal components, which explain the greatest amount of variability in the data set (a) scatter plot of the 1881 IR spectra and their absorption values at 1550 and 1650 cm\(^{-1}\) (b) scatter plot of the same data but mean centred (c) a scores plot of the 1\(^{st}\) and second PC and the new PC boundaries.
Fig. 28 (a) Two different classes of simulated spectra; Class 1 (red) and Class 2 (black). Clear spectral differences can be seen at two different wavenumber absorptions 1468 cm\(^{-1}\) and 1740 cm\(^{-1}\) (b) PCA scores plot of cell spectra from both Class 1 and Class 2 (c) plot of the PC2 loading vs. Wavenumbers / cm\(^{-1}\).
In Fig. 28 (a), synthetic IR spectra have been produced, 10 for class 1 (red) and 10 for class 2 (black). Visually it is obvious that there are spectral differences between the two classes at absorption bands located at 1648 and 1740 cm\(^{-1}\). In performing PCA on the synthetic data, the principle component scores can be visualised by displaying a two dimensional scores plot of the only two PCs (as there are only two variables there will only be two PCs). In Fig. 28 (b), it can be seen that 95.8% of the spectral variance contained within the original data set is contained within PC1. However, none of the variance contained within PC1 explains any differences between the two classes of data. The spectral variance which explains the differences between the two classes is found in PC2, with good separation along the PC2 axis.

As with the \( T \) matrix, the loadings matrix \( (P) \) has PC eigenvectors which are found in the columns of the matrix. The PC loading peak at a wavenumber variable is the cosine of the angle between the original variable and the new PC direction \(^{83}\). By plotting the PC loadings vs. the wavenumber variables, the wavenumber absorptions which contribute to the spectral absorption variance contained within the PC can be displayed. In displaying the PC2 loading plot in Fig. 28 (c), it is clear that the differences between the two classes come from the absorption bands 1468 and 1740 cm\(^{-1}\).

It is important to note here that the sign of loading bands have no meaning as the sign is invertible \(^{83}\). However, the magnitudes of the loadings peaks do have relevance and show which wavenumbers contribute more to the separation along the PC axis. By using PCA the relationship each spectrum has to the next can be seen. The relevant spectral variance contained within the data set will therefore be simplified and reduced to much more manageable numbers. This results in PCA being excellent for data reduction \(^{86}\). 

### 2.8.2 Supervised Multivariate Analysis

PCA results are a combination of both interclass (between class) and intraclass (within class) variability; this is because PCA is an unsupervised algorithm and is unable to decipher between the two. This can be a problem when analysing IR spectra recorded from biological samples as there tends to be a large amount of heterogeneity within each sample class. This is especially true when recording spectra of single cells due to the large biochemical changes of during its life cycle \(^{86}\).
Linear discriminant analysis (LDA) is a supervised multivariate algorithm which maximises the interclass variability while minimising the intraclass. LDA requires prior knowledge of the group classes beforehand and then finds the best linear hyperplane separation boundary between the class groups of recorded IR cell spectra. This is demonstrated in Fig. 29.

**Fig. 29** (a) Two different class groups of IR spectra are displayed in a scores plot containing absorption information for two variables (b) new linear hyperplane found that best explains the differences between the two classes of spectra (c) example of how each data points score is found on the new hyperplane variable (Z)
In Fig. 29 (a), two different class groups of IR spectra are displayed in a scores plot containing absorption information for two variables. It can be seen in Fig. 29 (a) that there is no complete separation of the two classes of IR spectra through either the X, or Y dimensions. However, when using LDA, the axes are rotated in order to find a new hyperplane dimension (Z) which shows the best linear separation between the two classes (Fig. 29 (b)). Through the new Z projection, good separation between the two class groups of IR spectra can be seen. Spectral loadings produced by the LDA algorithm will now show the wavenumber variables that contribute to spectral differences along the new Z axis.

**2.9 Cross-Validation**

The aim of creating classification models is to predict the classification of new data. Therefore, the question is how good are the created classification models at predicted new data? A way of answering this question is through cross-validation.

For cross-validation, the data is typically partitioned into two, one partition used to create the classification model (training data set) and the other (test data set; previously unseen by the created classification model) is used to test the performance of the created classification model. The performance of the created supervised classification model is judged on how well the created model predicts the new and previously unseen test set data. The percent of test data which are correctly classified (%CC) when projected into the classification model can be used as a way of judging the performance of a classification model.

There a number of different methods used to partition the data into training and test data sets. Hold-out cross-validation partitions the data set into two (test and training), with two thirds of the data being randomly selected as the training set, used to build the classification model, while the remaining third of the data set will be used to test the performance of the model created. Hold-out cross-validation is simple and fast, however, one of the negatives of hold-out cross-validation is that not all of the spectra in the data set has the chance of being used in both the training and test set once. With every spectrum not getting a chance of being used in both the train/test data sets there is a chance there may be
some bias in the cross-validation performance, i.e. the spectra partitioned into both training and test data sets may produce a good classification model performance, however, the result may not be reproducible if the data were to be partitioned again and cross-validation was performed. Therefore, hold-out cross-validation is completely dependable on what data is being split into either the train, or test sets. Through repeating hold-out cross-validation multiple times and recording the average validation performance result may go some way to solving the problem, however, there are other cross validation methods which utilise all of the data.\textsuperscript{88}

Leave one out cross-validation (LOOCV) uses every data point as the test set once. For example, for LOOCV the first data point in the data set is removed as the test spectrum while the remaining data points are used as the training set to make the classification model. This is repeated until all data points have been used as the test set once. The advantage of LOOCV being that every data point is used in the validation process and the performance result being the average validation performance result for the classification model. However, for large data sets, LOOCV can be computationally heavy and can also have high variance in the performance results.\textsuperscript{88}

K-fold cross-validation can be seen as a compromise between both hold out and LOOCV. For K-fold cross-validation, the data set can be split into, K, equally sized groups, with no data point being represented in any more than one group (Fig. 30). For example, if 5 fold cross validation was to be performed on the data, the data assigned to the first group will be used as the test set, while the remaining 4 will be used as the training set. Like LOOCV, the K-fold cross-validation will be repeated, being performed K times. Therefore, every data point is used in test set data partition once. K-fold cross-validation is very similar to LOOCV but with the advantage of not being as computationally intensive and as a result, faster. This makes it more favourable when validating large spectral data sets. Commonly 10-20 folds should be used for K-fold cross-validation as the number of data partitions has been show to produce the most stable cross-validation performance results.\textsuperscript{83, 88-90} However, there is another method which can evaluate the performance of a classification model, used as an alternative to cross validation. This method is known as bootstrapping. Bootstrapping is a method that employs sampling with replacement. The training set (bootstrap) is the same size as the original data set but with some spectra being duplicated, the number of distinct spectra in the training set is approximately two thirds of the original
dataset. The remaining spectra not used in the training set are then used as the test set. The test set, previously unseen, will then be projected into the training set model and the performance of the model evaluated \(^{82-83}\).

<table>
<thead>
<tr>
<th>1st Fold</th>
<th>2nd</th>
<th>3rd</th>
<th>4th Fold</th>
<th>5th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Training set**: Training data folds involved in making the classification model

**Test set**: The remaining data left out of the making of the classification model

*Fig. 30* Visual interpretation of how the data set is partitioned when using 5-fold cross-validation \(^{29}\).

### 2.9.1 Bootstrapping or K-Fold Cross-Validation

To test whether 10-fold or bootstrapping cross-validation provides the most reproducible cross-validation of IR spectral data, three different classes of randomly selected IR recorded spectra (100 spectra per class) were tested; LDA was used as the supervised classifying algorithm and the first 4 PCs were inputted. The 10-fold cross validation
produced an average %CC result of 92.6% and standard deviation (SD)/relative standard deviation (RSD) of 3.98 and 4.3 respectively (Fig. 31). The bootstrapping method, with 2000 bootstraps producing an average %CC of 93.5% and a SD and RSD of 2.32 and 2.49 respectively. Therefore, although bootstrapping cross-validation has more bias than 10-fold cross validation (due to sampling with replacement), bootstrapping cross-validation provides the best accuracy when evaluating the performance of a LDA classification model.

![Fig. 31 An LDA scores plot from the 10-fold cross-validation of 3 different classes of IR spectra (100 random selected IR spectra for each group). Training data is symbolised with a circle and test data is symbolised using a star. Also, 95 % confidence ellipses are displayed for each data group. Average % CC was 92.6%.

2.9.2 Predicted Residual Error Sum of Squares (PRESS)

PC-LDA can over-fit when too many are incorporated into the LDA classification algorithm; a high number of PCs incorporated into the algorithm will mean PCs with low eigenvalues are used, possibly containing spectral noise \(^{31}\). However, what is the correct number of PCs that should be used as spectral information to construct a supervised classification model like LDA? Unfortunately, there is no strict rule in for the selection of PCs to be used in PC-LDA. Previous studies have tried to suggest ways and rules to select the correct number of PCs, such as the Cumulative Percent Variance (plot of percent variance vs. PCs. PCs that explain 90-95 % variance are typically used) and Scree Test on Residual Percent Variance (RPV) (plot of RPV vs. PCs. Method looks for the knee point in the scree plot) \(^{90}\). However, these methods are subjective. An alternative method comes...
from a processing method known as Predicted Residual Error Sum of Squares (PRESS). The PRESS test attempts to come up with a method for PC selection in a non-subjective manner.

PRESS employs LOOCV to remove each spectrum from the data set while PCA is performed on the rest. The removed spectrum is then predicted based on the PC1 loading vector of the training set (the process is repeated for all data points). The error is the sum of squares difference between the true spectrum and the predicted. The process is then repeated, predicting the removed spectrum with both PC1 and PC2 loading vectors of the training set; this process continues until a set number of PCS selected have been evaluated. The PRESS error is then compared to the Residual Sum of Squares error (RSS) for each spectrum in the data set. The results are visualised by plotting a graph of the PRESS \((a+1)/\text{RSS}(a)\) (a = number of PCs being modelled) ratio vs. the number of PCs. If the ratio exceeds 1 then the PC which exceeds 1 contains noise \(^{82,90} \).
Chapter 3: Previous Studies using FT-IR Micro-Spectroscopy (FT-IRMS) to Investigate Stem Cell Differentiation

Over-recent years, the number of studies using FT-IR microspectroscopy to investigate stem cell differentiation has vastly increased due to the fact that the technique has great potential for recording spectral absorption fingerprints of cells, which can then be used to classify their phenotype. The benefit being that it can be done without the addition of potentially harmful biomarkers to cell culture, in a non-invasive manner; something which current industry standard techniques used to identify stem cells are unable to do.

A number of studies under the supervision Heraud and have used FT-IR microspectroscopy to look at stem cell differentiation, with a great variety of cell types investigated \(^91-96\). A study by Heraud et al.\(^92\) published in 2010, used FT-IRMS to see whether spectra recorded from undifferentiated hESCs and cells committed down mesodermal and ectodermal lineages could be distinguished (cells were left to differentiate for 4-5 days). Cells were directed to differentiate towards mesendoderm in serum-free medium supplemented with bone morphogenetic protein (BMP) 4 and activin A, and toward ectodermal lineages in medium supplemented with fibroblast growth factor (FGF) 2. Cells were cytospun spun onto MirrIR reflective slides for reflectance mode spectroscopy and average spectra were recorded of cells which had been air dried. The study results showed that all three classes of cell spectra could be classified using PLS-DA. Key spectral differences were seen from the absorption bands related to lipids and glycogen; undifferentiated cells having a greater lipid and glycogen content than differentiating.

Tanthanuch et al.\(^91\) looked at the spectral differences between Embryonic Stem-Derived Neural Cells (ESNCs (produced from mESCs)) and undifferentiated mESCs. Results showed that the main spectral differences were increases in lipid and \(\alpha\)-helix protein content, while a decrease in net \(\beta\)-sheet conformational proteins was seen from the differentiated progeny (ESNCs) when their spectra was compared against the spectra recorded from mESCs. A study by Zhao et al.\(^97\) also observed an increase in \(\alpha\)-helix protein absorption from differentiating HL60 cells; induced to differentiate by ATRA to producing mature granulocytes. A study by Thumanu et al.\(^95\) showed that spectral lipid intensities increased initially when hepatic progenitor cells differentiated, however, the intensities significantly decreased upon the final stage of differentiation, producing
hepatocyte-like cells. This may be an indication that lipid structural molecules are actively involved in the differentiation of embryonic derived hepatocytes. To produce hepatocyst-like cells, mESCs were left to form embryoid bodies for 14 days, then removed and cultured with basic fibroblast growth factor (bFGF) for two days to generate definitive endoderm cells. After another two days of bFGF treatment, cells were treated with hepatocyte growth factor (HGF) for three days to produce progenitor cells. A further culture step was then needed to produce mature hepatocyte cells from mESCs. The results from the study also showed an increase in α-helix content from mature hepatocyte-like cells when compared to the spectra recorded from progenitor cells. As well as this salient absorption feature, higher relative absorption levels of β-sheet structural proteins were also seen to be a discriminant absorption change when comparing the spectra recorded from progenitor cells to spectra recorded from mature hepatocyte-like cells; progenitor cells having increased absorption from absorption bands assigned β-sheet structural vibrations when compared to mature hepatocyte-like cells. This is in agreement with the structural changes of undifferentiated cells seen in the study by Tanthanuch et al. when investigating the pluripotent differentiation of mESCs. Therefore, these protein structural changes could be a possible mechanism/biomarker of the differentiation of mESCs.

A study by Ye et al. investigated the hepatocyte cells produced from rat bone marrow mesenchymal stem cells (rBM-MSCs). The main macromolecular changes of the cells during the differentiation process came from the lipids, with an absorption increase being seen from an absorption band associated with unsaturated fatty acid levels; cells at the later stages of hepatocyte differentiation showing increased absorption from the C=C vibration at 3010 cm\(^{-1}\). However, other macromolecular structural changes from differentiating cells from bands associated with nucleic acids and proteins were complex therefore, difficult to interpret.

Ami and co-workers looked at the differences in spectra recorded from mESCs and spectra recorded from cells at different stages of differentiation. The main spectral differences described from their study were seen in the nucleic acid (1000-860 cm\(^{-1}\)) and protein spectral ranges (1600-1800 cm\(^{-1}\)). Results showed that as cells differentiate the presence of absorption bands associated with the ribose ring (~914 cm\(^{-1}\)) and the deoxyribose ring (~899 cm\(^{-1}\)) appear in the mean spectra (after 4-7 days); the authors hypothesise that the presence of these bands indicates the switching of cell phenotype to the new differentiated phenotype. Average protein structural changes of cell spectra recorded from differentiating cells between 4-7 days showed a clear increase in α-helix.
protein and the presence of β-turn structures, around 1682 cm\(^{-1}\), suggesting a protein change to proteins typical of cardiomyocyte pre-cursors was taking place; authors state that it is known that these cells are rich alpha-myosin, a protein with α-helical structural formations thus, supporting the spectral changes witnessed as a result of differentiation. As well as this, undifferentiated cells could be distinguished from their differentiation progeny by their increased absorption from absorption bands typically associated with RNA molecules, this is in agreement with work by Notingher et al. \(^9\) (Raman spectroscopy), suggesting the involvement of RNA in protein synthesis, possible through messenger RNA (mRNA). DNA conformational changes were also seen from a study performed by Hughes et al. \(^9\) when investigating renal epithelial cell carcinoma side population cells displaying stem cell like characteristics. Results showed that vibrations from lipid and phosphodiesters to be particularly discriminating between spectra recorded from sub-populations of renal carcinoma cells (stem like cells hypothesised to produce the cancer), when compared to the more differentiated, proliferating cells, which make up the majority of the cell population.

Another study which resulted in bands primarily associated with nucleic acids being the main marker of differentiation came from a study by Walsh et al. \(^1\) In this study the symmetric phosphate vibrational band (PO\(_2\) \(\sim\)1080 cm\(^{-1}\)) was shown to be a significant marker of putative stem cells in human intestinal crypts, with changes to the 1080 cm\(^{-1}\) assumed to be possible DNA conformational changes in chromatin. Stem cells showing increased intensity from the 1080 cm\(^{-1}\) absorption band when compared to spectra recorded from differentiated and transit amplifying cells found along the paraffin-embedded human intestinal crypts. Authors state that the 1080 cm\(^{-1}\) marker proved to be more robust in the characterisation of the stem cells in the crypt than tissue stained with two different immunophenotypical markers, a significant advancement in the field as gastrointestinal stem cells lack specific biomarkers. Previous to the publication by Walsh and his colleagues, FT-IR microspectroscopy was successfully used by both German and Bentley et al. to investigate the spectral differences between corneal stem cells, their transit amplifying derivatives, and terminally differentiated cells. Spectral analysis results from both studies showed agreement that the greatest spectral differences between the different types of cells were due to absorptions bands primarily associated with nucleic acids \(^1\)\(^1\)\(^1\). In a recent investigation by the same group, the spectral signatures of stem cells, transit amplifying (TA) and terminally differentiated (TD) cells, found in human corneal epithelium, were distinguished using PCA. Spectral analysis results showed the DNA/RNA band at \(~\)1080 cm\(^{-1}\) to be a significant biomarker of stem cells, along with the phosphate
vibration at 1225 cm\(^{-1}\). Prominent phospholipid (1728 cm\(^{-1}\)) and protein changes (1558 cm\(^{-1}\)) were the significant biochemical differences between the TA and TD cells. The changes in nucleic acid associated with differentiated cells very much in agreement with previous work looking into putative stem cells and their differentiated derivatives\(^{103}\).

It can be seen in the literature that as of yet no absorption band can universally be described as a biomarker of stem cell differentiation. However, the studies looking into stem cells located in tissue niches seem to have found a reliable and reproducible biomarker of differentiation in the 1080 cm\(^{-1}\) DNA/RNA absorption band. With studies investigating the biochemical changes of embryonic stem cells though the 1080 cm\(^{-1}\) absorption band has not been seen as a specific marker of differentiation, with lipids and proteins being the main biochemical changes due to differentiation. This may be down to the differences in function between putative stem cells found in tissue crypts. Nevertheless, the identification of both stem cell and differentiated progeny through combining IR spectroscopy with supervised multivariate classification algorithms has been shown to be possible.
Chapter 4:

The Action of all-trans-Retinoic Acid (ATRA) and Synthetic Retinoid Analogues (EC19, EC23 and AH61) on Human Pluripotent Stem Cells; Differentiation Investigated using Single Cell IRMS

Publications Including Research Performed in this Chapter:


4.1 Abstract

The retinoids, ATRA, EC23 and AH61, all induce the differentiation of TERA2.cl.SP12 cells to produce neuronal cell phenotypes. However, even though ATRA and EC23 induce neuronal differentiation, differences in the differentiation ability of the different retinoids have been seen at the same concentrations; EC23 inducing larger numbers of neural cells with far less variability. Like EC23, AH61 has also been shown to have greater differentiation ability than ATRA, and EC19 produces mainly epithelial cell phenotypes through differentiation. Analytical methods previously used to monitor the change in the cell phenotype of this embryonic cell model have included immunofluorescence, flow cytometry, MALDI-TOF MS and reverse transcription polymerase chain reaction (RT-PCR) 4-7. However, all previous methods have their drawbacks when monitoring cell phenotype, with some analytical methods requiring the destruction of the cell being analysed and the addition of exogenous biomarkers to cell surfaces being required for analysis (see Chapter 1). This has led the belief that a non-destructive analytical method that has no or minimal impact on the cell biology is required for monitoring the status of the living stem cells and their different differentiation phenotypes. This is especially
necessary when monitoring the phenotype of living cells to be used for regenerative therapy.

IR spectroscopy has the potential for monitoring cell phenotype without the destruction of the cell or the need for the addition of fluorescent/magnetic biomarkers to cell cultures. Therefore, as a proof of concept study, the use of IRMS for identifying formalin fixed pluripotent TERA2.cl.SP12 stem cells from their differentiated progeny will be investigated in this chapter; cell populations induced to differentiate through the addition of ATRA, EC23, EC19 and AH61 to cell cultures. The study shows that IRMS coupled with appropriate scatter correction (RMieS-EMSC (see Chapter 2)) and multivariate analysis can be used as an effective tool to further investigate the differentiation of human pluripotent stem cells and monitor the alternative affects different retinoid compounds have on the induction of differentiation. IRMS detected differences between cell populations as early as 3 days of compound treatment. Populations of cells treated with different retinoid compounds could easily be distinguished from one another during the early stages of cell differentiation. As well as this, spectral analysis results showed that even though both ATRA and EC23 produce neuronal phenotypes at the 7 day period, biochemical differences can be seen between the cells treated with the different retinoids; in agreement with previous studies ⁵⁻⁷. These data results demonstrate that IRMS technology can be used as a sensitive screening technique to monitor the status of the stem cell phenotype and progression of differentiation along alternative pathways in response to different compounds, therefore, highlighting the potential of the technique for monitoring live cell phenotypes.

4.2 Cell Sample Preparations

All cell samples for this study were prepared by A. P. Henderson at Durham University.

4.2.1 Protocol for the Cell Culture of TERA2.cl.SP12 Cells

TERA2.cl.SP12 were cultured in a medium containing Dulbecco’s Modified Eagle Medium (DMEM), 10% (v/v) heat treated Foetal Bovine Serum (FBS (Gibco)), 0.22% (v/v) of penicillin/streptomycin ((Gibco) 1.1 ml penicillin/streptomycin in 500 ml of media) and 1% of 200 mM L-glutamine (Gibco).

Retinoids are naturally solid compounds. Therefore, there was a need to dissolve the compounds in order to deliver the drugs at the right concentration to the aqueous culture media. In order to achieve this, the retinoid stock solutions were prepared in the organic
For the study, TERA2.cl.SP12 cells were cultured in a humidified atmosphere of 5% CO\(_2\) in air and at 37 °C in a Sanyo CO\(_2\) incubator. The TERA2.cl.SP12 cells were brought up from frozen and transferred into a T-25 flask (BD Falcon) with 10 ml of culture media. The cells were then left in the T-25 flask for 2 days until confluent (Fig. 32). Once confluent, cells were passaged and put into a T-75 flask (BD Falcon) with 20 ml of culture media. For passaging, 3 mm acid-washed glass beads were used; acid washed beads were rolled over the adherent cells removing them from the surface of the flasks. Cells were then left for another 2-3 days in the T-75 flask to become confluent. The number of cells present in a confluent T-75 was approximately 24 million. This number was found through producing a single cell suspension of the cell population and performing cell counting. Single cell suspensions were produced by treating cell populations with 1 ml 0.25% trypsin/EDTA (Cambrex) solution for 2-3 minutes; this removed the cells from culture flask surfaces.

