Biochemical phenotyping of live single cells using confocal Raman spectroscopy

Julia Gala de Pablo

Department of Physics and Astronomy

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

September 12, 2019

This page is intentionally left blank.

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. Further details of the jointly-authored publications and the contributions of the candidate and the other authors to the work should be included below this statement.

The work in Chapters 4 (page 91) and 5 (page 111) of the thesis has appeared in publications [1] and [2] respectively. Work mentioned in Chapter 7 (page 192) was published in [3] and is currently in preparation to be published in a join author publication. The contributions of the author were as follows:

1. Julia Gala de Pablo, Fern J. Armistead, Sally A. Peyman, David Bonthron, Michael Lones, Stephen Smith, and Stephen D. Evans. Biochemical Fingerprint of Colorectal Cancer Cell Lines Using Label-Free Live Single-Cell Raman Spectroscopy. *Journal of Raman Spectroscopy* 49(8):1323-32, aug 2018. ISSN 03770486. doi:10.1002/jrs.5389. [1]

    - I was responsible for the cell preparation, all data acquisition and Raman spectra analysis, and the multivariate methods analysis for PCA, PCA/LDA and simple DTs. Additionally, I was responsible for the generation of the manuscript. Results for this publication appear in the Results chapter 4 on page 91.
    - Fern J. Armistead helped with cell culture maintenance. Michael Lones contributed with the C5.0 and the SVM analysis, and adviced in the choice of cross-validation and sampling. Sally A. Peyman, Stephen D. Evans, Stephen Smith and David Bonthron contributed with manuscript review and corrections and with data discussion.

2. Julia Gala de Pablo, David R. Chisholm, Andreas Steffen, Amanda K. Nelson, Christoph Mahler, Todd B. Marder, Sally A. Peyman, et al. Tandem Fluorescence and Raman (fluoRaman) Characterisation of a Novel Photosensitiser in Colorectal Cancer Cell Line SW480. *The Analyst* 143(24):6113-20, 2018. ISSN 0003-2654. doi:10.1039/C8AN01461B. [2]

- I was responsible for the cell culture of the SW480 cells, the incubation with the compound, the preparation, acquisition and analysis of all the Raman and confocal experimental data, and of the manuscript production. Results for this publication appear in the Results chapter 5 on page 111.
- The Chemistry Department of Durham University (David R. Chisholm, John M. Girkin, Carrie A. Ambler and Andrew Whiting) synthesized and purified the DC473 compound. The members of the Institute of Inorganic Chemistry of the Julius-Maximilians-Universität Würzburg (Andreas Steffen, Amanda K. Nelson, Christoph Mahler and Todd B. Marder) contributed with the fluorescence characterization of the DC473 compound. Sally A. Peyman and Stephen D. Evans contributed with data discussion. All authors contributed with manuscript review and corrections.

3. Armistead, F. J., Gala De Pablo, J., Gadêlha, H., Peyman, S. A., & Evans, S. D. (2019). Cells Under Stress: An Inertial-Shear Microfluidic Determination of Cell Behavior. *Biophysical Journal*, 116(6), 1127-1135. doi:10.1016/j.bpj.2019.01.034. [3]

    - I contributed with aid on cell culture and Matlab script production for aiding in analysis on the output of the image analysis of the deformation events. Results for this publication and a future following publication are discussed in the Conclusions chapter 7 (page 192).

In some parts of the thesis, figures have been reproduced from publications with permission from the copyright holder:

1. Fig. 1.2A (p6): permission from Elsevier, License: 4515260915862, 24/01/2019.
2. Fig. 1.2B (p6): permission from the ACS[1], 24/01/2019.
3. Fig. 1.6 (p19): permission from the RSC[2], License: 4515280035078, 24/01/2019.
4. Fig. 2.11 (p47) contains two pictures reproduced from Van Dyke's book from 1982 *An Album of Fluid Motion* [4]. Parabolic Press recommended contacting the author, who passed away in 2010, and his wife passed away in 2017 and could not be contacted for permission. His daughter, Nina Van Dyke, gave permission for using these images as an honour to his father.