In the study, the TERA2.cl.SP12 cells were treated with each different retinoid described (ATRA, AH61, EC23 and EC19) and DMSO (DMSO is needed as a vehicle to dissolve the retinoids). As well as this, un-treated TERA2.cl.SP12 cells were also spectroscopically investigated. This provided 6 different TERA2.cl.SP12 cell treatments for each time point to be studied (14, 7, 5 and 3 days); time points being the time after the initial addition of retinoids to cell culture flasks. Each cell sample was produced in triplicate so as to validate the IR spectroscopy result and to show that the culturing method was reproducible; each cell sample had three biological replicates with each replicate being produced in an individual T-75 flask. This resulted in 63 individual T-75 flasks in total for the study. To
produce cells for the 63 flasks needed, the confluent T-75 flask containing approximately 24 million cells was split 63 times into individual T-75 flasks, resulting in approximately 380,000 cells per flask. After the transferred cells had settled down and attached to the surface of the flasks, 20 µL of 10 mM retinoid stock solutions were added to 20 ml of culture media. This results in a final concentration of 10 µM for the retinoids and DMSO being added to culture media. All cultures were handled in reduced light conditions to account for the instability (photosensitivity) of ATRA.

EC23, AH61 and ATRA differentiate TERA2.cl.SP12 cells to produce neuronal cell types. Fig. 33 (a) shows undifferentiated TERA2.cl.SP12 cells. After the introduction of the retinoids to cell populations, cells underwent differentiation and their differentiated neuronal progeny had a different structural shape, like seen in Fig. 33 (b). This differentiated cell progeny now show a more spread out shape on the surface of the flask with a complex network of synapses connecting neuronal cells to one another; these synapses allow signals to be passed on to other cells. This structural change was also seen from retinoid treated cell populations when preparing cell samples for experiments in this thesis.

![Figure 33 (a) and (b)](image)

**Fig. 33 (a)** TERA2.cl.SP12 cells at low confluency; Scale bar: 100µm. **(b)** Phase micrographs of cells exposed to ATRA and EC23.

### 4.2.2 Cell Fixation

In order to perform IRMS studies on single cells appropriate sample preparation should be applied to preserve the biochemistry of the cells. Chemical fixatives can be used to preserve the biochemical structure of the cells by ultimately freezing them in time. In doing this, the biochemistry of the cells at the time of fixative addition is captured. Without fixation,
biological cells will denature as they dehydrate, the dehydration of the cells will lead to the collapse (autolysis) of internal macromolecular structures (proteins, lipids and carbohydrates) \(^{105}\). Therefore, chemical fixation enables different types of cell to be distinguished based on their spectral thus, chemical signals, and not because of biochemical differences due to cell deterioration.

Typical chemical fixatives used to preserve the chemistry of biological cell samples include ethanol (ethanol 70%) and formalin (4%) \(^{105-107}\). Air drying has also been used as a method to preserve the biochemistry of single eukaryotic cells. However, cells are naturally hydrated with water molecules bound to the macromolecules that make up the cell structure. The air drying affect can alter the osmotic pressure of the cells resulting in cell shrinkage/swelling, which can then result in the bursting of the cell membrane. As well as this, air drying can also induce autolytic processes from within the cells resulting in the breakdown and denaturing of phospholipids, nucleotides and proteins. Therefore, air-drying is generally not considered as a good process for preserving the biochemistry of eukaryotic cells. As a result, chemical fixatives have typically been favoured \(^{107}\).

A number of studies have investigated the chemical effects of typically used chemical fixatives on single eukaryotic cells using vibrational spectroscopy. The general consensus from the results the studies show formalin to be the best chemical fixative to preserve the biochemical integrity of a eukaryotic cell; formalin fixation producing a cellular spectrum which is the closest to that of the living cell spectrum \(^{105, 107-109}\). Formalin is a mixture of paraformaldehyde (approximately 37-40 %) and phosphate buffer saline solution (PBS). Commonly a concentration of 4% formalin solution is used as a fixative. The addition of formalin causes the cross linking of primary, secondary and tertiary structural amines of proteins molecules. The cross linkage is produced through the formation of a methylene bridge between proteins, denaturing cellular proteins, shrinking the sample volume and locking in other macromolecular structures, such as DNA, lipids and carbohydrates etc. A methylene bridge is created when the aldehyde functional group of aqueous formalin react with the N-H functional groups of amides, creating a CH\(_2\) methylene bridge between the nitrogen atoms of proteins \(^{105, 109-110}\).
4.2.2 Cytospinning

To record spectra from single cells using IRMS in transmission mode, cells must be placed on the surface of an IR substrate. For this study, the IR substrate to be used is Calcium Fluoride (CaF$_2$). These CaF$_2$ substrates to be used are 13 mm in diameter, circular in shape and 1 mm in thickness. The substrates are transparent to wavelength frequencies of IR in the MIR range and transparent to visible white light allowing cells to be visualised through an optical microscope. A method known as cytospinning was used to distribute the formalin fixed cells onto the CaF$_2$ windows surface.

For cytospinning, all sample and control culture solutions were firstly agitated to provide a homogeneous solution. After agitation, 100 µl aliquots of all culture solutions were then dispensed into individual labelled cytospinning cuvettes. The loaded cuvettes were spun at 950 rpm for five minutes firing cells onto the surface of CaF$_2$ substrate windows. The small CaF$_2$ sample slides were then left to dry overnight. After sufficient drying time had passed, each sample slide was dipped carefully into double distilled water to remove any residual phosphate buffer salts (PBS) and chemical fixative residues and again left to air dry once more.

4.2.3 FT-IRMS Instrument used and Experimental Setup

The infrared data were recorded on a Varian 670 FTIR spectrometer interfaced with a Varian-620 imaging infrared microscope. The microscope was equipped with a 128 × 128 liquid nitrogen cooled MCT focal plane array detector (FPA) with a pixel effective size on the sample of 5.5 µm. The infrared spectral images were collected in transmission mode (512 co-added scans at 4 cm$^{-1}$ resolution) and the background image was recorded from a clean CaF$_2$ slide. The background spectral images were recorded with 512 co-added scans at 4 cm$^{-1}$ resolution.

4.3 Flow Cytometry Analysis

To ascertain whether stem cell differentiation had taken place in the sample cultures treated with retinoid compounds and whether control cell populations (untreated control and DMSO treated) had remained pluripotent, flow cytometry analysis was used; flow cytometry analysis was performed by A. P. Henderson at Durham University$^8$. Flow
cytometry uses known antibodies to bind to certain cell surface markers indicative of the status of the cell phenotype. The stage-specific antigen 3 antibody (SSEA-3) recognizes specific cell surface antigens associated with globo-series glycolipids and glycoproteins. Along with SSEA-4, TRA-1-60 and TRA1-81, SSEA-3, is expressed highly in embryonic stem cells and down regulated during differentiation. The SSEA-3 marker has been shown to be an excellent in understanding the phenotype of EC cells. Antibodies for differentiation have also been produced with the A2B5 antibody (ganglioseries antigen marking early-stage neural cells) being a proven marker for the early stage production of neuronal cells during differentiation. Flow cytometric analysis was only performed on cell samples that had been left in experimental cultures for 7 days.

Fig. 34 Flow cytometry analysis results based on the SSEA-3 (pluripotent stem cell marker) and the A2B5 (neuronal cell marker) markers for cell culture samples left for 7 days. Flow cytometry results for each sample are based on the average result between triplicate cultures.

For the flow cytometry analysis (Fig. 34), a small fraction of suspended TERA2.cl.SP12 cells were added to a 96-well plate as a suspension in wash buffer (0.1% Bovine Serum Albumin (BSA) in PBS) for incubation with primary (1:10 SSEA-3 and 1:20 A2B5) and FITC-conjugated secondary antibody IgM (Sigma, 1:100). Labelled cells were analysed in a Guave EasyCyte Plus System (Millipore) flow cytometer. Thresholds determining the numbers of positively expressing cells were set against the negative control antibody P3X. Analysis of cell surface markers through flow cytometry (Fig. 34) shows that at 7 days, AH61, ATRA and EC23 treated cells have down-regulated the expression of the SSEA-3
marker to almost zero, whereas the Control, DMSO and EC19 treated cells still have a high percentage of SSEA-3 expression. For the A2B5 marker, the AH61, ATRA and EC23 show up-regulation of the neuronal cell marker, whereas for the control, DMSO treated and EC19 treated cells, only a few cells have up-regulated the A2B5 marker associated with neuronal cells. Flow cytometry results also show that the majority of the AH61, ATRA and EC23 treated TERA2.cl.SP12 cells have started differentiate at the 7 day period, differentiating to produce neuronal phenotypes. However, along with the control and DMSO cell samples, the EC19 treated cells have remained largely undifferentiated. This is in agreement with previous results produced by Christie et al. 5-6 who showed that both ATRA and EC23 were excellent inducers of neuronal differentiation when introduced to TERA2.cl.SP12 cells for 7 days, producing cell surface marker expressions similar to Fig. 34. Results also showed that only a small proportion of EC19 treated TERA2.cl.SP12 cells down-regulated the pluripotent stem cell marker, SSEA-3, after 7 days of treatment; results did show EC19 to induce differentiation but predominately epithelial like cells after 14 and 21 days after treatment to TERA2.cl.SP12 cells; epithelial like plaques being formed from EC19 treated cell populations.

4.4 Data Pre-Processing

4.4.1 Extracting Relevant Cell Information from IR Chemical Images

A recorded spectral/chemical image of single cells on an IR substrate can contain many spatial regions with no relevant cellular absorbing material. These areas are of little interest. Therefore, for this thesis there needed to be a method for selecting and extracting pixels that contained absorptions from single cells, while leaving un-interesting spectral information from recorded images. An in house function created using MATLAB was produced to achieve these goals; a step by step breakdown of the in house function is described in Fig. 35.
Method Used to Extract Single Cell Spectra from a Spectral Image

**Fig. 35** Flowchart explaining the steps involved in the function used to extract single cell information from IR spectral images.

**Step 1**
(Spectral image displayed as a heat map based on the area under the amide I band for each pixel)

**Step 2**
(Lower and upper amide area thresholds are selected)

**Step 3**
(Binary image is created based on the threshold selected)

**Step 4**
(Cell boundaries are found and the average spectrum of a cell boundary is extracted)

**Step 5**
(Average cell spectra outputted)

Finish
Step 1: After the recording of a spectral/chemical image of a sample the image is then transformed to an intensity map based on the area under the amide I absorption band of each pixel in the image (Fig. 36 (a)). This is done using the trapezoidal numerical integration algorithm “trapz” built into the MATLAB software package, which finds the integral under a curve. In Fig. 36 (a); single cells in the image can be seen as light blue to green in colour and cells that have been overlayed thus, not been distributed on the CaF$_2$ in a monolayer fashion, yellow to red in colour. Pixels representing no cellular information are represented as dark blue meaning very small amide I area values.

Step 2: Cells that have not been distributed on to the CaF$_2$ substrate in monolayer fashion and have clumped together (cells that have been stacked on top of other cells or large masses of cells that are all connecting) can cause problems with the in-house built function as the function is unable to distinguish/extract information from the individual cells present in the clump. Instead, the function will output an average spectra of all of the cells present in the clump. An average spectrum of many cells is not ideal considering the aim of the study is to investigate the biochemical changes of single cells after retinoid addition.

The reason for the clumping of cells on the substrate is due to the poor distribution of cells using the cytopinning method. Thus, to only extract spectral information from single cells present in the recorded chemical images, pixels representing clumps of cells (pixels which are yellow and red) and pixels representing no cellular absorption information need to be removed. To do this, lower and upper thresholds based on the area under the amide I absorption band need to be selected; the lower threshold value removes the pixel spectra that contain no cellular absorbing material (dark blue spatial regions of the image) and the upper threshold value is used to remove pixel spectra which have high amide I integral values as a consequence of greater concentration of cells stacked together in the clump (red and yellow spatial regions of the image). For example, Fig. 36 (c) shows spectra from three different pixels in the image; Fig. 36 (c1) shows an IR spectrum containing no cellular information (pixels in the image highlighted as dark blue), Fig. 36 (c2) is a spectrum from a clump of cells (red pixel) and Fig. 36 (c3) is a pixel spectrum from a single cell (light blue pixel).
Step 1

Fig. 36 (a), image displaying the area underneath the amide I absorption band for each pixel contained within the original spectral image. Pixels containing cells can be seen as light blue through to red (Approximately 20 – 70 on the scale bar).

Fig. 36 (b) is a TIFF image of the original spectral image recorded. It can be seen that there is good agreement between the two images in the position of the cells.
Step 2

(c1) Spectrum from a dark blue pixel in the image. The pixel spectrum represents an area of the image containing no cellular absorbing material; Amide I integral of roughly 0.

(c2) Spectrum from a red pixel in image - one of the pixels making up a clump of cells in the image; Amide I integral score of 63.5. Greater absorption due to cells being stacked on top of one another in the clump.

(c3) Spectrum from a light blue pixel in image - one of the pixels containing absorption information from a single cell; in the image; Amide I integral score of 26.2.
**Step 3**

**Binary Image:**

All pixels that are in between the lower and higher threshold values are accepted to be extracted from the image. All pixels to be extracted now become equal to one and are displayed as red in the image. All pixels rejected based on the threshold values set now

---

Cell boundaries overlayed on top of the TIFF image from (b)

Image showing each average cell spectrum extracted from the spectral image. Average spectrum extracted for every numbered cell displayed.
Not surprisingly, the main absorption contributions from the spectrum in Fig. 36 (c1) are from carbon dioxide as no cellular material is present in this spatial area of the recorded chemical image. Fig. 36 (c3) is a pixel spectrum from a single cell and has an amide I integral area value of 26.2 and Fig. 36 (c2) represents a pixel spectrum from a clump of cells thus, a larger area under the amide 1 band (63.2) than the pixel spectrum from a single cell. To choose the upper and lower threshold values, the amide I integral area values of pixels associated with single cells, pixels associated with no cellular absorption and pixels associated with clumps of cells need to be assessed. In the example in Fig. 36, lower and upper threshold values of 21 and 41 respectively were chosen as the amide I integral thresholds.

Step 3: Based on the thresholds values chosen, the function then finds all pixels within the spectral image that have an amide I integral area value between 21 and 41, accepted pixels are given a one and rejected a zero. A binary image map based on ones and zeros is then
produced (Fig. 36 (d)), with all red pixels being accepted for extraction from the image while all blue pixels are rejected. The red pixels represent spectral absorptions of mainly single cells distributed in the image, with spatial regions with no biological single cells and clumps of many single cells now being removed; each small clump of connecting red pixels represents a single cell (approximately 5 pixels), or small grouping of single cells (approximately 5-30 pixels).

**Step 4:** The function then finds the boundaries of all the single cells/small groupings of cells in the binary spectral image. Fig. 36 (e) shows the boundary edges overlaid on top of the TIFF image from Fig. 36 (b), this gives an indication of the cells that have been extracted from the spectral image recorded and cells that have not; single cells with poor S/N not being extracted. The spectra obtained from each pixel in a given cell area are averaged to give a single unique spectrum for that single cell. Each unique single cell spectrum to be extracted is then given an identification number (Fig. 36 (f)).

**Step 5:** Finally, average cell spectra are the outputted from the function Fig. 36 (g). In the example in Fig. 36, 167 unique single cell spectra were extracted for a chemical image that took ~30 min to record.

**4.4.2 Data Pre-Processing and Chemometrics**

Before spectral analysis, all single cell spectra extracted from the recorded chemical images were firstly subjected to the RMieS-EMSC correction algorithm (20 iterations) to remove RMieS scattering from the recorded cell spectra (see 2. 6. 2), noise reduced (retaining the first information contained in the first 70 PCs) (see 2. 7. 1) and vector normalised (see 2. 7. 2). Any cell spectra with poor S/N were removed through the use of an absorption threshold on the amide I band; any spectra with an amide I band maxima of less than 0.4 being removed. Due to a number of previous studies showing spectral differences from the protein amide I secondary structure to be salient features when distinguishing stem cell spectra from their differentiated phenotypes (Chapter 3) [84, 91-92], spectral data was transformed to second derivatives (see 2. 7. 3) to reveal amide I secondary structure; spectra transformed to second derivatives and are to be used throughout the thesis (Spectra transformed to second derivatives with 11 point Savitsky-Golay smoothing). Second derivative spectra were then cut to the 1000-3060 cm\(^{-1}\) spectral
frequency range and all variable intensities between the 1770–2830 cm\(^{-1}\) frequency range were set to zero to remove carbon dioxide absorption from the analysis. The second derivative data were then transformed to principal components (PCs) through PCA (see 2.8.1) and the PRESS algorithm was used to find the number of PCs which contain spectral variance information (see 2.9.2). Finally, principal component linear discriminant analysis (PC-LDA) (see 2.8.2) was then used to analyse the recorded cell spectra and the performance of the PC-LDA multivariate classification models created were tested through the use of bootstrapping cross validation (2000 iterations) (2.9 and 2.9.1). If LDA scores plots showed separations between the spectra recorded from the differently treated cell spectra then spectral loadings and mean spectra comparisons were used to evaluate which absorption bands in the spectrum were causing the separations seen the LDA scores plots (Fig. 37).
Fig. 37 Data processing and Chemometric analysis flow chart.

1. **RMieS-EMSC (20 iterations)**
2. **Noise reduction algorithm (70 PCs retained)**
3. **Vector normalised**
4. **Threshold on the amide I band to remove spectra with poor S/N**
5. **Second derivative (11 point smoothing)**
6. **Cut the spectrum to spectral region of interest**
7. **Set the 1770 – 2830 cm\(^{-1}\) to 0 to remove carbon dioxide**
8. **Multivariate analysis (PCA)**
9. **Predicted residual error sum of squares (PRESS)**
10. **PC fed linear discriminant analysis (LDA)**
11. **Validation of the PC-LDA classification model created using bootstrapping cross validation (2000 iterations)**
12. **Investigation of spectral loadings and mean spectra**
4.5 Control Sample Spectra Reproducibility

All samples and control have been studied in multiple replicates (generally triplicate) and verification for internal consistency of the control replicates was undertaken using PCA. Fig. 38 shows PCA scores plots (PC1 vs. PC3 and PC2 vs. PC3) of the cell spectra recorded from the DMSO control sample replicates left in culture for 5 days before formalin fixation. The figure shows no clear spectral separation of the cell spectra recorded from the different biological replicates. PCA was also used to verify the internal consistency of all untreated (control) and DMSO treated TERA2.cl.SP12 control cell sample replicates at 7, 5 and 3 days; PCA scores plot results indicating that all biological replicates were spectrascopically similar. Therefore, it was assumed that all control biological replicates were consistent and spontaneous differentiation had not taken place to any serious degree in control sample cultures.

![Fig. 38 (Day 5 sample data) comparing DMSO sample biological replicates 1 (blue) 2 (black) and 3 (green) using PCA (a) PC1 vs. PC3 and (b) PC2 vs. PC3.](image)

4.6 Comparison of Control (DMSO and Untreated TERA2.cl.SP12 Cells) and EC23, EC19, AH61 and ATRA Treated TERA2.cl.SP12 Cells at 7 Days

Flow cytometry results (Fig. 34) show that the majority of EC23, AH61 and ATRA treated EC cells have differentiated 7 days after retinoid introduction to cell cultures, EC cells treated with the retinoids up-regulating the A2B5 neuronal cell phenotype marker thus,
indicating the production of neuronal cell phenotypes. Flow cytometry results also show the majority of control cells, DMSO treated and untreated TERA2.cl.SP12 cells have retained the pluripotent stem cell marker SSEA-3 and are therefore still pluripotent. Fig. 39 shows the LDA scores plot of the spectra recorded from DMSO, untreated, EC23, AH61 and ATRA treated cells, 7 days after retinoid and DMSO introduction to sample cultures. The LDA model was created using the spectral variance information contained within the first 8 PCs. The LDA 1 vs. LDA 2 scores plot clearly shows the majority of cell spectra recorded from retinoid induced differentiated EC stem cells can be separated from the undifferentiated EC stem cells; undifferentiated EC stem cells (untreated control (pink) and DMSO treated (black) TERA2.cl.SP12 cells) are mainly situated in the negative regions of the LDA 1 and LDA 2 scores axis. The separation of the cell spectra in LDA scores space would be in agreement with the flow cytometry results, showing biochemical differences between retinoid induced differentiated cells and cells that have retained the pluripotent cell marker, SSEA-3, after 7 days in cell cultures. The scores plot in Fig. 39 also shows the AH61 and ATRA treated TERA2.cl.SP12 cells to be spectrally similar, while a large number of EC23 treated cells are separated from the control and ATRA/AH61 treated cells, albeit with some overlap of the AH61/ATRA spectral cluster. This result is in agreement with previous work by Maltman et al.\textsuperscript{7} who showed slight biochemical differences between ATRA and EC23 treated TERA2.cl.SP12 cells using MALDI-TOF MS.

![Fig. 39 LDA 1 vs. LDA 2 of spectra recorded from DMSO, untreated (control), ATRA, EC23 and AH61 treated cells, 8 PCs used as input data for LDA and 90% confidence ellipses plotted; PRESS/RSS ratio result showed the 31st PC to exceed 1 and the model achieved an average bootstrap score of 70% CC (correctly classified).](image-url)
Fig. 39 reveals that spectral analysis of IR recorded data using multivariate classification algorithms, coupled with appropriate scatter correction, is capable of distinguishing pluripotent stem cells from their differentiated cell phenotypes, with the advantage of separations being based on the biochemical changes to the retinoid treated EC cell spectra as a result of through differentiation.

4.7 Biochemical Differences between Control Cells; DMSO Treated TERA2.cl.SP12 and Untreated TERA2.cl.SP12 Cells

As it was necessary to initially dissolve the retinoid compounds in DMSO prior to adding them to the aqueous based media solution at the desired concentration, therefore, DMSO is a vehicle and it was necessary to identify any spectral changes affected by DMSO itself. Several studies have shown that DMSO can induce the differentiation of both ES and EC stem cells with key stem cell markers, such as OCT-4, being down regulated in gene expression tests at certain concentrations \(^{93-95}\). The flow cytometry results in this study show that the TERA2.cl.SP12 cells treated with DMSO at 10 µM concentration show no significant differences from the untreated TERA2.cl.SP12 cells (Fig. 34). There was however a small drop in the SSEA-3 stem cell marker in the DMSO treated cells but not large enough to indicate that differentiation had begun to any significant level; this is consistent with previous work reported by Christie \(^{5-6}\).

Fig. 39 shows the DMSO treated and untreated TERA2.cl.SP12 cell spectra to be similar. However, using PCA to compare the spectra recorded from just the untreated and DMSO treated TERA2.cl.SP12 cells shows that there are some spectral differences between the different cell samples at 7 days (Fig. 40); spectral differences were also seen at 3 and 5 days and can be found in the supplementary materials. Therefore, even though no apparent down regulation of the stem cell marker (SSEA-3) was observed by flow cytometry for the DMSO treated samples, there is clearly a visible cell response to DMSO in the infrared spectra. This difference must be considered for accurate interpretation of the spectra generated for ATRA and synthetic retinoid treated cells to dissect out the effect of the DMSO vehicle. As a result, DMSO treated samples were selected as the control and will be used to compare retinoid treated EC cells in this study.
4.8 Biochemical Differences between Retinoid and DMSO Treated (control pluripotent stem cells) TERA2.cl.SP12 Cells at 7 Days

With flow cytometry (Fig. 34) indicating that AH61, ATRA and EC23 treated EC cells left in cultures for 7 days have differentiated and IR spectroscopic analysis of cell samples (Fig. 39) at the same time period showing EC23, ATRA and AH61 treated cells to be spectroscopically different from the undifferentiated stem cells, it can be seen that both IR with flow cytometric analysis results are in agreement. It is therefore important to understand the spectroscopic differences between the differentiated and the pluripotent stem cells so as to understand what biochemical changes are happening as a result of differentiation. To do this, cells spectra recorded from retinoid and the DMSO cell samples are compared using PC-LDA, with spectral loadings produced by the classification model being investigated along with mean spectra comparisons to deduce spectral differences causing the separations of the different cell spectra in LDA scores space. The EC19 cell spectra will also be compared against the DMSO cell spectra to see if spectroscopic differences can be seen, even though flow cytometry results show that the majority of EC19 cells have not differentiated at the 7 day period. When performing LDA on only two different groups of sample spectra only one separation output boundary is allowed (number of different spectral groups -1). As a result, a one dimensional LDA scores histogram is to be displayed to reveal the separation along the LDA 1 separation boundary. Fig. 41 (a) shows the LDA 1 scores histogram when using PC-LDA to compare the spectra recorded for DMSO and EC23 treated EC cells, left in sample cultures for 7 days.
**Fig. 41** (a) PC-LDA scores histogram from the average bootstrap result (2000 bootstraps) of spectra recorded from DMSO and EC23 treated cells left in sample cultures for 7 days; 5 PCs entered into the LDA classification. Average bootstrap achieved 96% CC (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC23 (blue) DMSO (black)) (c1) PRESS/RSS ratio result; PRESS/RSS ratio exceeds 1 after 21 PCs (c2) percent variance explained vs. PCs.
Fig. 41 (a) shows that there is separation between the DMSO and EC23 treated TERA2.cl.SP12 cells along the LDA 1 separation boundary. To produce the histogram in Fig. 41 (a), the DMSO and EC23 spectral data set has been partitioned into training and test sets. The training set being used to construct the PC-LDA classification model and the remaining test set (spectral data not used in the construction of the model) is predicted by the training set model constructed; the performance result is the percentage of test set spectra that have been correctly classified by the training set constructed classification model. Fig. 41 (a) shows both the training and test set scores results from the average bootstrap out of the 2000 created during the cross validation; average bootstrap achieving a performance score of 96% (96% of the test set spectra being correctly classified). This performance result shows the training model to be a good model for explaining spectral differences between the EC23 and DMSO spectral data recorded from the day 7 cell samples.

Fig. 41 (b) reveals the LDA 1 spectral loading (upper curves) describing the wavenumber frequencies which contribute to the spectral differences along LDA 1; the larger the loading band the greater the wavenumber frequency contributions to the separation along the LDA 1 separation boundary. The spectral loading can then be correlated with mean spectral differences between the two groups of the cell spectra; if there is a significant spectral change from the majority of cells in the group population as a result of differentiation, then the changes should be represented in the mean spectrum.

PC-LDA has been used to compare the spectra recorded from retinoid treated and DMSO treated TERA2.cl.SP12 cells at each time point. PC-LDA scores plot histograms and spectral loading/mean spectral comparisons of the day 7 data can be found in the supplementary materials (Chapter 4 day 7). For days 5 and 3 data, only the spectral loading/mean spectral comparisons are described in the supplementary materials Chapter 5 days 5 and 3). All PRESS and bootstrapping cross validation results for each PC-LDA model is also described in the supplementary materials; PRESS being used to reveal the number of PCs which explain spectral variance and the PC number where spectral variance is stopped being explained and noise is introduced. Only PCs containing relevant spectral variance are to be included in the LDA classification model.
Table. 2 Spectral differences between retinoid treated cell spectra and DMSO control cell spectra at 7, 5 and 3 days.

<table>
<thead>
<tr>
<th></th>
<th>Increase in absorption or new band (cm⁻¹) from retinoid treated EC cells</th>
<th>Decrease (&lt;) in absorption from retinoid treated EC cells or new band (cm⁻¹) from pluripotent control cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMSO vs. EC23 day 7</td>
<td>1061, 1201, 1552, 1655, 2854, 2922</td>
<td>1032, 1084, 1153, 1172, 1336, 1367, 1392, 1439, 1452, 1468, 1630, 1643, 1691</td>
</tr>
<tr>
<td>DMSO vs. AH61 day 7</td>
<td>1084, 1203, 1655, 1693</td>
<td>1034, 1155, 1172, 1367, 1381, 1439, 1452, 1468, 1510, 1545, 1630, 1640, 2850, 2920, 2958</td>
</tr>
<tr>
<td>DMSO vs. ATRA day 7</td>
<td>1084, 1203, 1657, 1693</td>
<td>1034, 1155, 1173, 1367, 1381, 1439, 1452, 1468, 1545, 1637, 1653, 1673, 1693, 2850, 2920, 2958</td>
</tr>
<tr>
<td>DMSO vs. EC19 day 7</td>
<td>1061, 1203, ~1475, 1552, 1653, 2850, 2920</td>
<td>1032, 1155, 1172, 1311, 1367, 1381, 1390, 1439, 1452, 1468, 1630, 1643, 2850, 2920, 2958</td>
</tr>
<tr>
<td>DMSO vs. EC23 day 5</td>
<td>1155, 1475, 1637, 1651-1662, 1650, 2850, 2920</td>
<td>1030, 1084, 1153, 1412, 1450, 1512, 1530, 1549, 1626, 1637, 1653, 1693, 2850, 2920, 2958</td>
</tr>
<tr>
<td>DMSO vs. AH61 day 5</td>
<td>1201, 1396, 1693, 1673</td>
<td>1032, 1082, 1124, 1155, 1173, 1381, 1439, 1452, 1468, 1512, 1530, 1626, 1637, 1653, 2846</td>
</tr>
<tr>
<td>DMSO vs. EC23 day 3</td>
<td>1082, 1201, 1653, 1635, 1460, 1475, 1552, 1653, 2850, 2950</td>
<td>1032, 1155, 1173, 1437, 1452, 1468, 1549, 1630, 1643, 1693, 2856, 2950</td>
</tr>
<tr>
<td>DMSO vs. AH61 day 3</td>
<td>1082, 1201, 1460, 1475, 1552, 1635, 1653, 1673</td>
<td>1032, 1437, 1452, 1468, 1549, 1630, 1643, 1693, 2850, 2950</td>
</tr>
<tr>
<td>DMSO vs. ATRA day 3</td>
<td>1082, 1155, 1201, 1651</td>
<td>1032, 1439, 1452, 1468, 1527, 1643, 1691, 2850, 2920, 2958</td>
</tr>
<tr>
<td>DMSO vs. EC19 day 3</td>
<td>1396, 1201, 1460, 1475, 1541, 1635, 1653, 1673</td>
<td>1032, 1153, 1173, 1381, 1453, 1512, 1527, 1547, 1610, 1630, 1643, 1672, 1707</td>
</tr>
</tbody>
</table>

Key

Wavenumbers (wavenumbers), highlighted in bold, are the frequency positions of the loading bands that have the greatest intensity thus, the wavenumbers which describe the greatest spectral differences between the retinoid treated and the pluripotent control stem cell spectra.

Wavenumbers highlighted in yellow (yellow) underline the wavenumber loading bands which are involved in the separation of all retinoid treated and pluripotent control stem cell spectra at 7 days.

Wavenumbers highlighted in green (green) underline the wavenumber loading bands which are described causing all retinoid treated cell spectra and the pluripotent control stem cell spectra to be separated at 5 days.

Wavenumbers highlighted in purple (pink) underline the wavenumber loading bands which are described causing all retinoid treated cell spectra and the pluripotent control stem cell spectra to be separated at 3 days.
4.8.1 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 7 Cell Samples

Firstly, spectral analysis results show the majority of cells contained within the cell population of all retinoid treated cell samples left in cultures for 7 days to have spectral differences when compared to spectra recorded from the pluripotent control cells (DMSO treated TERA2.cl.SP12 cells) (see supplementary materials, Chapter 4 - Day 7 Analysis, for spectral analysis results). With flow cytometry results (Fig. 3) showing that the majority of TERA2.cl.SP12 cells treated with EC19 for 7 days to still be pluripotent (a large number of EC19 treated cells yet to drop the SSEA-3 stem cell marker) IR data analysis would suggest that the majority of EC19 cells have been induced to differentiate and are going through a gearing up process, but are yet to fully differentiate at 7 days. This would also suggest that IRMS is an extremely sensitive technique capable of distinguishing differentiating cells at early stages of differentiation, before cells have dropped their stem cell marker. The spectral differences seen between EC cells treated with EC19 for 7 days and the pluripotent control cells may be a result of treated cells becoming progenitor cells before full transformation to the epithelial cell types.

In Table 2 it can be seen that all retinoid treated cells left in sample cultures for 7 days have increases in relative spectral absorption (relative absorption due to spectral data being transformed to second derivatives) from ~1203 and ~1655 cm\(^{-1}\), and decreases in relative spectral absorption from 1032, 1152, 1173, 1439, 1452, 1468 and ~1640 cm\(^{-1}\), when compared against the spectra recorded from the control stem cells; spectral absorption band assignment to molecular vibrations can be found in Table 3. As well as these spectral differences, EC23, AH61 and EC19 treated cells all show a decrease in relative intensity from the absorption band at 1630 cm\(^{-1}\). These spectral differences suggest that the majority of retinoid treated TERA2.cl.SP12 cells show a net change in protein secondary structure as a result of differentiation; retinoid treated cells showing decreases in relative absorption from absorption bands associated with β-sheet structural proteins and an increase from the absorption band associated with α-helix structural proteins. Changes to protein secondary structure from differentiating/differentiated stem cells are in agreement with previous studies using IRMS to investigate stem cell differentiation\(^{84, 91-92, 96}\). A study Thanthanuch et al.\(^{91}\) showing similar protein secondary structure changes when investigating the differentiation of mESCs down neural lineages and Ami and colleagues\(^{84, 96}\) showing a reduction in relative absorbance at ~1639 cm\(^{-1}\) and an increase at 1658 cm\(^{-1}\) from differentiated embryonic stem cells. Therefore, from the spectral analysis results, it would
seem that the new neuronal cell phenotypes require a greater net concentration of α-helix structural proteins, while pluripotent stem cells require a greater net concentration of β-sheet proteins. The decrease in relative absorbance from absorption bands which have previously been associated with glycogen and carbohydrate molecules may be a direct result of extra work being done by the retinoid treated cells during the differentiation process; energy stores of the cells being decreased while new α-helix structural proteins are synthesised for the new neuronal phenotype being produced.

4.8.2 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 5 Cell Samples

For day 5 data, EC23 and AH61 recorded cell spectra showed spectral differences when being compared against the DMSO control cell spectra but ATRA and EC19 did not. As a result of this only EC23 and AH61 spectral comparisons are described for 5 day cell samples (see supplementary materials, Chapter 4 - Day 5 Analysis, for spectral analysis results); all retinoid treated and DMSO treated control cell spectral comparisons using PC-LDA are described in Table. 2.

The spectral differences showed that (Table. 2) AH61 and EC23 treated cells have relative increases in spectral absorption from 1473, 1651-1662 and 2920 cm\(^{-1}\), and decreases in relative spectral absorption from 1030, 1082-84, 1383-85, 1450, 1466, 1512-14, 1626-30 and 1637-40 cm\(^{-1}\), when compared to the control stem cell spectra. The increase in relative absorption from the band situated at 1651-1661 cm\(^{-1}\) (amide I α-helix band) and decreases in relative absorption from the absorption bands associated with β-sheet protein structures (1626-30 and 1637-40 cm\(^{-1}\)) were also seen when comparing spectra recorded 7 day cell samples. As well as these spectral differences, a decrease in relative spectral absorption was also seen from absorption bands situated at 1030, 1439 and 1452 cm\(^{-1}\) when investigating the spectral differences of spectra recorded from 7 day cell samples. This may show that these spectral changes are all necessary biochemical changes for the differentiation process. The loss of relative absorption intensity from the 1030 cm\(^{-1}\) absorption band again possibly indicating cell energy stores being used during the synthesis the new proteins needed for the phenotype being produced, as a consequence of differentiation.
4.8.3 Discussion of the Spectral Differences Found between Retinoid Treated Cells and Control cells; Day 3 Cell Samples

For day 3 cell samples, all retinoid treated TERA2.cl.SP12 cells showed spectral differences when compared to the control stem cell spectra, however, ATRA treated cells showed the least spectral separation, with the least number of cell spectra being able to be separated from the control cell spectra along LDA 1. This therefore resulted in the poorest bootstrapping model cross validation score of 75%, possibly indicating that ATRA is slow at starting the differentiation process compared to its synthesised chemical derivatives.

Retinoid treated cell spectra show relative increases in spectral absorption from 1201, 1653 cm\(^{-1}\), and decreases in relative spectral absorption from 1032, 1453, 1468 and 1643 cm\(^{-1}\), when compared against control stem cell spectra. Again, spectral analysis results show retinoid treated cell spectra to have a change in protein secondary structure when compared to the control stem cells; decreases in relative absorption from absorption bands associated with β-sheet proteins of the amide I absorption band.

The 1653 cm\(^{-1}\) loading band is also shown to be a significant spectral difference between the retinoid treated and the control cell spectra (increase in absorption or new band (cm\(^{-1}\)) from retinoid treated EC cells at 1653 cm\(^{-1}\) frequency). However, this spectral difference does not signify an increase in relative absorption of the α-helix amide I band for retinoid treated cells, but rather a change in α-helix band position (Fig. 42); retinoid treated cell spectra having an α-helix absorption band that has shifted to lower a vibrational frequency when compared to the control cell spectra. This net change in α-helix protein structure may be a spectral biomarker indicating the start of the differentiation process. Also, decreases in relative spectral absorption can be seen from absorption bands at 1032, 1453 and 1468 cm\(^{-1}\), these spectral changes have been witnessed when assessing the spectral differences between retinoid and control cell spectra at all time periods. Although it is difficult to characterise the 1453 and 1468 cm\(^{-1}\) absorption bands to a particular molecule vibration, it would seem that the molecular structures involved in these absorption bands play a significant role in differentiation, and all three absorption bands may be key spectral biomarkers for the differentiation process of this cell line.
### Table 3 Assigning spectral bands to molecular vibrations.

<table>
<thead>
<tr>
<th>Absorption bands (cm⁻¹)</th>
<th>Band Assignments to molecular vibrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1032</td>
<td>Previously been assigned to glycogen C-O vibrations of glycogen ⁹⁷.</td>
</tr>
<tr>
<td>1082-84</td>
<td>PO₂⁻ phosphate symmetric stretching vibration from nucleic acids (DNA and RNA) ⁹⁷.</td>
</tr>
<tr>
<td>1151</td>
<td>Previously been assigned to glycogen absorption due to C-O and C-C stretching and C-O-H deformation motions ⁹⁷.</td>
</tr>
<tr>
<td>1173</td>
<td>Previously been assigned to C-O stretching of serine, threosine and tyrosine cell proteins as well as carbohydrate molecules ⁹⁷.</td>
</tr>
<tr>
<td>1203</td>
<td>Previously been assigned to vibrational modes of collagen proteins - amide III absorption ⁹⁷.</td>
</tr>
<tr>
<td>1367</td>
<td>Previously been assigned to stretching C-O and C-H/N-H deformation vibrations ⁹⁷.</td>
</tr>
<tr>
<td>~1439</td>
<td>Previously been assigned to CH₂ vibrations of lipid and fatty acid molecules ⁹⁷.</td>
</tr>
<tr>
<td>1451</td>
<td>Previously been assigned to asymmetric CH₃ bending modes of the methyl groups of proteins and lipid molecular structures ⁹⁷.</td>
</tr>
<tr>
<td>1468</td>
<td>Previously been assigned to CH₂ bending vibrations of lipid and protein molecules ⁹⁷.</td>
</tr>
<tr>
<td>1630</td>
<td>Previously been assigned to the amide I absorption band of proteins with β-sheet structural conformations ⁵⁷, ⁹⁷-⁹⁸.</td>
</tr>
<tr>
<td>1637-1643</td>
<td>Previously been assigned to the amide I absorption band of proteins with β-sheet structural conformations ⁵⁷, ⁹⁷-⁹⁸.</td>
</tr>
<tr>
<td>1655-60</td>
<td>Previously been assigned to the amide I absorption band of proteins with α-helix structural conformations ⁹⁷.</td>
</tr>
<tr>
<td>1691</td>
<td>Previously been assigned to high frequency vibration of an antiparallel β-sheet of amide I ⁵⁷, ⁹⁷-⁹⁸.</td>
</tr>
<tr>
<td>2950</td>
<td>Previously assigned to the symmetric stretching vibration of CH₂ of acyl chains (lipids) ⁹⁷.</td>
</tr>
<tr>
<td>2922</td>
<td>Previously assigned to the asymmetric stretching vibration of CH₂ from acyl chains (lipids) ⁹⁷.</td>
</tr>
<tr>
<td>2958</td>
<td>Previously assigned to the asymmetric stretching vibration of CH₂ of acyl chains (lipids) ⁹⁷.</td>
</tr>
</tbody>
</table>
Comparing the mean spectra of all 3 day cell samples spectra (amide I spectral range 1600-1700 cm$^{-1}$); EC23 (blue), EC19 (red), ATRA (green) and AH61 (cyan). AH61, EC23 and EC19 mean cells spectra show a shift in α-helix band position, moving a lower frequency, which accounts for the large spectral loading band for the 1653 cm$^{-1}$ absorption band. ATRA shows a mean spectrum which is similar to the control cells; in agreement with the spectral analysis results that showed that not only ATRA treated cell spectra could be separated from the control spectra using LDA.

4.9 Comparing the Different Actions of the Retinoids on the TERA2.cl.SP12 Cells

Previous work has shown that the EC23 and ATRA treated TERA2.cl.SP12 cells produce very similar protein profiles during the cell differentiation process and therefore, a close functional relationship. However, subtle but significant difference in the response of the TERA2.cl.SP12 cells to both ATRA and EC23 was seen, primarily from the cellular retinoic acid binding protein 1 (CRABP1). A study by Christie et al. also looked at the similar biological effects the two retinoids had on the TERA2.cl.SP12 cells, their study showing that EC23 induced the production of neural tissue at a similar or possibly increased level to that seen when using ATRA. Fig. 39 shows that even though both ATRA and EC23 are known to induce the cellular differentiation of the TERA2.cl.SP12 cells to form the same neuronal cell phenotype, significant spectral differences can be seen between the EC23 and ATRA treated TERA2.cl.SP12 cells at the same relative time points. This is therefore consistent with previous observations seen using other techniques. However, interestingly, Fig. 39 shows the AH61 and ATRA treated TERA2.cl.SP12 cells...
to be biochemically similar 7 days after the introduction of the retinoids; indicated by the virtual overlap of the cyan and green data in the Fig. 39 LDA scores plot. The synthetic retinoid AH61 was designed to have a similar chemical structure to the natural ATRA, but with the structural stability of EC23. A recent study, using both flow cytometry and RT-PCR, showed AH61 to have a biological activity comparable or greater than that of ATRA.

Fig. 43, shows ATRA, AH61, EC19 and EC23 treated cell spectra recorded from 7 day cell samples and compared using PC-LDA; LDA 1 vs. LDA2 scores plot separating AH61/ATRA and EC23 treated TERA2.cl.SP12 cells along LDA 1 and the LDA 2 separation boundary separates neuronal differentiated from EC19 treated TERA2.cl.SP12 cells, which have been shown to produce predominately epithelial cell types but only after 14-21 days of retinoid treatment in cultures. The scores plot (Fig. 43) again shows biochemical differences between EC23 and ATRA/AH61 treated EC cells, which would be in agreement with previous work and also show the synthetic retinoid, AH61, to have a similar biochemical action on the TERA2.cl.SP12 cells to that of ATRA, but with the added advantage of chemical structure stability. This could result from a number of possibilities: (i) the EC23 and ATRA/AH61 treated TERA2.cl.SP12 cells follow different differentiation pathways: or (ii), the retinoid treated TERA2.cl.SP12 are following the same differentiation pathway but staggered relative to each other due to the differences in each initiators ability to start the differentiation process.

For the spectral differences between EC23 and ATRA/AH61 treated cells left in cultures for 7 days, Fig. 43 and supplementary materials (Chapter 4: Retinoid Treated Cell Spectra Comparisons), show the EC23 treated TERA2.cl.SP12 cell spectra to have greater relative absorption from the absorption bands associated with lipid molecular structures (2850 and 2920 cm⁻¹) and spectral differences in protein secondary structure when compared to AH61/ATRA cell spectra; EC23 cell spectra showing a shift in the α-helix amide I band to a lower vibrational frequency and AH61 and ATRA cell spectra showing greater relative absorption from absorption bands associated with β-sheet protein structures thus, indicating net protein structural changes.
Fig. 43 LDA 1 vs. LDA 2 of spectra recorded from ATRA, EC19, EC23 and AH61 treated cells, 10 PCs entered into the LDA classification. Average bootstrap achieved 75% CC and 90% confidence ellipses plotted; the wavenumber absorption bands responsible for the separations in LDA scores space are described (spectral analysis can be found in the supplementary materials).
Other studies looking into the spectral fingerprint changes of stem cells and their differentiated derivatives have showed decreases in lipid band intensities from differentiating and terminally differentiated cells. A study by Thumanu et al. showed that spectral lipid intensities increased initially when hepatic progenitor cells differentiated, however, the intensities significantly decreased upon the final stage of differentiation, producing terminally differentiated hepatocyte-like cells. This showed that lipid structures are actively involved in the differentiation of embryonic derived hepatocytes. This may also be true for TERA2.cl.SP12 cells, which may suggest that AH61, ATRA and EC23 treated EC cells, are all on the same differentiation pathway but EC23 treated TERA2.cl.SP12 cells are spectroscopically different due to the majority of treated cells being at different stage in the differentiation process.