[1] American Chemical Society
[2] Royal Society of Chemistry

Esta tesis se la dedico a mi madre, Milena, la mujer mas fuerte que he conocido y la primera que me habló de ciencia, a mi hermana, Eva, mi otra mitad, y a mi padre, Santiago, que me puso un teclado delante y me descubrió la programación.

This thesis is dedicated to my mother, Milena, the strongest woman I have ever known and the first one that talked to me about science, to my sister, Eva, my other half, and to my father, Santiago, that gave me a keyboard and introduced me into code writing.

This page is intentionally left blank.

Look at the resplendent colours on the soap bubbles!
Why is the sea blue?
What makes diamond glitter!
Ask the right questions, and nature will open the doors to her secrets

Sir Chandrashekhara Venkata Raman (1888-1970)

This page is intentionally left blank.

Acknowledgements

My thanks to Steve, for all his mentoring, support and advice. For the more than thirty drafts of my first paper. For actually reading this thesis. For turning my *spaniglish* into english. To Sal, because her door is always open, specially during my second year crises. For the trecking and for introducing me to whiskey.

My thanks to Fern, with whom I have been *sciencing* since the summer of 2013 and that has shared the ups and downs of this big project. To Twig, who has always been around for a coffee even when I despaired. To Victoria, the first one to invite me to an MNP night out. To Mr. Churchman, just for being himself. To Ellen and Abiral, for sharing the *Brexit corner* with me. To Holly for the lunges. To Liam, for his burnt SU8 wafer and other Rapid Prototyping banter. To Jamie, for his *Buenas mañanas*. To Jinyang, for his ethically dubious worm videos. To Matt H., for his games nights. To Mark, for his Taiwan golden T-shirt. To Grace, for being my Huel mate. To Ash, even though he killed the spinach plant. To George for the brownies. To my sister, for turning my awful *physics maths* into real maths. To the postdoc office team: Peng, Radwa, Sunjie and Anders. Because they did not kill me even in my most writing procrastinating moments. To inspirobot, for reminding me that everything is meaningless and we are all going to die.

In general, my thanks to all MNP, because I came to the UK alone and they made of these PhD years some of the best of my life. And to SMP, because they are also very cool. To the Fenton for existing, although they should thank me for all the money I spent there.

Last but not least, my thanks to a certain writer, with whom I shared some good and bad times. And to all the climbing people, particularly a ginger one, because they kept me sane for the last two years.

Finally, to whoever is reading this, for not taking this acknowledgements too seriously, as in the end this is just another book that will gather some dust in the deepest levels of Leeds library.

Hope you enjoy the science.

This page is intentionally left blank.

Abstract

Colorectal cancer is the 3rd most common cancer worldwide and has the second highest mortality rates, and around 20% of the tumours of diagnosed patients have metastasis. Understanding cancer development and progression represents a major challenge. As a proof of concept, hundreds of live single-cells derived from different stages of colorectal adenocarcinoma – from primary to metastasis (SW480, HT29, SW620) – were compared to a cell line from colorectal carcinoma (HCT116) and a leukaemia cell line (HL60) using Raman confocal microspectroscopy, obtaining >92% correct rate classification using principal component analysis and linear discriminant analysis, showing that Raman spectroscopy could detect differences between different colorectal cancer stages.

Photodynamic therapy is a treatment that uses photosensitisers that selectively accumulate in cancerous cells. Light activates the photosensitiser, killing the cancer cells while sparing the surrounding tissue, with promising results on colorectal cancer. This thesis aimed to prove if Raman spectroscopy could detect the presence of these agents and their effects in live single cells. Experiments were done to detect Protoporphyrin IX accumulation in live single colorectal adenocarcinoma cells. Moreover, the novel photosensitizer DC473 (provided by LightOX) accumulation was detected within fixed SW480 cells using Raman and fluorescence. Furthermore, the accumulation of DC473 was analysed in live single-cells, with higher accumulation at later stages of adenocarcinoma (HT29, SW620). DC473 accumulation caused a significant drop in the lipids bands of the adenocarcinoma cells. Additionally, the activation of the photosensitiser was tracked on live single-cells using Raman, observing the partial chemical reaction of the photosensitiser and a drop of the cells' viability over time. The bands of the fraction of DC473 that did not react were observed to broaden and shift to higher wavenumbers, indicating a change in location of the photosensitiser after activation.