With the majority of EC19 treated TERA2.cl.SP12 cells not being shown to have differentiated after 7 days of retinoid treatment it is not surprising that EC19 treated cells show separation from the neuronal differentiated cells in the LDA scores plot. The LDA 2 spectral loading (Fig. 43) describes increased relative absorption from bands associated with glycogen and carbohydrate molecules (1032, 1153 and 1173 cm$^{-1}$) and decreases in relative absorption from bands associated with β-sheet protein structures (1630 and 1690 cm$^{-1}$) and lipid molecules (2850, 2920 and 2958 cm$^{-1}$) for the EC19 treated TERA2.cl.SP12 cells, when compared against other retinoid treated cells.

Spectral comparisons of EC19 cells with control cells showed EC cells treated EC19 for 7 days to have similar spectral differences to other retinoid cellular differentiating models, i.e. increase in relative absorption from the α-helix band and decreases in relative absorption from bands associated with β-sheet proteins, lipids and carbohydrates. Based on this spectroscopic evidence, it may suggest that the majority of EC19 treated cells are on the differentiation pathway after 7 days of retinoid treatment but are yet to fully differentiate. Therefore, these small but significant differences may describe the biochemical differences of a differentiated cell and a cell which is on the differentiation pathway but not fully dropped its pluripotency yet, possible cells at the progenitor cell stage.
4.10 Monitoring the Spectroscopic Signals of Differentiating Cells through Time using IRMS

Fig. 44 shows the PC-LDA comparison of spectra recorded from AH61 treated TERA2.cl.SP12 cells at 3, 5, 7, and 14 days; LDA classification model produced using 5 PCs and LDA scores plot showing 3 distinct clusters, 3 day cell spectra cluster, 5 and 7 day cell cluster and 14 day cell cluster. This may suggest that AH61 is quick to induce the differentiation process, with the majority of AH61 treated cells to be differentiated after 5 days of AH61 treatment (5 and 7 day cell spectra clusters overlapped in LDA scores plot). This shows the potential of IRMS for monitoring cell phenotype.

Fig. 44 PC-LDA scores plot (LDA 1 vs. LDA 2) of spectra recorded from AH61 treated cells samples at the 3 day (black), 5 day (blue), 7 day (green) and 14 day (red) time points, 5 PCs entered into the LDA classification with an average bootstrap achieving 68% CC.

Spectral comparisons of cell spectra recorded from DMSO, EC23 and ATRA treated TERA2.cl.SP12 cells at 3, 5, 7 and 14 days after retinoid treatment can be seen in the supplementary materials (Chapter 4 - Monitoring the Spectroscopic Signals of Differentiating Cells through Time using IRMS). Interestingly, Spectral differences can also be seen between DMSO cell spectra recorded from cells at 3, 5 and 7 days after
DMSO treatment. This is may not be particularly surprising considering the complex biochemical changes of a cell during its life cycle; one also has to consider the doubling rates of the cells and how this is affected by the nutrients in the culture media being used up when cultures start to become crowded (doubling time for these cells is approximately 18-24 h).

The changing biochemistry of undifferentiated stem cells in culture is something which has been overlooked in a number of previous studies when using IRMS to investigate the changes of differentiating stem cells\textsuperscript{57, 84, 92, 96}. For example, a study by Ami and colleagues\textsuperscript{96} only investigated the biochemical differences of average recorded spectra at the different time points during the differentiation process, without any consideration for the changes of the pluripotent stem cells through time. Spectral analysis results of the DMSO cell spectra in the supplementary materials show that the biochemical changes of the pluripotent stem cells are not trivial and without adequate control cells to compare against, i.e. stem cells in culture for the same period of time as the differentiating cells, spectral differences between differentiating and undifferentiated cells will be due to both cell cycle biochemical changes and biochemical changes due to differentiation. Thus, a lack of adequate control cells to compare against will lead to the wrong conclusions being made when investigating spectral differences of differentiating cells.

In the supplementary materials (Chapter 4 - Day 14 Retinoid Treated EC Cells), spectra recorded from ATRA, AH61 and EC23 treated TERA2.cl.SP12 cells left in sample cultures for 14 days are compared against spectra recorded from DMSO treated control cells left in culture for 7 days using PC-LDA (DMSO treated control cells became over-confluent after 7 days and the sample cultures were subsequently stopped as a result thus, no control cell samples at day 14). Spectral comparisons using LDA show all the neuronal differentiating cells to be biochemically similar due to the majority of retinoid treated cell spectra being situated in the negative scores space of the LDA 1 separation boundary, and control cell spectra being situated in the positive region of LDA 1 scores space. This would suggest that the majority of neuronal differentiating TERA2.cl.SP12 cells are terminally differentiated to the new neuronal cell phenotype at 14 days. Spectral loadings for LDA 1 (Chapter 4 - Day 14 Retinoid Treated EC Cells (Sup Fig. 10 (b))) show the retinoid treated cells left in culture for 14 days to have new spectral bands at wavenumbers 1024 and 1041 cm\textsuperscript{-1}, possible due to the cells taking on a new source of glucose or carbohydrate (1024 cm\textsuperscript{-1})
has been associated C-O stretching vibrations of glycogen and 1041 cm\(^{-1}\) associated with C-O bending frequencies of the C-OH groups of carbohydrates)\(^{97}\). Retinoid cell spectra also show increases in relative absorptions from the 1643, ~1670 and 1693 cm\(^{-1}\) absorption bands, associated with β-structural proteins (~1670 cm\(^{-1}\) associated with β-turn protein structures)\(^{57}\), the 1658 cm\(^{-1}\) α-helix structural band, and absorption bands at 1155, 1236, 1313, ~1400, 1458, 1473 2848, 2910, 2929, 2960, 3010 and 1730/1750 cm\(^{-1}\). The absorption bands at 1313 and ~1400 cm\(^{-1}\) previously being associated with the amide III band of proteins and the symmetric stretching vibration of the COO-functional group, thought to be from primarily fatty and amino acids respectively\(^{97}\). Mean spectral comparisons show all retinoid treated cell spectra to have a significant spectral change when compared to the control cells in the 1350-1420 cm\(^{-1}\) spectral range (Chapter 4 - Day 14 Retinoid Treated EC Cells (Sup Fig. 10 (c1))). Although it is difficult to assign vibrations in this spectral range to a specific molecule or functional group, these biochemical changes could be significant markers of new neuronal cell phenotype produced through differentiation. As well as these data results findings, retinoid treated cells at day 14 also reduced relative intensity at 2920 cm\(^{-1}\) (asymmetric stretching vibration of CH\(_2\) from acyl chains) and increased at ~2960 cm\(^{-1}\) (assigned to the asymmetric stretching vibration of CH\(_3\) of acyl chains). This may give insight into the chain lengths of saturated lipid molecules needed for differentiated and pluripotent cells; results suggest that terminally differentiated cells needing longer molecular chain lengths, whereas and pluripotent stem cells contain smaller chain lengths. The ratio could be used as a potential biomarker for terminally differentiated cells.

4.11 Conclusions

To conclude, in this study IRMS has been shown to be an effective tool that can distinguish between stem cells and their differentiating derivatives without the need for fluorescent or magnetic biomarkers, identifying cell phenotypes in a label free manner. The use of appropriate scatter correction algorithms mean that the spectra observed are indicative of the cellular biochemistry present and are not distorted by RMieS scattering. With the addition of multivariate chemometric algorithms, pluripotent stem cells can be characterised from their induced derivatives based on the spectral differences thus, highlighting the ability of the technique for potentially discovering spectral biomarkers of the differentiation process; for this study, the amide I secondary structure has been a particularly salient absorption band when investigating the spectral changes as a result of
differentiation, in agreement with previous studies using IRMS to investigate the
differentiation process of single cells. This study therefore highlights the potential of IRMS
as a screening tool in future stem cell therapy, distinguishing undifferentiated cells in cell
populations before implantation therefore reducing the risk of tumour formation.
Chapter 5

The Action of Retinoids on Human Pluripotent Stem Cells at Early Time Points after Retinoid Addition and the Investigation of the RMieS-EMSC Correction algorithm

5.1 Abstract

In this investigation the spectra recorded from AH61, ATRA and EC23 treated TERA2.cl.SP12 cells have been compared using PC-LDA at time points 6 h, 24 h, 1, 2, 3, 5, 7 and 9 days after retinoid introduction to cell cultures; retinoid concentrations of 1 µm being used to induce differentiation. Spectral analysis results show poor separation of the retinoid treated and control stem cell spectra from 3 and 5 day data, with some separations being seen from data recorded from days 7 and 9 samples. However, no separations were seen from spectra recorded from the earlier sample time points. A combination of lower retinoid concentration, reduced cell spectra extracted from chemical images and poor RMieS corrected spectra may have all been contributing factors to the poor separations seen from the data recorded for this investigation. To understand why poor corrected spectra were produced when using RMieS-EMSC correction algorithm, the RMieS-EMSC algorithm and its correction parameters were evaluated in this investigation. Results show that the RMieS-EMSC should not be treated like a black-box, with recommended correction parameter values not being appropriate for all cell sample types. These investigation results therefore highlight the continued problem of RMieS when recording spectra of cells in air, with also providing a potential solution through altering the correction parameter value of the RMieS-EMSC correction algorithm.
5.2 Introduction and Aims

In the previous chapter it was shown that it was possible to distinguish the biochemical changes of retinoid treated TERA2.cl.SP12 cells from pluripotent cells using IRMS, with spectral differences being seen as early as three days after retinoid addition to cell cultures. Therefore, can the action of potent inducing agents (ATRA, AH61 and EC23) be seen earlier than 3 days after retinoid addition to cell cultures? Also, by having a greater number of time point samples can the biochemical changes of the TERA2.cl.SP12 cell line be mapped spectroscopically, possible enabling spectral biomarkers of neuronal differentiation to be found? To investigate this, the action of ATRA, AH61 and EC23 on pluripotent TERA2.cl.SP12 cells is again probed with IRMS coupled with multivariate algorithms and appropriate scatter correction, however, pluripotent control spectra are now to be compared with retinoid treated cell spectra at 6 h, 24 h, 1, 2, 3, 5, 7 and 9 days. The recording of data from the earlier times after retinoid introduction will show whether the retinoids induce a biochemical action initially after being introduced to cell population and will also provide a greater number of spectroscopic time points so as to help deduce spectral biomarkers of cellular differentiation.

5.3 Cell Sample Preparations

All cell samples for this study were prepared by B. A. Murray at Durham University.

5.3.1 Protocol for the Cell Culture of TERA2.cl.SP12 Cells

The culturing procedure for the TERA2.cl.SP12 EC cells was the same as explained in 4.2.1. The only differences being that the final concentration for both retinoids/DMSO in cell cultures was 1 µm and not 10 µm, the samples numbers were increased spanning a time point range of 6 h to 9 days and no EC19 cell samples were produced as it is a pure inducer of cellular differentiation; the retinoid concentration was changed from 10 to 1 µm as a result of the synthetic retinoid, EC23, being shown to be more potent inducer of cellular differentiation at lower concentrations than the natural retinoid, ATRA. The reason for this is thought to be as a result of the greater stability of the synthetic analogous of ATRA.

Again, cells were chemically fixed using 4% formalin solution and transported to Manchester University where they were deposited onto CaF₂ IR sample substrates using the cytospinning method for spectroscopic data recording.
5.4 Data Pre-Processing

5.4.1 Improvements Made to the Single Cell Finding Function
In the previous chapter the in-house developed single cell finding function, created using MATLAB, was successfully used to extract spectral information from single cells present in chemical images. However, the cytopinning method used to distribute cells onto the CaF$_2$ surfaces was not reproducible; some cell samples having a good distribution of single cells, whereas other cell samples had more clumps of cells than well distributed single cells on the surface. The problem with the clumping of cells on the IR substrates is that the in-house function is unable to distinguish/extract information from the individual cells present in the clump. Instead, the function will output an average spectra of all of the cells present in the clump, which is not ideal considering the aim of the study is to investigate the biochemical changes of single cells as a result of cellular differentiation. Therefore, for this experiment, the concentration of cells being distributed onto CaF$_2$ slides using the cytopinning method was lowered (lowering the cell concentration significantly reduced the number of cell clumps produced, which resulted in well separated single cells) and the in-house function was updated so that any cell clumps present in the image could be removed from the image; the contrast of chemical images was also improved so that the cell boundaries of small single cells could be found (See Fig. 45 for an explanation into the improvements made to the single cell finding function).

**Improvement Step 1**
Firstly, lowering the concentration of formalin fixed cells being used for the cytopinning method reduced the number of cell clumps being produced on cell samples; Fig. 45 (a) showing only two clumps of cells. The improvement in the spread of cells in the image has also reduced the number of single cells near to neighbouring cells; therefore, clear absorption gaps can be seen between cells making it easier for the function to find the boundaries of the individual cells present. The improvement in image contrast has also helped visualise some of the small single cells which have a lower recorded IR signal.
**Improvement Step 1**

Normal spectral image but now with improved contrast helps in the visualisation of the single cells present in the image. It can also be seen that with the reduced cell density the cells are more spread out with minimal contact to other cells. However, a couple of cells in the image are in contact with others and one two clump of cells can be seen.

**Improvement Step 2**

To split up the cells that are in contact with others, and remove clumps from the spectral image, a function has been made, which enables the user to create a mask. This enables sections of the spectral image to be removed. Clumps of cells can now be removed and absorption gaps can be made, splitting up two or more cells that are connected.

**Improvement Step 3**

Mask created to remove the pixels situated within the spatial regions selected to be removed from the recorded image.
The boundaries for each cell are then found before with absorption thresholds used to remove pixels which do not contain any absorbing matter. This image displays the boundaries found in blue, with the cells in white.

The cell extraction map shows 72 single cell spectra extracted from the image, producing 72 average single cell spectra for the recorded chemical image. The combination of adapting the function and the reduced cell density used for the cytospinning procedure has reduced the problems seen in Chapter 4.

The 72 extracted mean cell spectra from the image.

Fig. 45 Graphical step by step of the improvements made to in-house developed single cell finding function.
**Improvement Step 2**

Fig 45 (a) shows two clumps of single cells and some single cell spectra with close neighbouring cells; close neighbouring cells in a spectral image can also be difficult for the single cell finding function because if cells are situated very close to one another in the image there will be no clear absorption gap defined between the cells, and as a consequence, the function will not be able to find the individual cell boundaries. Therefore, to split up neighbouring cells and remove cell clusters, the function was adapted so that spatial regions of the image could be removed. This allows spatial gaps to be created between neighbouring cells and whole cell clusters can now be removed entirely (Fig. 45 (b)).

**Improvement Step 3**

After the regions of the spectral image to be removed have been selected a mask is created (Fig. 45 (c)); the white regions of the mask show the pixel in the image which are to be kept and the black are the pixels to be removed from the recorded image before single cell finding.

After the spatial regions of the chemical image had been removed (spectra contained within the spatial regions selected set to 0) the single cell finding function worked as explained previously (4. 4. 1); absorption threshold used to remove spatial regions of the recorded image with no absorbing matter present, cell boundaries found from each single cell remaining (Fig. 45 (d)) with an average cell spectrum recorded from each single cell present in the image (Fig. 45 (e)). Fig. 45 showing that 72 single cell spectra could be extracted from the recorded image (Fig. 45 (f)).

**5. 4. 2 Data Pre-Processing and Chemometrics**

The data pre-processing and Chemometric analysis used for data in this study were the same as described in 4. 4. 2, the only difference being an increase in number of iterations for the RMieS-EMSC algorithm; increasing from 20 to 200. This increase in the number of iterations was due to greater RMieS scattering being seen from the single cell spectra recorded. A previous study by Bassan and colleagues$^{77, 81}$ stated that each iteration of the RMieS-EMSC algorithm results in a more accurate correction, moving more towards the pure spectrum with increased numbers of iterations. Therefore, with increasing the number
of iterations used in the RMieS-EMSC correction algorithm the quality of the corrected spectra should be improved and the greater RMieS scattering should be accounted for.

5.5 Comparison of Control (DMSO treated TERA2.cl.SP12 Cells) and EC23, AH61 and ATRA Treated TERA2.cl.SP12 Cell Spectra

ATRA, EC23 and AH61 all induce the differentiation of the TERA2.cl.SP12 cell line to produce neuronal cell phenotypes. When comparing the spectra recorded from the EC23, ATRA, AH61 and control cells (DMSO treated) left in sample cultures for 9 days using PC-LDA, Fig. 46 shows that the majority of AH61 and EC23 treated cells spectra can be separated from the cluster of control cell spectra, whereas a fewer ATRA treated cell spectra are separated from the scores area defined by the control cell spectra. This would suggest that the majority ATRA treated TERA2.cl.SP12 cells are biochemically similar to the control stem cells, even after 9 days of retinoid treatment. This may possibly indicate that a large number of ATRA treated TERA2.cl.SP12 cells are yet to fully differentiate, whereas results suggest that the majority of the EC cells treated with the synthetic analogues of ATRA to either be differentiating or differentiated.

![Fig. 46 PC-LDA scores plot (LDA 1 vs. LDA 2) of all ATRA, EC23, AH61 and DMSO treated (control) cell spectra recorded from the 9 day cell samples; 9 PCs inputted into the LDA algorithm and 90% confidence ellipses plotted. PRESS/RSS ratio result showed the 9th PC to exceed 1 and 9 pcs was equal to 75% of the original spectral variance. LDA model achieved an average bootstrap score of 74% CC; 90% confidence ellipses are plotted just for the visual effect and are not used in the statistical analysis of data.](image-url)
The analysis result in Fig. 46 is in agreement with previous studies which have showed ATRA to be a less effective cellular differentiation agent when compared to EC23 at lower concentrations. In addition, the spectroscopic result in Fig. 46 suggests that the synthetic retinoid, AH61, is also a more potent inducer of cellular differentiation than ATRA at lower concentrations; LDA loading plots and mean spectra comparisons are found in the supplementary materials (Chapter 5 – Day 9 data analysis (Sup Fig. 11)). With both EC23 and AH61 molecules shown to have greater stability than ATRA, the analysis results in Fig. 46 would corroborate the theory that the greater potency shown by the synthetic analogues of ATRA are a result of greater compound stability.

The major loading bands produced as a result of the LDA are described in Table. 4. Through comparing the major loading bands produced from the analysis of 9 day cell spectra with the mean spectra of each unique sample (Chapter 5 – Day 9 data analysis (Sup Fig. 11)), results show the EC23 and AH61 treated cell spectra to have increases in relative absorption from bands primarily associated with lipid molecular structures (2847 and 2915 cm⁻¹), the glycogen band at 1032 cm⁻¹, and decreases in relative absorption from bands associated β-structural proteins (1630, 1640, 1674 and 1693 cm⁻¹) plus the bands situated at 1451 and 1466 cm⁻¹; mean spectral comparisons also show the EC23/AH61 treated cell spectra to have a greater relative absorption from the α-helix band, situated at ~1656 cm⁻¹ when compared to both ATRA/DMSO cell spectra; in agreement with the previous data analysis in chapter 4. Comparing these spectral differences to the spectral differences found when investigating the data recorded from cells treated with 10 μM DMSO/retinoids reveals the biochemical differences found between retinoid treated and control cell spectra to be similar to the differences when comparing data from day 7 samples (chapter 4); drop in relative intensities from bands associated with amide I β-structural proteins being seen for both data analysis. The change in protein secondary structure indicating a change in net protein for differentiated TERA2.cl.SP12 cells, possible as a result of the down regulation of β-rich proteins needed for the stem cell phenotype. As the changes seen from separated 9 day data are similar to differences found from 7 day data in the previous chapter (chapter 4), it may be assumed that the separations are a direct result of biochemical changes due to differentiation.

Results also show the ATRA cell spectra to have reductions in relative absorption from the 1630 and 1692 cm⁻¹ β-structural protein bands, but not from the β-sheet and antiparallel β-sheet bands at 1640 and 1674 cm⁻¹ respectively. As well as these spectral differences,
ATRA cell spectra also show increases in absorption from the lipid bands situated at 2849 and 2920 and a decrease at 2961 cm\(^{-1}\); these spectral differences are in agreement with the differences separating the majority of EC23 and AH61 cells from the cluster of control stem cell spectra along LDA 1. This may suggest that these biochemical changes of the retinoid treated cells are necessary for differentiation.

Table. 4 Spectral differences when comparing spectra recorded from retinoid treated cells and pluripotent stem cells using PC-LDA; 5, 7 and 9 day cell sample spectra.

<table>
<thead>
<tr>
<th></th>
<th>LDA 1: major loading bands present in the LDA 1 loading plot</th>
<th>LDA 2: major loading bands present in the LDA 2 loading plot</th>
</tr>
</thead>
</table>
| **DMSO vs. EC23 vs. AH61 vs. ATRA cell spectra; spectra recorded from 9 day cell samples** | -1032, -1061, -1175, -1420, -1460, -1474, -1557, -2847, -2915 and -3068 cm\(^{-1}\)  
+1385, +1451, +1466, +1630, +1640, +1674 and +1693 cm\(^{-1}\) | -1092, -1236, -1385, -1400, -1458, -1474, -1541, -1630, -1692 and -2961 cm\(^{-1}\)  
+1032, +1061, +1171, +1410, +1451, +1466, +1551, +1643, +1673, +2849, +2920 and +3067 cm\(^{-1}\) |
| **DMSO vs. EC23 vs. AH61 vs. ATRA cell spectra; spectra recorded from 7 day cell samples** | +1092, +1236, +1420, +1441, +1452, +1466, +1543, +1630, +1693, +1717, +1734 and +2853 cm\(^{-1}\)  
-1032, -1061, -1082, -1155, -1551, -1638, -2920, -2957 and -3065 cm\(^{-1}\) | +1451, +1466, +1547, +1630, +1690 and +2958 cm\(^{-1}\)  
-1030, -1061, -1080, -1155, -1238, -1474, -1638, -2859, -2920 and -3066 cm\(^{-1}\) |
| **DMSO vs. EC23 vs. AH61 vs. ATRA cell spectra; spectra recorded from 5 day cell samples** | +1061, +1082, +1155, +2849 and +2920 cm\(^{-1}\)  
-1552, -1693 and -2959 cm\(^{-1}\) | +1552, +1634, +1656, +1693, +2849 and +2959 cm\(^{-1}\)  
-1032, -1061, -1082, -1155, -1475 and -2920 cm\(^{-1}\) |
Fig. 47 PC-LDA scores plots (LDA 1 vs. LDA 2) of DMSO, AH61, ATRA and EC23 cell spectra recorded from the day 7 (a) day 5 (b) and day 3 (c) cell samples. Day 7 LDA classification model produced using 10 PCs, equal to 68% of the spectral variance explained, and PRESS result showed that the 11th PC exceeded 1. Day 5 LDA classification model produced using 6 PCs, equal to 54% of the spectral variance explained, and PRESS result showed that the 9th PC exceeded 1. Day 3 LDA classification model produced using 6 PCs, equal to 47% of the spectral variance explained, and PRESS result showed that the 7th PC exceeded 1.
In Fig. 47 (a-c), LDA scores plots of days 7, 5 and 3 data are presented. LDA scores plots results show some spectral differences between the retinoid treated and the control data from days 7 and 5 samples, but there is very little separations seen from day 3 data; only some EC23 and AH61 day 3 spectra are separated from DMSO/ATRA cluster in scores space.