To test if Raman spectroscopy could be pushed to higher throughput using microfluidics, Raman acquisition was made in Polydimethylsiloxane microfluidic devices on the same cell before and following incubation with DC473. Results showed that Raman spectroscopy has great potential for single-cell time-dependent drug treatment studies.

This page is intentionally left blank.

Abbreviations

The following mathematical notation was used in the thesis.

Mathematical notation

h Planck's constant

k Boltzmann's constant

$\hbar$ $h/2\pi$

μ Molecular dipole moment

α Polarizability tensor

E Electric Field

q_v Vibrational coordinate

ν_0 Frequency of the incident light

ν_v Frequency of the vibration

μ_m Reduced mass

m Mass

The following abbreviations are used throughout the thesis, in addition to the three letter abbreviation for amino acids. They are indicated as footnotes the first time they are mentioned in each section.

Acronyms and Abbreviations

5-ALA 5-aminolevulinic acid

ABC ATP binding cassette

ABCG2 ATP-binding cassette super-family G member 2

ACS American Chemical Society

ALAS Aminolevulinic acid synthase

ANN Artificial neural networks

APC Adenomatous polyposis coli

ATRA All trans-retinoic acid

ATR-IR Attenuated total reflection infrared

BP Band pass (filter)

BRAF v-Raf murine sarcoma viral oncogene homolog B

BSA Bovine serum albumin

CA Cluster analysis

CARS Coherent anti-Stokes Raman spectroscopy

CCD Charge-coupled device

CI Chromosomal instability

CIMP CpG island methylator phenotype

COX2 Cyclooxygenase-2

CRC Colorectal Cancer

CTC Circulating Tumour Cells

cyt C Cytochrome C

DEP Dielectrophoresis

DMEM Dulbecco's Modified Eagle Medium

DMSO Dimethyl sulfoxide

DNA Deoxyribonucleic acid

DPBS Dulbecco's Phosphate-Buffered Saline

DT Discrimination Trees

EGFR Endothelial growth factor receptor

EMCCD Electron multiplying charged-coupled device

EMSC Extended Multiplicative Scatter Correction

FA Fatty acid

FBS Foetal Bovine Serum

FDA Fluorescein diacetate

F.D.A. Food and Drug Administration

FGF Fibroblast growth factor

FTIR Fourier-transform infrared

FWHM Full-width half maximum

GCPQ Quaternary ammonium palmitoyl glycol chitosan

GFP Green Fluorescent Protein

G-CSF Granulocyte colony stimulating factor

GLS Gaussian-Lorentzian sum

GLP Gaussian-Lorentzian product

GM-CSF or CSF2 Granulocyte-Macrophage colony stimulating factor

HTS High-throughput screening

HCS High-content screening

HMBS hydroxymethylbilane synthase

HSES High Speed Encoded Stage

HUVECs Human umbilical vein endothelial cells

IPA Isopropanol

IR Infrared

ITO Indium tin oxide

KRAS Kirsten RAt Sarcoma virus oncogene

LB Luria-Bertani/Lysogeny Broth (medium)

LD Linear Discriminant

LDA Linear Discriminant Analysis

LDH Lactate dehydrogenase

LED Light Emitting Diode

LOND Lipid oil nanodroplets

LP Long pass (filter)

LTRS Laser tweezers Raman spectroscopy

LUT Look-up table

MSI Microsatellite instability

NA Nucleic acids

ND Neutral density (filter)

NP Nanoparticle

OLI-MB Oil layer inside microbubbles

PC Principal coefficient

PCA Principal Component Analysis

PDMS Polydimethylsiloxane

PDT Photodynamic Therapy

PEI Polyethylenimine

PET Polyethylene Terephtalate

PFA Paraformaldehyde

PI Propidium Iodide

PI3K Phosphatidylinositol-3-kinase

PLGA Poly(lactide-co-glycolide)