When comparing LDA investigations of day 7 data to the data recorded for the previous chapter, separations between retinoid treated cell spectra and the control stem cell spectra are small; scores plot separations of retinoid treated and control cell spectra getting worse when comparing earlier time point data. As a result, no cross-validation was performed to for 7, 5 and 3 day recorded spectral data because it can be seen that the created classification models are poor.

When investigating the data recorded from 7 day samples (Sup Fig. 12, supplementary materials), spectral analysis results do show the separated EC23/AH61 cell spectra to have changes to the amide II absorption bands. Comparison of both LDA loading plot and mean spectra revealing that the amide II spectral change is a result of a shift in band position for the separated EC23/AH61 cells; amide II band moving to a lower vibrational frequency when compared to the control cell spectra. Results also show reductions in relative intensities from the 1643 cm\(^{-1}\) \(\beta\)-sheet protein band, the 1030 cm\(^{-1}\) glycogen band, the 1082 cm\(^{-1}\) band (primarily associated with the phosphate vibrations from nucleic acid molecules)\(^{54}\) and absorption bands associated with C-H vibrations of lipid molecules (2859 and 2920 cm\(^{-1}\)); these spectral differences are were also seen when comparing the retinoid treated 9 day cell data to the data recorded from the control cells. When comparing ATRA cell spectra, ATRA cell spectra generally show increases in relative intensities when compared against the other spectra suggesting a different stage of differentiation for the ATRA treated cell spectra.

For LDA of 5 day data, only some AH61 and ATRA cell spectra can be separated from the control cluster of data; separated AH61/ATRA cell spectra showing increases in relative absorption from lipid absorption bands, and a decrease from the amide II absorption band. Absorption bands tentatively assigned to molecular vibrations of lipid molecules were previously shown (chapter 4) to be actively involved in the differentiation process, with
spectral data recorded from retinoid treated EC cells being salient differences when compared to the control cell data recorded from the different sample time points, in agreement with previous research \(^95\). Day 5 spectral analysis is therefore further proof of their activity in cellular differentiation.

For LDA of 3 day data, only a small number of EC23/AH61 cell spectra can be seen to be separated in scores space. The spectral differences causing the separation of some treated cells are absorption intensity increases from bands situated at 1201, 1268, 1450, 1468 cm\(^{-1}\), frequency shift in the amide II absorption band (some retinoid treated cell spectra shifting to a higher wavenumber (1548 to 1552 cm\(^{-1}\)), increases in relative absorption from the \(\beta\)-associated amide I bands (~1630, ~1643 and 1674 cm\(^{-1}\)) \(^54\), and a relative intensity decrease from the 1693 cm\(^{-1}\) and the 2958 cm\(^{-1}\) bands. Some of these spectral differences being seen when investigated the biochemical actions of the different retinoids on the EC cell line at 9, 7 and 5 days thus, possible indication of some treated cells being differentiated/differentiating at 3 days.

When comparing the spectra recorded from 6 h, 24 h, 1 and 2 day cell samples, using both PCA and PC-LDA, no significant spectral differences could be seen between the control stem cells and the retinoid treated EC cells; based on the data results this is an indication that retinoids ATRA, AH61 and EC23 do not induce significant biochemical changes the TERA2.cl.SP12 cells soon after their introduction to cell cultures.

### 5.6 Discussion of the Results Found

The analysis results of the IRMS recorded cell data for this experiment have not provided the same results as the previous chapter, i.e. previous chapter data analysis using LDA generally showing good scores separations of retinoid treated TERA2.cl.SP12 cell spectra when compared to the control spectra, whereas for data recorded for this investigation, data analysis results have generally showed poor separations of retinoid treated and stem cell data. However, for the retinoid treated cell spectra which could be separated from the cluster of control cell spectra in LDA scores space, the spectral differences causing the separations did show similarities to the spectral differences seen from day 7 data (chapter 4), suggesting that these biochemical differences found are necessary for the cellular
differentiation of this EC cell line. As well as poor spectral differences being seen from
day 3, 5, 7 and 9 data, no significant spectral changes could be seen from the spectra
recorded from retinoid treated cells and the control cells at the earlier time points (6 h, 24 h
1 and 2 day cell samples); possible suggesting that little biochemical changes are induced
at the earlier time points after retinoid addition to cell cultures.

One possible reason for the poor analysis results could have been the reduction in retinoid
concentration used, reducing from 10 to 1 µm. This concentration reduction may have
reduced the action of the retinoid compounds, resulting in fewer TERA2.cl.SP12 EC cells
differentiating and differentiation not being started for some EC cells in the cell population
until later treatment time periods. Another possible reason may be that fewer cell spectra
were being compared in this investigation; lower number of cells distributed onto CaF₂
sample substrates, therefore, lower number of cell spectra extracted from the chemical
images recorded from each sample slide - much fewer numbers than the previous
experiment (chapter 4).

Fewer numbers of treated cells will be differentiated/differentiating in cell cultures stopped
at the earlier time points after retinoid introduction. With fewer numbers of cell spectra
being recorded in this investigation, only a snapshot of the cell population biochemistry is
captured thus, greater probability that cells actually differentiating may not be
spectroscopically represented in the data analysis. Therefore, although reducing the
concentration of cells distributed onto the CaF₂ samples improved the single cell finding
function, enabling individual cell spectra to be extracted from single chemical images, it
may have possible had a negative effect on the spectral analysis. This would also suggest
that a large number of cell spectra should be recorded to reflect the biochemical changes of
the cell culture population.

5.7 Evaluation of the RMieS-EMSC Correction Algorithm

The cell spectra recorded for this spectral investigation show greater RMieS than the cell
spectra recorded for the previous spectral investigation, see Fig 48 (a) and Fig. 36 chapter
4 for a comparison. Therefore, one other possible reason for the poor spectral analysis
results seen from this investigation could be the increased RMieS seen from the recorded
cell spectra (Fig. 48 (a)). Fig. 48 (b-c) show the RMieS-EMSC corrected spectra (corrected using 200 iterations and matrigel as a reference spectrum), which should correct the cell spectra for RMieS, leaving the pure cell absorption spectrum as the output.\textsuperscript{76-77, 79, 81} However, it is clear that both Fig. 48 (b-c) exhibit a population of very heterogeneous cell spectra, with some spectra in the recorded population not looking like a typical cell spectrum. The 2600-3700 cm\(^{-1}\) spectral range looks particularly heterogeneous with a lot of variance between individual cell spectra recorded; similar poor and heterogeneous cell spectra can be seen from all individual cell spectra recorded for this chapter investigation. Having some corrected cell spectra which exhibit absorption band structures that are strange when compared to the cell population can be seen as outliers and these outliers can have a negative impact on multivariate analysis results.\textsuperscript{111} Increased spectral variance may also have a negative effect on the LDA multivariate analysis due to the intra-class variance becoming much greater than the inter-class spectral variance explaining the biochemical differences between the different cell phenotypes. Consequently, the poor RMieS-EMSC corrected spectra may have had an influence on the poor spectral analysis results produced from this investigation.

An explanation for the increase in RMieS seen from recorded spectra may be a result of the distribution of single cells onto the IR CaF\(_2\) sample slides; for this experimental investigation the distribution of single cells onto IR sample slides was good, with many single cells without close neighbouring cells, however, this was not the case for the majority sample slides in the previous experiment (chapter 4). An increase in RMieS can often be seen when recording an IR spectrum from more rounded cells.\textsuperscript{76} Certainly a single cell with no neighbouring cells represents a more rounded shape than a single cell with many close neighbouring cells. Therefore, it would seem likely that the distribution of single cells onto IR sample slides has resulted in an increase in RMieS present in the cell spectra.

To investigate why the RMieS-EMSC correction algorithm produces poor RMieS corrected spectra, two cell spectra showing a high level of RMieS have been corrected with the RMieS-EMSC correction algorithm using both 8 and 200 iterations (Fig. 49), these corrected spectra are then compared against the corrected spectrum produced from a cell spectrum containing a low level of RMieS (cell spectrum exhibiting a low level of RMieS taken from the data recorded for the experimental investigation in chapter 4).
Fig. 48 (a) All un-processed cell spectra recorded for the day 7 spectral comparisons (b) the same cell spectra as in (a) now corrected for RMieS using the RMieS-EMSC correction algorithm and 200 iterations (c) the same spectra contained in (b) but cut to the 2600-3700 cm$^{-1}$ spectral range.
All spectra revealed in (a) were corrected using the RMieS-EMSC correction algorithm using 200 iterations.

The blue spectrum: looks like a typical IR spectrum of a biological sample, however, the phospholipid band present in un-processed spectrum is now not present in the corrected.

All spectra revealed in (a) were corrected using the RMieS-EMSC correction algorithm using 8 iterations.

It would seem that increasing the number of iterations from 8 to 200 only worsens problem.

Fig. 49 (a) Two cell spectrums exhibiting a greater level of RMieS (red and green) and a spectrum exhibiting a low level of RMieS (blue) (b) spectra in (a) corrected using the RMieS-EMSC algorithm and 8 iterations (c) using the RMieS-EMSC algorithm and 200 iterations.

Fig. 49 (a) shows the un-processed cell spectra containing RMieS; both red and green showing visible absorption band shifts from both the amide I and II bands and absorption
bands situated in the 2800-3000 cm\(^{-1}\) frequency range. As well as these changes, the phospholipid absorption band, situated in the 1720-1745 cm\(^{-1}\) spectral range, has also been significantly changed as a result of the increased RMieS. These spectroscopic changes as a result of increased RMieS have been explained previously \(^{76}\).

After RMieS-EMSC correction using 8 iterations (Fig. 49 (b)) the blue spectrum looks as if it has been corrected properly showing a flat spectral baseline free from any broad sinusoidal oscillations and no derivative-like distortion on the high wavenumber frequency side of the amide I band; typical signs of RMieS. However, the blue corrected spectrum does now show the phospholipid band to be missing from the corrected, whereas it was distinctly present in the un-processed spectrum. The red and green corrected spectra have strange absorption band features after correction, unlike typical spectra recorded from biological samples in the MIR frequency range; red and green spectra exhibiting an amide I absorption band that has been split in two, the spectral baseline is not free from broad sinusoidal oscillations, a strange dip in absorption is present around \(~1400\) cm\(^{-1}\), the absorption bands typical associated with C-H vibrations from lipid molecules (found in the 2800-3000 cm\(^{-1}\) spectral range) look deformed when compared against the blue spectrum, and all red and green absorption bands present in the 2800-3000 cm\(^{-1}\) spectral range look to have shifted band frequency positions (red and green spectral absorption bands shifted to a higher wavenumber frequency position when compared against the blue corrected spectrum.

Fig. 49 (c) shows the spectra after RMieS-EMSC correction using 200 iterations. The red and green spectra exhibited in the figure now look even stranger after the increase in number of iterations; the blue spectrum looks ok, however, the red and green spectra look very different to a typical MIR spectrum of a biological sample. Therefore, it would seem that increasing the number of iterations used in the RMieS algorithm actually worsens the problem when correcting spectra exhibit a high level of RMieS, resulting in corrected spectra which do not have a typical absorption band structure. As previous stated at the beginning of the chapter, the cell spectra recorded for this investigation were corrected with the RMieS-EMSC correction using an increased number of iterations due to the greater RMieS observed from recorded spectra, and a study by Bassan and colleagues stating that each iteration of the algorithm results in a more accurate correction, closer to the pure absorption spectrum \(^{77, 81}\). However, the results in Fig. 49 actually show that for some recorded cell spectra the increase in iterations actually worsens the problem
producing strange cell spectra. A reason for this happening may be a result of the corrected spectrum from the first iteration of the RMieS-EMSC correction algorithm becoming the reference spectrum for the next loop of the RMieS-EMSC algorithm. If the corrected spectrum from the first iteration is poor then the corrected spectra produced from further iterations will also be poor, progressively getting worse. This may be an explanation as to why the corrected spectra look worse with increasing iterations, rather than a more accurate correction, closer to the pure absorption spectrum.

The RMieS-EMSC algorithm has a number of correction options, such as the range of the scattering particle radius, the number of iterations, the number of PCs, i.e. the number of PCs containing decomposed scatter scattering curves information to be used in the algorithm and the range of the scattering particle average refractive index. These correction options are used to produce the number of potential $Q$ scatter curves, i.e. each particle radius size and average refractive index value will produce a unique $Q$ scatter curve thus, by having a large ranges for both options will produce a large number of unique potential $Q$ scatter curves, which are then decomposed using PCA as a way of reducing the scatter curve information in small PC features; least squares regression is then used to find the $Q$ scatter curve which best represent the RMieS present in the IR spectrum $^{77, 79, 81}$.

The typical recommended correction parameters for the RMieS-EMSC algorithm are 2-8 $\mu$m for the scattering particle radius, 1.1 to 1.5 for the average refractive index of the scattering particle, and as previously discussed, 7-8 PCs of scatter information to be entered into the algorithm. It is believed that the majority of RMieS is caused by the cell nucleus $^{81}$ when being interrogated by a MIR beam. A typical cell nucleus may be between 6-8 $\mu$m in radius from a eukaryotic cell. However, the cytopinning method does cause the cells to round up after distribution to sample slides. Therefore, the combination of the more rounded shape of the individual cells and the good distribution onto sample slides (very few neighbouring cells thus, more rounded in shape when the light interacts) may have resulted in RMieS being predominately caused by the edges of the whole cells. Cells typically have a greater radius than 8 $\mu$m, possible between 20-30 $\mu$m. Also, is all of the scatter information retained in the first 8 PCs?
Fig. 50 (a) Cell spectra from Fig. 49 (a) corrected for scattering using the RMieS-EMSC algorithm with the first 10 PCs selected and 3 iterations. Fig. 50 (b) corrected using 30 pcs, 3 iterations, the upper scattering index increased to 1.8 and the upper scattering diameter increased to 15 microns.

To investigate whether changing the recommended correction parameters has an effect on the corrected spectra produced by the RMieS-EMSC algorithm, the same spectra shown in Fig. 50 (same colour labelling) are corrected using the algorithm with the number of PCs extended to the first 10; only 3 iterations of the RMieS-EMSC (Fig. 50 (a)). Fig. 50 (a) shows that through increasing the number of PCs the red and green corrected spectra have been improved when compared to the corrected spectra displayed in Fig. 49; corrected spectra representing more typical absorption band structures seen from typical IR spectra recorded from biological matter. However, RMieS scattering is still present (broad oscillating baseline still present) in the red and green corrected spectra and the absorption bands in the 2800-300 cm\(^{-1}\) spectral range are still shifted in terms of the frequency of absorption positions; red and green spectral absorption bands shifted to a higher wavenumber frequency position when compared against the blue corrected spectrum. As
well as this, the broad absorption in the 3100-3600 cm\(^{-1}\) spectral range has become untypically elongated.

In Fig. 50 (b), cell spectra have been corrected using the first 30 PCs as well as increasing the upper scattering particle radius to 20 µm, possible a more accurate radius range to the cellular population under investigation. The corrected spectra produced when using these altered correction parameters now look much better, with the red and green spectra now looking much more like the blue corrected cell spectrum. Interestingly, both the red and green spectra absorption bands in the 2800-3000 cm\(^{-1}\) spectral have the same wavenumber frequencies band positions as the blue spectrum thus, the change in correction parameters has fixed this frequency shift as a result of RMieS. Also, corrected spectra now have flat baselines, absorption bands now look to be well structured and curiously, the small absorption band at ~970 cm\(^{-1}\) has become more prominent for all corrected spectra.

Consequently it would seem that the RMieS-EMSC correction should not be treated like black box, with the recommend correction parameter values not being appropriate for all cell types and environments. To my knowledge, this is the first time that the RMieS-EMSC correction algorithm has ever been investigated in this manner and investigation results show that through tweaking the correction parameter values better quality cell spectra can be produced, which may lead to better multivariate analysis results.

### 5.8 Conclusions

To conclude, the analysis results of this experimental investigation are poor when compared to the results from the data recorded in the previous chapter, with no real spectral differences being seen from the spectra recorded from treated cells at the earlier time points after retinoid addition to cell cultures. As well as this, poor spectral differences were seen when comparing spectra recording from retinoid treated cells and the control stem cells at the later time points, which in the previous chapter produced good separation in agreement with flow cytometry results. In addition, it would seem that the cell spectra recorded in this investigation have a greater level of RMieS scattering, which may have resulted in poor RMieS corrected spectra when using the RMieS-EMSC correction algorithm. A combination of the low number of cell spectra being recorded from IR sample
slides when compared to the previous data analysis, reduced retinoid concentration used to induce the differentiation process and poor RMieS corrected cell spectra may have resulted in the poor spectral analysis results witnessed. Spectral analysis results would suggest that TERA2.cl.SP12 cells treated with ATRA, EC23 and AH61 show no biochemical changes at the earlier time points after retinoid addition to cell cultures. However, because poor spectral differences were also seen at the later time points it is difficult to say whether this is actually the case. However, through altering the correction parameter values of the RMieS-EMSC algorithm the corrected spectra can be improved, possible producing corrected spectra more like their pure absorption spectra; extending the number of PCs and scattering particle radius possibly holding the answer to problems witnessed in the investigation. Therefore, investigation results highlight the continued problem of RMieS present in single cell spectra recorded in air. Although the RMieS-EMSC correction algorithm has been championed as the answer to RMieS, it is clear that there are still issues with RMieS when recording IR spectra from cells in air and further research is needed to accurately remove RMieS from all recorded cell spectra.

If this experiment was performed again, the number of cell spectra extracted from the chemical images should be maximised to capture a good statistical number of cell spectra for analysis, and the correction parameters of the RMieS-EMSC algorithm will be altered as demonstrated. Increasing the number of cell spectra extracted from each sample image could be achieved by either recording more spectral images from each cell sample, or seeding a high number of cells onto the calcium fluoride substrates and find a better method to extract the single cell spectra from the chemical mages recorded; cells being formalin fixed after they have attached to the surface of the IR sample slides. Seeding living cells to CaF$_2$ sample slides will provide a high density of cells onto the sample surfaces with lower numbers of cell clumps present. However, this method will still result in cells being situated close to neighbouring cells on the sample slide, which means there will be no clear absorption gap between cells in recorded images and the single cell finding function described will struggle to distinguish/extract single cell spectra. A computational processing approach employed by Filik and colleagues$^{112}$ has effectively been shown to solve this problem though, separating single cells from their cell neighbours and ensuring only single cell spectra is extracted. Therefore this processing method should be employed to extract high numbers of single cells from spectral images, taking advantage of the speed advantages of spectral imaging with also ensuring single cell spectra are used in the IR analysis.
Chapter 6:

The Recording of IR Spectra from Single Cells Immersed in Aqueous Solution

6.1 Abstract

If IR spectroscopy is ever going to reach its potential for monitoring cell phenotype, then there needs to be studies where IR spectroscopy is used to investigate different cell phenotypes of living cells, with also studies elucidating the problems involved with spectroscopically investigating hydrated cells which is an environment much closer to the natural environment of living cells. Although recording IR spectra from living cells in transmission mode is difficult due to the strong absorption of water when using IR spectroscopy, it is possible if an adequate water barrier is maintained to reduce water absorption so as to stop the bending vibration of the water molecule becoming saturated at the detector. To do this, micro-devices producing reproducible water pathlengths are needed. This study indeed shows that coupling both a micro-device with SR-IRMS and using an appropriate processing methods to model and remove the concentration of aqueous water contributing to the hydrated cell spectrum does provide a methodology where hydrated cell spectra can be recorded using IR. The study also shows the spectral changes of both formalin fixed and living cells when investigated using IR spectroscopy. Importantly, this may be the first time that has been shown that formalin fixed cell also show increased nucleic acid absorption signal, similar investigations of living cells. Finally, the study presents a new processing method to accurately model the aqueous buffer contribution contained from a hydrated cell spectrum, results possible revealing that the new buffer removal function described is possible better at predicting the aqueous buffer contribution than other processing methods successfully used, with living ATRA and DMSO treated TERA2.cl.SP12 cells being discriminated using IR spectroscopy; spectroscopic results showing nucleic acid absorption bands to be the most salient bands providing the spectral separation thus, the increase in nucleic acid signals from hydrated cell spectra may possible provide clearer spectral biomarkers of cellular differentiation for this cell line, and possible other cell phenotypes. Further work is needed though to fully corroborate the data results of this study.
6.2 Introduction and Aims

Chemical fixation using formalin solution provides a method for preserving the biochemistry of the eukaryotic cells for spectroscopic measurement. However, a study by Vaccari et al. found chemical differences between spectra recorded from live U937 leukemic monocytes that had been air dried and chemically fixed using formalin and ethanol. Spectroscopic results revealing that the use of both chemical fixatives and air drying all had an effect on the chemical composition of the eukaryotic cells when compared to the spectra recorded from living U937 leukemic monocytes; air dried and chemically fixed cells showing significantly less nucleic acid contributions when compared to the spectra recorded from living cells. Authors theorised that the reduction in intensity from absorption bands associated with nucleic acid molecular structures possible signified the denaturing of both DNA and RNA molecules contained within the eukaryotic cell as a result of the air drying and chemical fixing procedures. Therefore, studies where the chemical changes of nucleic acids are of interest could be compromised when using either air dried or chemically fixed eukaryotic cells. Chemical fixation is also a form of cell death and once cells have become chemically fixed they become un-viable. Critically, cell phenotype should be understood from living cells with the biochemistry being maintained after information is recorded so that single cells can retain their biological function. This may then enable a methodology process where the detection of pluripotent stem cells from a differentiating cell population is possible without compromising the viability of the cell population, therefore, reducing the risk of tumour development after cell transplantation during regenerative therapy studies.

By coupling IRMS with a micro device, a single cell spectrum can be recorded of a biological cell in a small volume of aqueous solution, sandwiched between two infrared windows. When doing this there is a closer matching of refractive indexes between biological cells and water than when compared to that of air and biological cells, thus, much reduced RMieS.

Through using this sampling method, a number of studies have been successful in investigating proteins and living cells in aqueous buffer solution using IR spectroscopy. Therefore, as well as living cell investigations using IR being a more appropriate model to the real-life goal of using IRMS for monitoring cell phenotype, the
A previous study by Bambery et al.\textsuperscript{118} has also investigated the use of the RMieS-EMSC. Study results showed that using an average RMieS corrected spectrum as a reference spectrum produced corrected spectra which gave closer correspondence to the original spectra than when using a Matrigel spectrum as a reference. Therefore, the study result by Bambery and colleagues raises more questions about the setup of the RMieS-EMSC correction algorithm. As a consequence, there may be more confidence in the recorded cell spectra when using the micro-device sampling method than RMieS-EMSC corrected spectra. However, when recording an IR signal from a cell immersed aqueous buffer solution there is the obvious problem of very strong water contributions to the overall recorded spectrum (Fig. 51).

Water has three absorption bands in the MIR spectral range with two of them overlapping absorptions that would normally be seen when recording a spectrum from biological material (Fig. 51 and 52). The first is the largest absorption band for water in the MIR
spectral range and centred at ~3400 cm\(^{-1}\) (Fig. 52 (a)). This absorption band is a combination of the symmetric stretching vibration (\(v_1\)), asymmetric stretching vibration (\(v_3\)) and the overtone of the bending vibration (\(v_2\)) of water molecules. The second is the weakest of the three, and appears as a broad ‘bump’ in the baseline centred at ~2127.5 cm\(^{-1}\) (Fig. 52 (c)). This band is a combination of the (\(v_2\)) vibrational mode and a water librational mode. Since it appears in a region of the spectrum containing no relevant biological information its presence is not a particular problem. However, the third water band occurs at ~1643 cm\(^{-1}\) and is the result of the bending mode vibration of the water molecule (\(v_2\)) (Fig. 52 (b)). The bending mode absorption of the water molecule overlaps the amide I and partially the amide II absorption bands, produced through absorption of mainly protein macromolecules contained within a biological sample. Therefore, it is important to be able to accurately remove the water contribution from an IR cell spectrum and a number of published papers have incorporated different processing methods to achieve this \(^{114-117}\).

One method (method 1) of removing aqueous water contributions from a recorded cell spectrum is to simply collect a background spectrum through the aqueous buffer solution in a location which is close to a cell in the micro device, applying apertures in the microscope roughly the size of the average cell. Then a cell plus buffer spectrum is recorded in close vicinity to the location of the background spectrum. By subtracting the buffer spectrum from the cell plus buffer spectrum a pure cell spectrum is left. This method is relatively simple and will produce a cell spectrum. However, this method does not consider that the buffer absorption spectrum will contain more water contribution than that of the cell plus buffer spectrum. This is clearly shown in Fig. 52; buffer spectrum displaying greater absorption from the \(v_1 + v_2 + v_3\) water combination band and the \(v_2 +\) librational combination band when compared to the cell plus buffer spectrum. Therefore, by subtracting the aqueous buffer spectrum from the cell plus buffer spectrum over-correction of the water absorption band at ~1643 cm\(^{-1}\) will occur, resulting in a reduction in the amide I absorption band as a result of over-subtraction \(^{117}\). Nevertheless, the resulting pure cell spectrum produced from this method can be used for analysis as long as the amide absorption bands are not included in the spectral analysis.
Fig. 52 (a) The \((v_1 + v_2 + v_3)\) water combination band (combination of symmetric stretching vibration \((v_1)\), asymmetric stretching vibration \((v_3)\) and the overtone of the bending vibration \((v_2)\) of the water molecule) (b) the \(v_2\) water bending mode absorption band (c) the \(v_2 +\) librational mode combination band. Spectrum recorded of buffer solution (black) and a spectrum recorded of a single cell in buffer solution (Raw cell plus buffer spectrum) (blue).
Another processing method (method 2) has successfully been used in the literature to remove the buffer contributions from the recorded cell spectra. This method incorporates a background spectrum recorded in air, a cell plus aqueous buffer spectrum and a buffer spectrum in close vicinity of the cell being recorded. The buffer spectrum is then removed from the cell plus buffer spectrum, as with the previous method. However, in this method a 1:1 ratio subtraction is not assumed. A scaling factor is applied based on the differences in absorption peak height of the water v$_2$ + librational combination band at ~2125 cm$^{-1}$, for both the buffer and the cell plus buffer spectra. Since this water absorption band appears in a region of the spectrum containing no relevant biological information, it can therefore be used as a measure of how much water absorption is contained in the raw cell plus buffer spectrum. This processing method has been successfully used in a number of studies $^{114-116}$. However, it must be understood that in order for any processing method to work, the absorption due to the bending vibration of the water must not become saturated. To achieve this, an experimental pathlength in the micro-device of 8 µm or less is recommended $^{119}$. It should also be mentioned that the problem of the experimental pathlength being less than 8 µm can be partially avoided using Attenuated Total Reflectance (ATR). ATR-FT-IR has been successfully used in a number of studies investigating living cells $^{120-121}$. With ATR-FTIR, the IR beam internally reflects off a crystal with a high refractive index Internal Reflection Element (IRE) after being directed into the crystal at a certain angle. As the beam comes into contact with the sample, an evanescent wave is created at the sample surface. The reflected IR beam is attenuated as certain frequencies of IR being are absorbed by the sample; the attenuated beam is then measured at the detector and analysed $^{51,53}$. The evanescent wave has a typical penetration depth of $<1$ µm; this is dependent on wavelength though $^{53}$. Penetration depths of less than 1 µm certainly reduce the water absorption contributions in the recorded spectra as a result of much reduced optical pathlength, however, it does not completely solve the water problem as water absorption contributions will still exist in ATR-FT-IR recorded cell spectra recorded, and the penetration depth may not be sufficient enough to probe all of the macromolecules present in cells thus, valuable biochemical information not being recorded as a consequence. Therefore, transmission IR spectroscopy will be used in this study coupled with micro-devices to provide suitable experimental pathlengths to minimise the water absorption contributions in raw cell plus aqueous buffer spectra.

Although method 2 is a well-established processing method and gives satisfactory results in the literature, in this study a third processing method is described. This new processing
method employs a least squares fitting algorithm to accurately model the buffer contribution contained within in a raw cell plus buffer spectrum. Therefore, the aim of this study is to further demonstrate that single cell can spectra can be recorded from an aqueous buffer solution, when coupling IRMS with microfluidic devices. Additionally, the study will also evaluate a new processing method for accurately removing the water contributions from recorded raw cell buffer spectra against current methods successfully used provide pure cell absorption spectra. To do this, firstly formalin fixed TERA2.cl.SP12 cells will be investigated as a proof of concept study, and then living TERA2.cl.SP12 cells will be investigated.

6.3 Cell Sample Preparations

6.3.1 Formalin Fixed Cells
For this study, formalin fixed TERA2.cl.SP12 cells used in chapter 4 investigations were used. As day 7 cell samples provided good spectral differences they were chosen to be used in this study; for the cell culture protocol see 4. 2. 1.

6.3.2 Live A498 cells
In this study, living A498 cell spectra processed using the aqueous buffer removal function described are compared against raw cell plus buffer spectra recorded from formalin fixed TERA2.cl.SP12 cells (the same formalin fixed cells used in Chapter 4 (day 7 cell samples), which have also been processed using the aqueous buffer removal function. The A498 cells were kindly provided the Paterson Institute of Cancer Research, with cell culture being performed by myself and Dr. Caryn Hughes at Manchester University. The A498 cell line is a kidney carcinoma cell line and cells are adherent. Cells were cultured using a media broth containing DMEM (Sigma) supplemented with 1% L-glutamine and 15% of FCS (Sigma) in T-25 BD Falcon flasks. Cultures were kept in a humidified atmosphere containing 5% CO$_2$ at 37°C until roughly 75% confluent. Cells were removed from the culture flasks using a cell scraper.

6.3.3 IRMS Investigation of Stem Cell Differentiation using Living TERA2.cl.SP12 Cells
TERA2.cl.SP12 cells were kindly provided by Prof. Stefan Przyborski at Durham University $^{122}$. TERA2.cl.SP12 cells were brought up from frozen and then split into four
T-25 BD Falcon flasks and left for 24 h to settle down and attach to the flask surfaces. After 24 h, two of the flasks were treated with DMSO (retinoid vehicle) and the other two with ATRA, the final DMSO and ATRA concentration being 10 µm; the culture media used has previously been explained in 4. 2. 1. Treated TERA2.cl.SP12 cells were then left for 3 days. After this time period cell populations had grown to a high density and as a result cell cultures were passaged and transferred to T-75 BD Falcon flasks; a cell scraper was used to remove attached TERA2.cl.SP12 cells. Cell cultures were then left for a further 4 days, with ATRA treated TERA2.cl.SP12 spreading out on the flask surface and changing shape as a result of the retinoid addition to cultures; an indication that differentiation had started for the majority of treated cells 7 days after ATRA addition. Cells were then removed from their flasks, again with the use of a cell scraper, and cell numbers were determined using a haemocytometer. Results showed that DMSO treated cell cultures contained approximately 12 million cells per 1 ml of media and ATRA treated cell cultures contained approximately 1 million per 1 ml; the reduction in cell numbers being an indication that cellular division had slowed down as a result of cellular differentiation, for ATRA treated TERA2.cl.SP12 cells.

Through some prior experimental tests with fixed cells, it was found that a only large clumps of cells contained in the micro-static cell provided an absorption signal with acceptable S/N when using the Globar light source. Therefore, relatively high densities of live cells were needed for the live cell experiment. The cells were seeded onto CaF$_2$ using 100 µl of media. Through experiment, 100, 000 cells per 100 µl of culture media provided the best cell concentration for seeding. Cells were then left for 1 h in media to allow cells to attach to CaF$_2$ surfaces. To remove the media from the top of the CaF$_2$ substrates the cell samples were placed in a well plate and washed with PBS twice. Once dry, 1 µl of PBS was added to the cells along with the 6 micron PTFE spacer and another CaF$_2$ substrate placed on top; this is the setup for the micro-static device.

**6.4 Materials and Methods**

**6.4.1 Micro-fluidic Device Design**

Micro-fluidic device used in this study was designed and created in-house by Francis Ball and Stephan Mohr at Manchester University.
The device was designed to transport the cells from an input reservoir to a point where they could be investigated with the IR beam. Once sufficient spectra had been recorded, cells were then transported out of the device into an output reservoir. To achieve an 8 µm experimental pathlength, a spacer unit made from a photoresist film (experimental positive photoresist (X AR-P 3100/10) manufactured by Allresist in Germany designed specifically for increased adhesion) was sandwiched between two Calcium Fluoride (CaF$_2$) infrared windows (Bottom: 13 × 1 mm diameter and thickness respectively, Top: CaF$_2$ cut to 6 mm square, 1 mm in depth) (Fig. 53 (a)); CaF$_2$ was selected as the substrate because of its low solubility in water, consequently the material was less likely to deteriorate under routine use with aqueous buffer solution. The photoresist film was accurately spin-coated onto the bottom CaF$_2$ window at 7500 rpm, coating the surface of the window to thickness of 8 µm. The positive photoresist film and substrate were then baked causing the resist to polymerise and harden. An accurately defined channel was then removed from the photoresist using UV lithography (exposure was 20 milliwatts per square centimetre (mW/cm$^2$) for a period of 90 seconds (s)) £124. However, since a single cell is larger than 8 µm, prolonged compression of the cell across the entire channel may result in forced biochemical changes to a cell. One example of this was demonstrated by Birarda et al.£114 who studied the response of monocytes to compression forces and found that such forces can indeed induce biochemical changes to a cell. For this reason, the depth of both the inlet and outlet channels were increased to 60 µm so as to ease the mechanical stress experienced by the cells, resulting in a raised analysis plateau that was significantly higher than the floor of the channel. The gap between the analysis plateau and the channel floor was bridged using a channel ramp with a shallow incline (Fig. 53 (b)). This was made using a CAT3D M6 CNC milling machine (Datron Technologies Ltd., Milton Keynes, UK) with a 0.3 mm diameter tungsten carbide milling tool (Toolex Ltd, Trowbridge, UK). Finally, the top window was heat bonded onto the photoresist at 120°C for 72 h.

The design of the device was such that the cells could be transported from a cell input reservoir to a point where they could then be probed with an IR source beam on the analysis plateau (Fig. 53 (c)). The cells were delivered to the device in aqueous buffer using a micro-infusion pump with accurately controlled low flow rates (Cole-parmer 74900 series syringe pump). Through the use of a microscope (32x lens), once cells were seen passing along the analysis plateau of the device, the flow was stopped so that cells would be situated on the analysis plateau for IR measurement. Once sufficient spectra had been recorded, cells are then transported out of the device into an output reservoir. During
routine use of the micro-fluidic device it was noticed that the channel depth within the analysis plateau slightly increased due to pressure being exerted on the device as a result of the flow of fluid traversing through.

**Fig. 53** (a) Schematic of the micro-fluidic device (b) graphical display of the raised analysis plateau situated in the centre of accurately defined channel in the photoresist (c) side on view of the analysis plateau with cells flowing through the device.
The increase, although very small, did result in the saturation of the (ν₂) bending water absorption band at the detector. This problem was resolved by the addition of a metal plate that was screwed in to the device holder (Fig. 3 (a)), the increased in downward pressure on the upper CaF₂ substrate was sufficient enough to counter this effect. This gave the final device design shown in Fig. 53 (a). The device holder was constructed so that the device could fit on both the Bruker and Agilent FT-IR microscope stages used for data collection in this project.

6.4.2 Micro-static Device

A commercial available micro-static device was used for the live cell study using the benchtop IRMS instrument. The reason for this being that the in-house developed micro-fluidic device previously described in this chapter was not structurally reliable to be used for all experimental investigations. Therefore, a commercial micro-static device was purchased to reproducible produce an experimental pathlength of less than 8 µm. The device requires two spherical IR substrate windows and a spacer; the spacer has a hole in the middle and sits between the two IR windows, effectively sandwiching the biological cells (biological cells seeded to the surface of one CaF₂ substrate and left for an hour in media; just before recording data, media removed and cells washed with PBS and the surface is air dried) and a small volume of aqueous buffer solution creating a small optical pathlength. A spacer with a 6 µm thickness was used therefore, providing an optical pathlength of 6 µm; the 6 µm spacer was made from Polytetrafluoroethylene (PTFE) and supplied by Harricks Scientific Products Inc.

6.4.3 Synchrotron Fourier Transform Infrared Microspectroscopy (SR-FT-IRMS or SR-IRMS) Instrument and Experimental Setup

For the experiment investigating formalin fixed TERA2.cl.SP12 cells (EC cells treated with DMSO for 7 days vs. EC cells treated with EC23 for 7 days) using the in-house developed micro-fluidic device coupled with IRMS, synchrotron light was used as the IR source radiation. The synchrotron infrared microspectroscopy (SR-IRMS) was carried out using the infrared beamline (MIRIAM) at the Diamond Light source, Oxfordshire. Data was recorded on a Bruker Vertex 80V vacuum spectrometer coupled to a Hyperion 3000 infrared microscope. The infrared microbeam from the synchrotron was applied via 36x objective/condenser optics. To compensate for the problem of CaF₂ and water dispersion, the IR illumination path was optimised for enhancing the signal in the 1800-1300 cm⁻¹
spectral range. The microscope was also set up in a quasi-confocal arrangement (improving sensitivity and S/N) with front and back apertures set to $14 \times 14 \, \mu m^2$, approximately the size of a single cell. The microscope is equipped with a mid-band high sensitivity MCT detector. Both the background and sample spectra were recorded at $4 \, cm^{-1}$ spectral resolution and adding 512 scans.

### 6.4.4 Attenuated Total Reflectance Fourier Transform Infrared Micro-spectroscopy (ATR-FT-IRMS or ATR-IRMS) Instrumentation

ATR-IRMS was used to investigate the spectral changes between cell spectra recorded in aqueous and dry environments and to provide an RMieS free cell spectrum when investigating difference reference spectra used in the aqueous buffer removal function. ATR spectra being recorded on a FTS 7000 series Digilab spectrometer coupled with a Varian 600 UMA FT-IR microscope and using a Varian slide on micro-ATR Germanium accessory. Cell spectra recorded using 128 co-added scans and $4 \, cm^{-1}$ spectral resolution.

### 6.4.5 Benchtop Fourier Transform Infrared Micro-spectroscopy (FT-IRMS) Instrumentation and Experimental Setup

Data was recorded from live DMSO and ATRA treated TERA2.cl.SP12 cells using a benchtop Varian 670 FTIR spectrometer interfaced with a Varian-620 imaging infrared microscope. The microscope was equipped with a MCT single point detector. The living cell spectra were collected in transmission mode (256 co-scans at $4 \, cm^{-1}$ spectral resolution) coupled with the commercially bought micro-static device; the background spectra being recorded from a clean CaF$_2$ substrate slide (background was recorded using 512 co-scans at $4 \, cm^{-1}$ spectral resolution on a 2mm CaF$_2$ IR window). The beam throughput was set to 100% (no attenuation).

### 6.4.6 Aqueous Buffer Removal Function

The aqueous buffer removal program (method 3) is an in house function created using MATLAB. The function incorporates the ordinary least squares (OLS) fitting algorithm to model the correct amount of aqueous buffer absorption contained within a raw cell plus buffer spectrum. OLS is a regression algorithm that finds the best linear fit through minimising the sum of squared errors (residual ($r$)) between the two spectrums.
To model the correct amount of buffer absorption contained within a cell plus buffer spectrum, a raw cell plus buffer spectrum \( (Y) \) (Fig. 54 (a)) and a scatter free Matrigel reference spectrum \((X_1)\) (Fig. 54 (b)) and a buffer spectrum \((X_2)\) (Fig. 54 (c)) are compared using the OLS algorithm. Then the OLS algorithm is used to fit both \( X \) spectra to the \( Y \) spectrum. Firstly, \( X_1 \) is fitted to \( Y \) and then \( X_2 \) to \( Y \), producing a scaling coefficient \((b)\) for both \( X \) spectra; \( b_1 \) for the Matrigel reference spectrum and \( b_2 \) for the buffer spectrum (Eq. 9 (a)) (it should be explained that the spectra were cut to the 1500-1700 cm\(^{-1}\) spectral range before OLS comparison; OLS comparing spectra over this range). The aqueous buffer spectrum is multiplied by \((b_2)\) to produce a predicted buffer spectrum \((\tilde{Y})\) (Eq. 9 (b)) that models the water absorption contribution contained in the recorded raw cell plus buffer spectrum. The predicted \( \tilde{Y} \) is then subtracted from the \( Y \) spectrum to produce a pure cell absorption spectrum, free from aqueous buffer contributions (Fig. 54 (d))\(^{126-127}\).

\[
b = \frac{(Y-r)}{X} \quad \text{Eq. 9 (a)}
\]

\[
\tilde{Y} = (X \times b) \quad \text{Eq. 9 (b)}
\]

The Matrigel reference spectrum is incorporated into the aqueous buffer removal function so as to ensure that the aqueous buffer spectrum is fitted to the raw cell spectrum \( (Y) \) in a manner which provides reasonable spectral differences between the two spectra in the amide I and II spectral range, i.e. sufficient differences so as to produce a pure cell absorption spectrum with a good amide I and II absorption band ratio, like seen from the Matrigel reference spectrum. By doing this, a good model of the aqueous buffer contribution in the recorded cell spectra is seen. This is illustrated in Fig. 55 where (a) shows a raw cell plus buffer spectrum (blue), the predicted \( \tilde{Y} \) buffer spectrum produced when incorporating the Matrigel reference spectrum into the algorithm (red), and the predicted \( \tilde{Y} \) buffer spectrum produced when no reference spectrum is used (black). Results in Fig. 55 (a) clearly shows the red predicted buffer spectrum provides a better fit to raw cell plus buffer spectrum than when it is not included in the OLS comparison; the black spectrum overlapping the raw cell plus buffer spectrum in the amide II spectral range (the black spectrum being wrongly fitted high up the \(~1643\) band) and this poor fitting has consequently resulted in there being little spectral differences between the two spectra in the amide I spectral range, whereas the red spectrum fits the raw cell plus buffer spectrum much better, the new predicted buffer spectrum is placed underneath to the amide II absorption band in the raw cell plus buffer spectrum and there is now no overlap of the raw cell and predicted buffer spectrum.
Fig. 54 (a) Raw cell plus buffer spectrum (b) Matrigel reference spectrum (c) aqueous buffer spectrum (d) pure cell spectrum produced as the output of the buffer removal function described.
Fig. 55 (b) shows the resulting pure cell spectra produced by both the red and black predicted aqueous buffer spectra in Fig. 55 (a); the blue pure cell spectrum produced by the red predicted buffer spectrum and the black produced by the black predicted buffer spectrum. When the black spectrum is subtracted from the blue spectrum a poor pure cell spectrum with obvious over-subtraction of the amide I band is produced (Fig. 55 (b)). As well as this, the water $v_2 +$ librational combination band has also been over-subtracted, therefore, suggesting that the predicted black spectrum was an over estimate of the water absorption contained within the raw cell plus buffer spectrum. However, when the red spectrum is subtracted from the blue a flat line is seen where the water $v_2 +$ librational combination band is situated. This suggests that the red spectrum was a perfect model of the water contribution in the raw cell plus buffer spectrum.

Fig. 55 (a) The predicted $\tilde{V}$ spectrum (red) when incorporating the Matrigel reference spectrum, the predicted $\tilde{V}$ spectrum when no reference spectrum is incorporated (black) and the cell plus buffer (blue) spectrum (b) pure cell spectrum produced when a Matrigel buffer is incorporated into the buffer removal program (blue) and when Matrigel is not (black).
6.5 Evaluation of the Aqueous Buffer Removal Function

Fig. 56 (a) describes the spectral differences between the pure absorption spectra produced when using each processing method (method 1 (red), method 2 (black) and method 3 (blue)) on the same raw cell plus buffer spectrum. Method 1, the red spectrum shows some over-subtraction of the amide I band and the water $v_2 +$ librational combination band; an indication that predicted buffer absorption spectrum contains more water contribution than the raw cell plus buffer spectrum. Method 2, the black pure cell spectrum has experienced some under-subtraction of the predicted aqueous buffer spectrum, with the black spectrum showing a strange amide I and II band ratio (amide I band much greater intensity than the amide II band) and some under subtraction of the water $v_2 +$ librational combination band can be seen in Fig. 56 (b). Method 3, the blue pure cell absorption spectrum produced through the use of method 3 has a flat baseline in the water $v_2 +$ librational combination band spectral region and no obvious under/over subtraction seen. Therefore, the aqueous buffer removal function described seems to be better than processing methods 1 and 2 at predicting the correct aqueous buffer contributions found in the raw cell plus buffer spectrum.

One may question the use of Matrigel as a reference spectrum due to it being a mixture of proteins and not containing the complex mixture of diverse molecules that make up a biological eukaryotic cell. However, Matrigel has been purposely chosen to show that Matrigel is a perfectly good reference for the universal buffer removal function described, and is not incorporated into the pure cell spectra produced. It is therefore not necessary to acquire a scatter free cell spectrum when using the function. To emphasise this point, in Fig. 57 (a) two pure cell spectra have been produced using the aqueous buffer removal function and the same raw cell plus buffer ($Y$) and aqueous buffer spectra ($X$), the only difference being that each pure cell spectrum has been produced using a different reference spectrum; one produced when incorporating a Matrigel spectrum as the reference (black), and the other using a RMieS free cell spectrum recorded from formalin fixed TERA2.cl.SP12 cells (blue) (RMieS free cell spectrum recorded from a formalin fixed TERA2.cl.SP12 cell using ATR-FT-IR (6. 4. 4)).
Fig. 56 (a) Pure cell absorption spectra produced from one single cell plus buffer spectrum using all three processing methods; method 1 (red), method 2 (black) and method 3 (blue) (b) blow up of the pure cell spectrum produced using method 2 (water ν₂ + librational combination band spectral region); some slight under subtraction of the water ν₂ + librational absorption band can be seen in the figure.
Fig. 57 (a) Comparison of the outputted pure cell absorption spectra produced using the same raw cell plus buffer spectrum, but different reference spectra: matrigel (black) and a cell spectrum recorded using attenuated total reflectance (ATR) (blue) (b) PCA scores plot (PC1 vs. PC3) of spectra recorded from formalin fixed TERA2.cl.SP12 cells and live A498 cells in aqueous buffer. Buffer removed using both matrigel and an ATR recorded cell spectrum as references (c) mean second derivative spectra.
Fig. 57 (a) clearly shows that reference spectra incorporated into the aqueous buffer removal function have had very little effect on the outputted buffer free cell spectra produced. However, Fig. 57 (a) is only comparing the spectral differences of one recorded cell spectrum. Fig. 57 (b) shows a two dimensional PCA scores plot (PC1 vs. PC3) of 14 TERA2.cl.SP12 cell spectra (the same formalin fixed TERA2.cl.SP12 cell used in chapter 4 (day 7 samples), spectra recorded through coupling SR-IRMS with the commercially bought micro-static device) and 14 living A498 kidney carcinoma cell spectra (again, spectra recorded through coupling SR-IRMS with the commercially bought micro-static device); all cell spectra were recorded in an aqueous environment (phosphate buffer solution (PBS)) and pure cell absorption spectra produced through using the aqueous buffer removal function; the black formalin fixed TERA2.cl.SP12 cell spectra and green live A498 cell spectra represent pure cell absorption spectra corrected using the aqueous buffer removal function incorporating a Matrigel reference spectrum, and blue (TERA2.cl.SP12 cells) and cyan (living A498 cells) pure cell absorption spectra represent cell spectra corrected using the aqueous buffer removal function and a RMieS free ATR-FT-IR recorded cell spectrum as a reference (the same ATR-FT-IR RMieS cell spectrum used in Fig. 57 (a)).

Comparing all pure cell absorption spectra produced using PCA (Fig. 57 (a) (pure absorption cell spectra were transformed to second derivatives to eliminate baseline slopes found in some pure cell spectra) clearly shows that representative spectra recorded from the different cell samples cluster together in PCA scores space. Therefore, even though cell spectra are not exactly the same after incorporating different reference spectra in the function, indicated by most cell spectra produced using the different references not being completely overlapped in the PCA scores space, the spectral differences are small when compared to the spectral differences of the different cell phenotypes thus, different reference cell spectra has little effect on the output spectra of the function, as long as the reference spectrum provides a good amide I and II band ratio. Consequently, Matrigel is an adequate reference spectrum to use for this universal buffer removal function and should be the reference of choice when a cell spectrum free from RMieS is difficult to obtain.

To further evaluate the aqueous buffer removal function cell spectra recorded from two different groups of cells were again investigated, however, this time spectral differences were evaluated from both hydrated recorded cell spectra (cell spectra recorded from
formalin fixed cells re-suspended in PBS using the in-house developed micro-fluidic device described) and cell spectra recorded in air (chemically fixed using 4% formalin and cytospun on to CaF₂ substrates); this time formalin fixed DMSO treated TERA2.cl. SP12 cells (treated for 7 days – chapter 4 data) were compared against formalin fixed EC23 treated TERA2.cl.SP12 cells (treated for 7 days – chapter 4 data) as good spectra differences were seen from these data in the study discussed in chapter 4.

Fig. 58 (a) Comparison of mean cell spectra recorded in air on a CaF₂ slide and recorded in aqueous buffer in the micro-fluidic device (b) comparison of mean cell spectra (shown as the second derivative) recorded in air and recorded in aqueous buffer.

Fig. 58 (a) shows the resultant mean spectra of the dried cells and spectra recorded using the micro-fluidic device with the buffer removal function (incorporating a Matrigel reference spectrum) being applied. As you can see, the spectra recorded of the cells under
the different environments are broadly similar, but importantly, it can be seen that there are some significant differences; amide I band shifts to a lower wavenumber frequency (1658 to 1653 cm\(^{-1}\)) and the amide II shifts to a higher wavenumber (frequency 1547 to 1551 cm\(^{-1}\)) for cell spectra recorded in an aqueous environment. These spectral differences are again highlighted in Fig. 58 (b), which shows the spectra transformed to second derivatives.

It is important to establish if the difference in the amide I and II absorption bands seen in Fig. 58 (a) is due to a problem with the buffer removal model, or if the buffer removal model is fine and the shifts in absorption bands represents real differences in chemistry between the dried cells and hydrated cells. To do this, firstly, an ATR recorded cell spectrum of a number of cells was collected in air. Then, without moving the sample, a drop of aqueous buffer solution was pipetted onto the cells cytopun onto CaF\(_2\) and the ATR spectrum of the same cells recorded both immediately after hydration, and then at several time points after hydration. Data collection continued until the water had evaporated and the cell had returned to its original dried state. The ATR cell spectra (Fig. 59 (a) and (b)) show very similar structural changes to those seen when comparing hydrated cell spectra recorded in the microfluidic device to those of dehydrated cells recorded in air; once the cells have become hydrated there is an immediate shift in both the amide I and II absorption bands when compared to that of cell spectra recorded in air, amide I shifting from 1658 to 1650 cm\(^{-1}\) and amide II shifting from 1518 to 1551 cm\(^{-1}\). It can also be seen that as the cells dry out, both amide absorption bands go back to their original positions. Fig. 59 (c) and (d) show the same ATR recorded cell data as in Fig. 59 (a) and (b) but transformed to second derivatives thus, displaying the amide I and II secondary structures. It can be seen from the secondary structure changes, from dry to hydrated, that it is not so much a shift in absorption band peak position, but a redistribution of absorption intensities which is causing the band shifts. Absorption bands 1516, 1558 and 1630 cm\(^{-1}\) becoming less intense and bands at 1553, 1650 and 1692 cm\(^{-1}\) becoming more intense when the cells are hydrated.
Fig. 59 Spectra recorded of formalin fixed TERA2.c1.SP12 cells, dry and in air, and hydrated in aqueous buffer using ATR spectroscopy. Cell spectrum recorded in air (blue), cell spectra recorded 5 minutes after aqueous buffer added (green), 10 minutes (red), 20 minutes (aqua), 40 minutes (pink) and 2 hours (yellow) (a) Amide II absorption band (b) Amide I absorption band (c) Amide II absorption band transformed to second derivative spectra (d) Amide I absorption band transformed to second derivative spectra.
Fig. 6 Spectra recorded of formalin fixed TERA2.c1.SP12 cells, dry and in air, and hydrated in aqueous buffer using ATR spectroscopy. Spectra transformed to second derivatives. Cell spectrum recorded in air (blue), cell spectra recorded 5 minutes after aqueous buffer added (green), 10 minutes (red), 20 minutes (aqua), 30 minutes (pink) and 40 minutes (yellow).

These spectral changes of proteins in an aqueous environment have been previously witnessed before by Wellner et al.\textsuperscript{128} when investigating the structural changes of hydrated proteins. A number of other studies\textsuperscript{57-58, 129-130} have also seen shifts in absorption bands associated with proteins when investigated hydrated samples using IR, with it being suggested that the shifts are a direct result of hydrogen bonding, i.e. the stronger the hydrogen bond involving the amide C=O carbonyl the lower the electron density in the C=O group and the lower absorbing frequency\textsuperscript{113}. The peak positions of the amide I secondary structure are a direct result of hydrogen bonding, with intermolecular β-sheet proteins containing the strongest hydrogen bonds thus, absorbing at 1610-1625 cm\textsuperscript{-1}, and α-helical proteins, which have weaker hydrogen bonding, absorbing at the higher frequencies of 1653-1660 cm\textsuperscript{-1}\textsuperscript{129}.

Hydrogen bonding is said to lower the absorbing frequency of stretching vibrations because it lowers the bond stiffness, however, for bending vibrations, hydrogen bonding produces additional restoring forces, thus, raising the frequency of absorption\textsuperscript{58}. As the
amide II absorption band is made up of ~60% N-H bending vibrations from protein macromolecules \(^5^1\), this could explain the displacements of the amide absorption bands seen in our results for hydrated cells (Fig. 59). The structural changes seen from cell spectra recorded in an aqueous environment are therefore a direct result of hydration with the differences being due to real chemical differences and not because of a problem with the aqueous buffer removal function. Therefore, any spectral differences between the pure cell spectra produced using the buffer removal function can be confidently seen as true structural changes.

Whelan and colleagues \(^1^3^1\) have previously looked at the spectroscopic changes in DNA structure from hydrated eukaryotic cells when compared to that of dehydrated. Their results confirmed the presence of B-like DNA conformation once eukaryotic cells were hydrated, which was then lost and transformed back to the A-like DNA conformation when dehydrated. The hydrated B-DNA showed much sharper and more pronounced absorptions intensities than the dehydrated eukaryotic cell A-DNA, in agreement with the spectroscopic investigation by Vaccari and colleagues \(^1^0^6\). This can also be seen from the hydrated cell spectra shown in Fig. 60, where hydrated cell spectra show a pronounced increase in absorption at bands 972, 1086 and 1422 cm\(^{-1}\), associated with hydrated B-DNA. Therefore, by hydrating the cells, an enhancement in nucleic acid structural information can be seen when compared to that of cell spectra recorded in a dry environment. Whelan \textit{et al.} \(^1^3^1\) recorded cell spectra from living cells. Interestingly, the results of this study have shown that the same is also true for formalin fixed cells. This dramatic improvement in DNA information recorded from hydrated cells could have significant implications for studies looking at disease and cellular differentiation, death and survival. Our study reveals that this enhancement in DNA information in recorded IR spectra can also be seen from formalin fixed cells as well. To the best of my knowledge i believe this is the first time this has been shown.

### 6.6 Comparison of Hydrated Formalin Fixed DMSO and EC23 Treated TERA2.cl.SP12 Cell Spectra

As a proof of concept study, IR spectra recorded from formalin fixed DMSO and EC23 cells immersed in aqueous buffer solution were compared using PC-LDA to see whether
spectral differences could be found between the different cell types; raw cell plus buffer spectra recorded corrected using the aqueous buffer removal function described. Day 7 cell samples used in the experimental investigation described in chapter 4 were chosen because the data recorded from these samples (6. 3. 1) provided good spectral differences when compared using multivariate analysis, therefore, these cell samples represented good model samples to test and compare the aqueous buffer removal function. All raw cell plus buffer spectra were recorded using synchrotron light (6. 4. 3) and cells were immersed in buffer solution using the micro-fluidic device described (6. 4. 1).

For pre-processing, the pure absorption cell spectra produced by the aqueous buffer removal function were transformed to second derivatives to remove baseline differences between recorded spectra and spectra, spectra were then cut to the 1380-3030 cm\(^{-1}\) spectral range as this range in the spectrum provided acceptable S/N and data was compared using PC-LDA; 40 spectra were recorded from each cell sample (DMSO and EC23 cells) and the first 3 PCs were entered into the LDA algorithm.

The LDA scores histogram displayed in Fig. 61 (a) shows that are spectral differences between the different cell spectra recorded from hydrated cells along the LDA 1 separation boundary. Fig. 61 (b) shows LDA 1 loading plot describing the spectral differences providing the separations LDA 1 and the mean spectral differences between DMSO and EC23 hydrated cell spectra; spectral differences represented in Table. 5. The greatest spectral differences between DMSO and EC23 hydrated cell spectra being from the amide I band; EC23 treated cells showing a decrease in relative intensity from the \(\alpha\)-helix band at 1655 cm\(^{-1}\) and absorption bands associated with \(\beta\)-structural proteins at 1628 and 1638 cm\(^{-1}\), with increases in relative intensity being seen from EC23 treated cells at absorption bands situated at 1633 and 1644 cm\(^{-1}\). Due to cell hydration the frequencies of the amide I secondary structure absorptions have shifted and as a consequence the 1630 cm\(^{-1}\) \(\beta\)-sheet absorption band is now the 1628 cm\(^{-1}\), the 1643 cm\(^{-1}\) is the 1638 cm\(^{-1}\) and the 1637 is the 1633 cm\(^{-1}\) absorption band present in the hydrated cell spectra show in Fig. 61 (b). The comparison of the mean spectra show the EC23 to have reduced absorption intensity from the \(\alpha\)-helix band, however the spectral difference at \(\sim1647\) cm\(^{-1}\) may be indicative of a shift in amide I band for EC23 treated stem cells; the \(\alpha\)-helix band for EC23 treated TERA2.cl.SP12 cells shifting to a lower vibrational frequency when compared to control pluripotent stem cell spectra.
Fig. 61 PC-LDA comparisons of cell spectra recorded at the synchrotron facility in aqueous buffer, spectra recorded from the DMSO and EC23 formalin fixed cells from the 7 day sample cultures. Buffer contribution removed using processing method 3 (a) PC-LDA scores histogram of spectra recorded from DMSO and EC23 treated cells in aqueous buffer, which have been transformed to second derivatives prior to PCA and cut to the 1380-3030 cm\(^{-1}\) spectral range; 3 PCs entered into the LDA classification (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC23 (blue) and DMSO (black)). PRESS/RSS ratio result showed the 6\(^{th}\) PC to exceed 1. Also, the first 3 PCs contained 40% of the spectral variance from the original data set.

Table 5 Spectral differences between formalin fixed hydrated EC23 treated cells and control (DMSO treated) stem cells at 7 days. Data recorded from cells in aqueous buffer, with buffer removed using processing method 3.

<table>
<thead>
<tr>
<th>EC23 (&gt;) increase in absorption or new band (cm(^{-1}))</th>
<th>EC23 (&lt;) decrease in absorption or new band from control cells (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1421, 1468, 1489, 1543, 1558, 1576, 1633, 1647, 1674, 1699</td>
<td>1441, 1475, 1550, 1566, 1589, 1610, 1629, 1639, 1655, 1716, 2853, 2923, 2958</td>
</tr>
</tbody>
</table>
The shifting of the α-helix band to a lower frequency, coupled with decreases in relative intensity from the 1630 cm⁻¹, 1643 cm⁻¹, and the C-H stretching vibrational bands between 2830-3000 cm⁻¹, were all seen when EC23 treated cells spectra, recorded in air, were compared to the DMSO treated cell spectra; day 7 data analysis chapter 4. Therefore, even though a small number of hydrated cell spectra were recorded for this analysis, LDA has shown that there is some separation from recorded cell spectra and some spectral differences are in agreement with spectral differences separating DMSO treated from EC23 treated TERA2.cl.SP12 cell spectra in the study described in chapter 4. However, there are also some spectral differences which are not the same, such as the reduction in absorption in relative intensity from the α-helix band for EC23 treated EC cells. This may be a consequence of the small number of spectra being compared, i.e. not enough cell spectra to represent the global biochemical changes of the cell culture population undergoing cellular differentiation. Nevertheless, results seem to show that the aqueous buffer removal function can produce accurate cell spectra, free from both RMieS and water contributions, allowing the discrimination of different cell phenotypes as a result of their different biochemical compositions.

6.7 The Investigation of Living TERA2.cl.SP12 using FT-IRMS

Although the proof of concept study investigating formal fixed cells with SR-IRMS coupled with a micro-fluidic device and aqueous buffer removal function was successful, and showed that pure cell absorption spectra can be extracted from raw cell plus buffer spectra, the investigation of living cells represents a more challenging prospect due to the fragility of living cells. As a consequence, for this experimental investigation spectra were recorded from living cells quickly after the cell media had been removed from cells attached to the CaF₂ substrates, otherwise, there is the chance that cell apoptosis may start and the spectral differences between recorded cell spectra may be as a result of the cells at different stages of cell death rather than the chemical constituent differences as a result of retinoid induced differentiation.

For this experimental investigation, the hydrated cell spectra recorded from DMSO treated TERA2.cl.SP12 cells are compared against the cellular IR spectra recorded from ATRA treated TERA2.cl.SP12 cells. The spectral analysis results from previous experimental investigations (chapters 4 and 5) have shown that the majority of TERA2.cl.SP12 cells
treated with EC23, AH61 and ATRA for 7 days or more are likely to be biochemical
different as a result of neuronal differentiation. Therefore, to make sure that the majority of
ATRA treated EC cells had differentiated for this investigation, TERA2.cl.SP12 cells were
treated with both DMSO and ATRA for 7 days before spectral data recording using IRMS
(see 6. 3. 3 for live TERA2.cl.SP12 cell culture and sample preparation for this
experiment)

Ideally, with the excellent S/N produced using synchrotron radiation as an IR source, SR-
IRMS is an ideal method for investigating single cells in an aqueous environment.
However, unfortunately SR-IRMS instrumentation was not available for this experimental
investigation. Therefore, data was recorded from aqueous hydrated living cells using a
benchtop IRMS instrument (see 6. 4. 5 for instrumentation used to record cell spectra and
experimental setup). Through some experimental tests using formalin fixed cells (6. 3. 1),
it was found that the Globar IR light source used in the benchtop IRMS was not bright
enough to record data from single cells, i.e. poor single cell IR absorption signal recorded
when recording in single point mode and employing apertures 14 x 14 µm² in size; roughly
the dimensions of the single cells (cells found in the micro-static device immersed in
aqueous buffer and aperture sizes roughly the dimensions of the single cells incorporated
before a single cell spectrum recorded). Consequently only large clumps of cells found in
the commercial bought micro-static cell provided an absorption signal with adequate S/N
when using the Globar IR light source. Therefore, clumps of living cells were investigated
for this experimental study (apertures roughly the size of the cell clumps (∼50 x 50 µm²)
being employed to provide an average IR spectrum for all cells involved in the mass of the
cells). The advantage of this increase in aperture dimensions is that this setup is now not
diffraction limited thus, better S/N seen from the longer wavelengths (1250-900 cm⁻¹) and
a greater amount of light will reach the detector, which will also increase the SNR over the
whole spectral range. Although this meant that it was not possible to understand the IR
spectral differences of individual living stem cells from living differentiated cells, the
investigation would still be able to provide the average biochemical differences of the
retinoid treated EC cells from the pluripotent stem cells.
Fig. 62 (a) Average cell spectrum (blue) and aqueous buffer spectrum (black) recorded in close vicinity on the benchtop FT-IR Micro-spectrometer (b) absorption cell spectrum after the buffer contribution was removed using processing method 3 (clear over subtraction of both the amide I and II bands) (c) comparison of the mean second derivative spectra, ATRA cell spectra (green) and DMSO (black).
Fig. 62 (a) shows an average cell spectrum recorded from a living cell (blue) and an aqueous buffer spectrum (black) recorded using a benchtop FT-IRMS instrument (data vector normalised). Pure cell absorption spectra were produced using the aqueous buffer removal function and Fig. 62 (b) displays one of the pure average cell absorption spectra recorded after removing the aqueous buffer contribution. One of the observations from the hydrated cell spectrum shown in Fig. 62 (b) is that the amide I absorption band has been severely over-subtracted. This is probably due to the aqueous buffer spectrum having a water bending absorption band which has the same absorption intensity as the raw cell plus buffer spectrum, even though both spectra were recorded within a close spatial vicinity to each other in the micro-static cell. From previous results in this experimental investigation it would be expected that the raw cell plus buffer spectrum to have a greater intensity in the 1600-1700 cm\(^{-1}\) spectral range than the buffer spectrum as the amide I absorption band will be added to the absorption signal of the water bending band; even though the buffer spectrum should contain a slightly higher water concentration than the raw cell spectrum, and this has previously been witnessed in this study. However, with both spectra having no spectral differences in this spectral range, when the correction algorithm predicts the aqueous water contribution in the raw spectrum it will not be possible to fit the two spectra so as to provide spectral differences where the amide I and II band ratio will be maintained for the pure cell absorption spectrum, and this is why an over-subtracted amide I band is produced as a compromise (Fig. 62 (b)).

The reason as to why the raw cell plus buffer spectrum and the aqueous buffer spectrum have almost the same absorption intensity in the 1600-1700 cm\(^{-1}\) spectral range may be a result of the absorption bands in this spectral range from both spectra being on the limit of detector saturation. Absorption signal becomes non-linear when on the limit of signal saturation, therefore, there may have been small spectral differences in this spectral range between the recorded spectra but these have not represented due to non-linear signal detection. As a result, it is possible that the amide II absorption band is also over-subtracted due to a poor fitting of the two spectra. Nevertheless, excellent cell absorption information is found from the remaining absorption bands in the pure cell spectrum produced, especially from absorption bands primarily associated with nucleic acids. The results in Fig. 62 are therefore further evidence of enhanced nucleic acid absorptions from hydrated cell spectra.
Fig. 62 (b) also shows the pure cell absorption spectrum to exhibit a sloped baseline, possible a direct result of the poor fitting of the aqueous buffer spectrum to the raw cell plus buffer spectrum recorded; sloped baselines were also exhibited for other pure cell absorption spectra produced using processing method 3, with baselines being varied between the recorded cell spectra. To resolve the baseline issue, live pure cell absorption spectra were transformed to second derivatives. As well as this, due to the over-subtraction of the amide absorption bands the spectral range of the transformed living cell spectra were also cut so that the amide I and II bands were removed from the spectra before multivariate analysis using PCA.

Fig 62 (c) shows the comparison of the mean second derivative spectra, which depicts the average control stem cell spectra to have greater relative absorption intensity from the 968, 1051, 1088 and 1225 cm$^{-1}$ absorption bands; all associated with nucleic acids. Fig. 62 (c) also shows the ATRA treated average cell spectra to have greater relative absorption intensities from the C-H vibrations in the 2800-3000 cm$^{-1}$ spectral range.

Fig. 63 (a) shows PCA score plot results of the living average cell spectra. PCA score plot result shows that there are clear spectral differences between ATRA and DMSO treated live TERA2.cl.SP12 spectra; clear separations along the PC1 linear separation boundary between ATRA and DMSO treated TERA2.cl.SP12 living cell spectra being seen (EC cells treated for 7 days). On investigation of the PC1 loading band and mean second derivative spectra the greatest biochemical differences between the differently treated EC cells come from the 970, 1051 and 1088 cm$^{-1}$ absorption bands, all previously associated with nucleic acid molecules, and C-H molecular vibrations in the 2800-3000 cm$^{-1}$ spectral range, which are tentatively assigned to lipid molecules present in the cell structures. ATRA treated cell spectra showing much reduced relative absorption from these bands when compared to the control cell spectra, in combination with decreases in relative intensity from the 1223 and 1420 cm$^{-1}$ absorption bands, previously associated with B-DNA, and increases at 990, 1024, 1155, 1383, 1468 cm$^{-1}$ absorption bands (Fig. 63 (b)).
In previous studies\textsuperscript{98, 100, 132}, the $\sim$1084 absorption band has been seen as a discriminating spectral marker of stem cells. Notingher and colleagues\textsuperscript{98, 132} theorising that the decrease in $\sim$1084 cm$^{-1}$ absorption band signal from differentiated cell spectra is a sign of mRNA stores of the cells being used as they synthesise the new proteins needed for the new cellular phenotype being produced. Due to the phosphate symmetric stretching vibration
band (1082 – 1084 cm\(^{-1}\)) being absorbed by both DNA and RNA molecules present in a cell, it is difficult to definitively say whether the reduction in the relative intensity of the 1084 cm\(^{-1}\) absorption band is as a result of mRNA stores being used up when using solely IRMS. However, the results in Fig. 62 and 63 would be in agreement with previous studies, with ATRA treated cells showing reduced absorption from the ~1084-88 cm\(^{-1}\) band as well as other bands previously associated with DNA/RNA molecules (970 and 1051 cm\(^{-1}\) absorption bands). Therefore, the results of this live cell investigation are in agreement with previous spectroscopy studies, showing that a reduction in nucleic acids may well be appropriate spectral biomarkers to discriminate differentiating/differentiated stem cells. However, a study by Flower et al. showed that spectral differences of average cell spectra using multivariate analysis is easier than when comparing single cells, with the average cell spectrum representing the biochemical centroid of the cell population interrogated thus, the number of cells in different cell cycle phases in the cell population also affecting the separations as well as biochemical differences as a result of cellular differentiation. Therefore, good spectral separations using PCA should be expected if the raw live cell plus buffer spectra have been corrected properly using aqueous buffer removal function. As a result, the mean spectral differences and PCA loadings should be treated with caution due to the spectral separation being a result of biochemical changes of the group of cells being interrogated and not individual cell spectra. Consequently, spectral biomarkers found as a result of comparing average cell spectra should be further investigated using single cell spectra to confidently find the biochemical changes of single cell as a result of cellular differentiation.

In conclusion, even though the amide I and II bands had to be removed from the live TERA2.cl.SP12 spectra, excellent separation can be seen between the ATRA and DMSO treated cells that left in cultures for 7 days. The enhancement of nucleic acid signals from cell spectra recorded from living cells, possible as a result of DNA switching to the B-DNA hydrated form, has provided good spectral differences between the ATRA treated cell spectra and the pluripotent control cell spectra. These results therefore back up the previous study by Whelan and colleagues, and also shows the potential of investigating hydrated living cell spectra for similar studies; the increase in nucleic acid signal possible providing spectral biomarkers of cellular differentiation. However, only a few live cell spectra are being compared through PCA and as shown by Flower et al. finding spectral differences of average cell spectra is easier than when comparing single cell spectra using multivariate analysis.
6.8 Conclusion

This study has further shown that it is possible to record hydrated cell spectra from cells immersed in aqueous solution when coupling a micro-device with IRMS; that is to say as long as a suitable water barrier is maintained. The study also presents a new function that accurately removes the aqueous buffer contribution from cell spectra recorded in an aqueous environment. The pure cell spectra produced have no scattering, good baselines, and results of this study suggest that the presented function is possibly better than previously used processing methods used to accurately predict the aqueous absorption contributions in recorded raw cell plus buffer spectra. Through the use of ATR-IMS the study also elucidated the spectroscopic changes of cell spectra recorded in different physiological environments, therefore, validating the buffer removal function and providing confidence in the quality of the pure cell spectra produced. Hydrated cell spectra showing not only enhanced signal from absorption bands associated with nucleic acid molecules, something which has been previously seen in the literature, but also spectroscopic changes from amide I and II bands when compared against cell spectra recorded in air. Importantly these spectral changes were even seen from formalin fixed cells.

When coupling SR-IRMS with a micro-device to provide an adequate water barrier, spectral differences could be found between formalin fixed EC23 and DMSO treated EC stem cells immersed in buffer solution; pure cell absorption spectra produced through the use of the aqueous buffer removal function presented. Only a small number of single cell spectra were compared, nonetheless, spectroscopic data results revealed spectral differences similar to that seen from the previous spectroscopic data results (chapter) of the same cell samples, but cell spectra recorded in air. As a consequence of the good results of seen from the proof of concept study, living cells were also successfully investigated a benchtop FT-IRMS instrumentation and a commercially bought micro-static device. Although with the experimental setup poor S/N was achieved when recording data from single living cells, data results from spectra recorded from cell clumps provided good spectral separations of ATRA and DMSO treated TERA2.cl.SP12 cells for 7 days. Flow cytometry results of ATRA and DMSO treated EC cells in this thesis investigation showed that after 7 days of treatment, the majority of ATRA treated cells have differentiated, therefore, the spectral differences between ATRA and DMSO treated average living cell spectra can be considered as a result of cellular differentiation; ATRA inducing neuronal
cellular differentiation. Spectral differences of living cell spectra, found through comparing the PC1 loading plot and mean second derivative spectra, revealed the nucleic acid bands to be the salient spectral differences between ATRA and DMSO living cell spectra. Absorption bands associated with nucleic acid molecular structures were not seen to have large influences in the spectral differences of retinoid treated TERA2.cl.SP12 cell spectra, when compared against the control stem cell spectra. This may have been down to the increase in IR signal from nucleic acids when samples are hydrated. Therefore, not only does recording IR spectra from hydrated cell spectra remove the problem of RMieS, but the increase in nucleic acid signal may play an important role in discriminating cell phenotype, possible enabling spectral biomarkers for cell phenotype to be found and improving the classification results. In conclusion, this study highlights the potential of coupling FT-IRMS with microfluidic devices in detecting subtle biochemical changes of either fixed or live cells without the need for labels in a quick and relatively inexpensive manner. With careful consideration of the IR beam and appropriate processing methods to remove the buffer contributions being applied, there is no reason why FT-IRMS screening of live cells cannot be used in conjunction with immunological expression analyses and other biological assays in providing complementary biochemical data for use in research, or a clinical capacity.
Chapter 7:

Overall Conclusions from the Study and Future Work

7.1 Conclusions

To conclude, the results from the experimental investigation contained in chapter 4 show that IRMS, coupled with appropriate scatter correction and multivariate chemometric classification algorithms, is capable of distinguishing the retinoid induced biochemical changes of the TERA2.cl.SP12 EC stem cells as a result of cellular differentiation; TERA2.cl.SP12 cells differentiating to produce neuronal cell phenotypes upon the addition of ATRA, EC23 and AH61 retinoids to cell sample cultures (AH61 and EC23 are synthetic retinoids synthesised at Durham University). Results also show that biochemical differences can be captured using IRMS as early as 3 days after retinoid treatment, and spectral analysis results showed that even though both ATRA and EC23 produce neuronal cell phenotypes at the 7 day period, biochemical differences can be seen between the cells treated with the different retinoids; in agreement with previous studies using biological assays and mass spectrometry. Spectroscopy results reveal the amide I secondary structure to be a particularly salient absorption band when investigating the spectral changes as a result of differentiation, in agreement with other studies using IRMS to investigate the differentiation process of single cells. Results therefore demonstrate the sensitivity of IRMS technology as well as highlighting the potential for its use to monitor the status of cell phenotype, capturing the discrete molecular changes are a result of differentiation.

Further investigation of the TERA2.cl.SP12 cell line in this study (chapter 5), with the aim being to see how early retinoids initiated a biochemical response and to try and model the spectroscopic signal changes of the EC cell line through to differentiation process, proved to be less successful (chapter 5), with no real spectroscopic differences being seen at the early sample time points and poor spectral separations being seen at the later time point samples. The poor analysis results were probably a result of a change in the experimental methodology, which resulted in a greater level of RMieS being contained in the majority of
cell spectra recorded. The RMieS-EMSC correction found correcting the cell spectra with increased RMieS difficult and poor cell spectra with a high level of variance were produced as a result. Results in chapter 5 are therefore a reminder that although IRMS has a lot of potential for investigating sing cells, potentially enabling the understanding the cell phenotype without the need for harmful and invasive labels and biomarkers, there are still fundamental issues which have a negative impact on the reproducibility of the analytical technique. Only when these are completely understood will the potential of IR spectroscopy for cell phenotype screening be realised. However, although the RMieS-EMSC correction algorithm produced some very strange pure absorption cell spectra in the experiment described in chapter 5, on investigation of the correction function it would seem that through altering correction parameter values of the RMieS-EMSC algorithm the corrected spectra can be improved. To my knowledge this is the first time that the correction algorithm has been investigated in this manner.

Although RMieS was a problem for the experimental investigation described in chapter 5, the investigation involved chemically fixed cells and cell spectra recorded in air using transmission mode spectroscopy. However, there is a closer matching of refractive indexes between biological cells and water than when compared to that of air and biological cells. Thus, investigating single cells immersed in an aqueous liquid with IR vastly reduces RMieS. Therefore, one may have more confidence in the cell spectra recorded using this methodology than cell spectra recorded in air. Another advantage to this methodology being that an aqueous environment is much closer to the natural environment of living eukaryotic cells, and allows living cells to be probed with IR.

The experimental investigation described in chapter 6 shows that it is indeed possible to record IR spectra from cells immersed in an aqueous environment, as long as an appropriate water pathlength is maintained combined with a processing method for modelling/removing the aqueous absorption contribution from recorded spectra. In this study, firstly, formalin fixed cells and then living cells were successfully investigated using IRMS coupled with micro-devices to provide a suitable water barrier. The study also describes a new processing method to model the aqueous buffer contribution contained in raw cell plus buffer spectra, and pure cell absorption spectra produced show no evidence of RMieS, therefore, highlighting the potential of IRMS for the screening of living cells. The study also elucidates the changes to recorded spectra when investigated cells in the
different environments (air and aqueous solution). Results show that not only do you see an increase in nucleic acid signal from hydrated cell spectra, but also spectroscopic changes are seen from the amide I and II bands as a result of hydration and increased hydrogen bonding; even from formalin fixed cells. The spectroscopic investigation of living DMSO and ATRA treated TERA2.cl.SP12 cells (treated for 7 days before IR data recorded) revealed good spectral separation when using PCA, all be it from a small number of spectra and average cell spectra recorded from cell clumps. While the amide I and II absorption bands had to be removed prior to multivariate analysis as a result of over-subtraction, clear spectral differences were seen between the different cell phenotypes; mainly from absorption bands associated with nucleic acid molecules. With absorption bands associated with nucleic acid molecules not being prominent in spectral separations from studies investigating cell spectra recorded in air, this would suggest that only spectra recorded from hydrated living cells provides reliable information on both DNA and RNA molecules present in the cell; nucleic acid vibrations being more difficult to detect and distinguish in dehydrated cells due to lower absorption and band broadening, whereas the hydrated B-like DNA are more easily observed. As nucleic acids are heavily involved in cellular differentiation then it would seem that investigating living cells with IR will be an advantage. Investigating stem cell differentiation through this method may also bring a better understanding of the biochemical changes of living stem cells, possible producing spectral biomarkers of cellular differentiation. It remains to be seen as to whether the same can be said of formalin fixed cells also investigated in this manner.

7.2 Future work

Critically, cell phenotype should be understood from living cells with the biochemistry being maintained after information is recorded so that single cells can retain their biological function. At this present time, current methods used to probe cell phenotype are not able to investigate single cells without compromising their biological viability. Results from this study show that IR spectroscopy can record cell spectra from living cells. In the experimental investigation described in chapter 6, the benchtop IRMS instrumentation was unable to provide acceptable S/N when applying apertures roughly the size dimensions of the single cells immersed in aqueous solution, therefore, apertures were opened up to let more light through to the detector and cell clumps were probed to produce an average living cell spectrum. To solve this problem the high brightness of synchrotron light can be
used. However, is the sequential recording of single cell spectra using synchrotron light the best method?

If living cells are not immersed in their culture media then they are going to be unhappy and will start to undergo apoptosis after some time. Although PBS provides an environment where cells will not burst due to osmotic pressure, there are no nutrients in the buffer as the nutrients will contribute to the spectrum recorded, therefore, eukaryotic cells should not be investigated in this environment for long periods of time. Unfortunately, recording single cell spectra sequentially is time consuming. A better method is to record chemical images of the cells immersed in the aqueous solution, with the image dimensions spanning a wide field of view. This will enable more cell spectra to be recorded in a much quicker time. However, thermal Globar IR sources may still not be bright enough resulting in long image collection times with increased co-added scans. Synchrotron light can provide a brighter light source but access to synchrotron facilities for extensive live cell investigations using IR is not ideal.

The answer to low S/N may be solved by lasers, such as Quantum cascade lasers (QCL’s). QCL’s, employing broadly tunable lasers to produce IR radiation over broad frequency ranges are under rapid development and can produce a high brightness rivalling synchrotron light, but with the advantage of the instrumentation being employed on the bench. The high brightness of these broadly tunable QCL’s can speed up image spectral acquisition times, ideal for investigating living cells with IR light. Therefore, in the future, I hope to investigate the use of QCL IR imaging as method to investigate living single cells. Investigations will also probe the viability of living cells after spectroscopic measurement so as to see whether IR spectroscopy can be a truly non-invasive analytical method to probe living cells.
Chapter 8: References


[70] G. L. Carr, L. M. Miller and P. Dumas, Chapter 7 - Synchrotron Radiation as a Source for Infrared Microspectroscopic Imaging with 2D Multi-Element Detection in *Biomedical application of synchrotron infrared microspectroscopy* (A Practical Approach); (RSC, 2011).


Chapter 9: Supplementary Material for

9.1 Chapter 4

Sup Fig. 1 PCA of spectra recorded from all biological replicates of untreated (pink) and DMSO treated TERA2.cl.SP12 cell samples (black) (a) PCA of 5 day samples (b) PCA of 3 day samples.
9.1.1 Chapter 4 – Day 7 Data Analysis
Sup Fig. 2 (a) PC-LDA scores histogram from the average bootstrap result (2000 bootstraps) of spectra recorded from DMSO and AH61 treated cells left in sample cultures for 7 days; 5 PCs entered into the LDA classification. Average bootstrap achieved 96% CC (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves AH61 (cyan) DMSO (black)) (c) Mean spectra differences between the DMSO and AH61 treated cell spectra (1) 1000-1770 cm\(^{-1}\) spectral range (2) 2800-3100 cm\(^{-1}\) spectral range. PRESS/RSS ratio result showed the 20\(^{th}\) PC to exceed 1 and the first 5 PCs contained 94% of the spectral variance contained in the original data set.
Sup Fig. 3 (a) PC-LDA scores histogram from the average bootstrap result (2000 bootstraps) of spectra recorded from DMSO and EC19 treated cells left in sample cultures for 7 days; 5 PCs entered into the LDA classification. Average bootstrap achieved 94% CC (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC19 (red) DMSO (black)) (c) Mean spectra differences between the DMSO and AH61 treated cell spectra (1) 1000-1770 cm⁻¹ spectral range (2) 2800-3100 cm⁻¹ spectral range. PRESS/RSS ratio result showed the 21st PC to exceed 1 and the first 5 PCs contained 95% of the spectral variance contained in the original data set.
Sup Fig 4 (a) PC-LDA scores histogram from the average bootstrap result (2000 bootstraps) of spectra recorded from DMSO and ATRA treated cells left in sample cultures for 7 days; 5 PCs entered into the LDA classification. Average bootstrap achieved 94% CC (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves ATRA (green) DMSO (black)) (c) Mean spectra differences between the DMSO and AH61 treated cell spectra (1) 1000-1770 cm⁻¹ spectral range (2) 2800-3100 cm⁻¹ spectral range. PRESS/RSS ratio result showed the 25th PC to exceed 1 and the first 5 PCs contained 95% of the spectral variance contained in the original data set.
9.1.2 Chapter 4 – Day 5 Data Analysis

Sup Fig 5 (spectra recorded from cell samples at 5 days) 5 PCs entered into the LDA classification with average bootstrap achieving 86% CC (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC23 (blue) DMSO (black)). PRESS/RSS ratio result showed the 21st PC to exceed 1. Also, the first 5 PCs contained 95% of the spectral variance from the original data set (b) LDA 1 loading plot (upper curves (pink)) from the average bootstrap of the PC-LDA analysis of AH61 and DMSO treated cells and mean spectra comparisons (lower curves ATRA (cyan) DMSO (black)), 91% CC was achieved with 4 PCs. PRESS/RSS ratio result showed the 22nd PC to exceed 1. Also, the first 5 PCs contained 94% of the spectral variance from the original data set.
9.1.3 Chapter 4 – Day 3 Data Analysis

(a)

(b)
Sup Fig. 6  (spectra recorded from cell samples at 3 days) 5 PCs entered into the LDA classification with the average bootstrap achieving 88% CC (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC23 (blue) DMSO (black)). PRESS/RSS ratio result showed the 21st PC to exceed 1. Also, the first 5 PCs contained 95% of the spectral variance from the original data set (b) PC-LDA comparisons of ATRA (green) and DMSO (black) samples with 5 PCs inputted. Average bootstrap achieved 75% CC, PRESS/RSS ratio result showed the 18th PC to exceed 1 (c) PC-LDA comparisons of AH61 (cyan) and DMSO (black) samples with 5 PCs inputted. Average bootstrap achieved 93% CC, PRESS/RSS ratio result showed the 18th PC to exceed 1 (d) PC-LDA comparisons of EC19 (red) and DMSO (black) samples with 6 PCs inputted. Average bootstrap achieved 96% CC, PRESS/RSS ratio result showed the 19th PC to exceed 1.
9.1.4 Chapter 4 - Retinoid Treated Cell Spectra Comparisons

Sup Fig. 7 (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves ATRA (green), AH61 (cyan) and EC23 (blue)) (day 7 sample spectra) (b) blow up of amide I spectral range (c) LDA 2 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves ATRA (green), AH61 (cyan), EC23 (blue) and EC19 (red)).
9.1.5 Chapter 4 - Monitoring the Spectroscopic Signals of Differentiating Cells through Time using IRMS
Sup Fig. 8 (a1) PC-LDA scores plot (LDA 1 vs. LDA 2) of spectra recorded from EC23 treated cells samples at the 3 day (black), 5 day (blue), 7 day (green) and 14 day (red) time points, 8 PCs entered into the LDA classification. Average bootstrap achieved 80% CC. Mean spectral comparisons, 1000-1758 cm\(^{-1}\) spectral range (a2) and lipid spectral range (a3) (b1) PC-LDA scores plot (LDA 1 vs. LDA 2) of spectra recorded from ATRA treated cells samples at the 3 day (black), 5 day (blue), 7 day (green) and 14 day (red) time points, 5 PCs entered into the LDA classification with an average bootstrap achieving 74% CC. Mean spectral comparisons, 1000-1758 cm\(^{-1}\) spectral range (b2) and lipid spectral range (b3)
Sup Fig. 9 (a) PC-LDA scores plot (LDA 1 vs. LDA 2) of spectra recorded from DMSO treated cells samples at the 3, 5 and 7 time points, 6 PCs entered into the LDA classification. Average bootstrap achieved 63 % CC (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves DMSO day 3 (black) DMSO day 5 (blue) DMSO day 7 (green)) (c) LDA 2 loading plot. PRESS/RSS ratio result showed the 25th PC to exceed 1. Also, the first 5 PCs contained 95.8% of the spectral variance from the original data set.
9.1.6 Chapter 4 - Day 14 Retinoid Treated EC Cells
Sup Fig. 10 PCA comparisons of spectra recorded from the retinoid treated cells at 14 days vs. control stem cells at the 7 day period (a) PC-LDA scores plot of ATRA, EC23, AH61 and DMSO treated cells with 10 PCs entered into the LDA classification algorithm and 90% confidence ellipses shown (b) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves EC23 (blue), ATRA (green), AH61 (cyan) and DMSO (black)). PRESS/RSS ratio result showed the 25th PC to exceed 1. Also, the first 10 PCs contained 96% of the spectral variance from the original data set (c) Mean spectral comparisons ((c1) 1000-1800 cm⁻¹ spectral range) (c2) (lipids 2830 – 3100 cm⁻¹).

Table. S1 Spectral Differences between ATRA, AH61 and EC23 Spectra Recorded from Retinoid 14 Day samples and 7 Day Control Samples

<table>
<thead>
<tr>
<th>Day 14 cells (&gt;) increase in absorption or new band (cm⁻¹)</th>
<th>Day 14 cells (&lt;) decrease in absorption or new band from control cells at 7 days (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024, 1041, 1155, 1236, 1313, 1402, 1458, 1473, 1640, 1670, 1730, 1750, 2848, 2910, 2929, 2960, 3010</td>
<td>1032, 1450, 1552 (signifying a change in band position), 1628, 2920</td>
</tr>
</tbody>
</table>

(c2)
Sup Fig. 11 PC-LDA spectral loadings and mean spectra of all 5 day data. 9 PCs inputted into the LDA algorithm (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves DMSO (black) ATRA (green), AH61 (cyan) and EC23 cell spectra (blue)) (b) LDA 2 loading plot (upper curves (pink)) and mean spectral comparisons.
Sup Fig. 12 PC-LDA spectral loadings and mean spectra of all 7 day data, 10 PCs inputted into the LDA algorithm (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves DMSO (black) ATRA (green), AH61 (cyan) and EC23 cell spectra (blue)) (b) LDA 2 loading plot (upper curves (pink)) and mean spectral comparisons.
Sup Fig. 13 PC-LDA spectral loadings and mean spectra of all 5 day data, 6 PCs inputted into the LDA algorithm (a) LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons (lower curves DMSO (black) ATRA (green), AH61 (cyan) and EC23 cell spectra (blue)) (c) LDA 2 loading plot (upper curves (pink)) and mean spectral comparisons.
Sup Fig. 14 (a) PC-LDA spectral loadings and mean spectra of all 3 day data, 6 PCs inputted into the LDA algorithm. LDA 1 loading plot (upper curves (pink)) and mean spectra comparisons displayed in the figure (lower curves DMSO (black) ATRA (green), AH61 (cyan) and EC23 cell spectra (blue).