PLSR Partial Least Squares Regression

PMMA Poly(methyl methacrylate)

PPARγ Peroxisome proliferator-activated receptor γ

PPIX Protoporphyrin IX

PS Photosensitiser

PUFA Polyunsaturated fatty acid

PVA Polyvinyl alcohol

RACE Raman activated cell ejection

RACS Raman activated cell sorter

RACC Raman activated cell cytometer/cytometry

RAR Retinoic acid receptor

RBC Red blood cell

Re Reynold's number

ROS Reactive oxygen species

RR Resonant Raman

RS Raman spectroscopy

RSC Royal Society of Chemistry

RXR Retinoic X receptor

SAV Surface area to volume ratio

SERRS Surface-enhanced resonance Raman scattering

SERS Surface enhanced Raman spectroscopy

SRS Stimulated Raman scattering

STR Short Tandem Repeat (profiling)

SVM Support Vector Machines

TE Thermoelectric

UV Ultraviolet

VBNC viable but non-culturable (bacteria)

VEGF Vascular endothelial growth factor

WBC White blood cell

WT Wild type

Contents

Page

Statements . . . i
Dedication . . . iii
Acknowledgements . . . vii
Abstract . . . ix
Abbreviations . . . xi

I Introduction and Theory **1**

1 Introduction **2**
1.1 Background to Single Cell Analysis: Combining Raman and Microfluidics . . 2
1.2 Raman Spectroscopy . . . 3
1.2.1 Raman Spectra Analysis . . . 4
1.3 Raman Spectroscopy of Biological Systems . . . 4
1.3.1 Raman Spectroscopy of Single Cells . . . 4
1.3.2 Raman Spectroscopy for Medical Diagnosis . . . 6
1.3.3 Live versus Fixed Cells . . . 7
1.3.4 Colorectal Cancer as a Model System . . . 9
1.3.5 Model System . . . 11
1.4 Raman Spectroscopic Single-Cell Detection of Drugs . . . 12
1.5 Photodynamic Therapy . . . 14
1.5.1 Porphyrins . . . 15
1.5.2 Novel Photosensitisers . . . 17
1.6 Raman Spectroscopy and Photodynamic Therapy . . . 17
1.7 Raman *on a chip* . . . 18
1.8 Thesis Overview . . . 22

2 Background Theory **24**
2.1 Theory of Raman spectroscopy . . . 24
2.1.1 Classical Theory . . . 25
2.1.2 Quantum Theory . . . 28
2.1.3 Intensity of Raman scattering and Raman cross-section . . . 29
2.1.4 Bandwidths . . . 30
2.1.5 Resonant Raman . . . 32
2.2 Theory of Chemometric Analysis . . . 32
2.2.1 Unsupervised Multivariate Methods . . . 33
2.2.1.1 Theory of Principal Component Analysis . . . 33
2.2.1.2 Partial Least Squares Regression . . . 37
2.2.2 Supervised Multivariate Methods . . . 38
2.2.2.1 Training and Validation . . . 38
2.2.2.2 Linear Discriminant Analysis . . . 38
2.2.2.3 Discrimination Trees . . . 41
2.2.2.4 Support Vector Machines . . . 42
2.2.2.5 Evaluation of a Classifier and Confusion Matrices . . . 44

2.3 Theory of Photodynamic therapy . . . 45
2.4 Microfluidic Theory . . . 46

II Methods 48

3 Methods 49
3.1 Microscopy and Spectroscopy Facilities . . . 49
3.2 Raman System Characterisation and Experimental Design . . . 52
3.2.1 Raman System Calibration to the Silicon Peak . . . 52
3.2.2 Raman Signal of Potential Substrates . . . 53
3.2.3 System Confocality . . . 55
3.2.4 Comparison of Excitation Lasers Available . . . 56
3.2.5 Comparison of Detectors and Scanning Options . . . 59
3.2.6 Beam Expander . . . 61
3.2.7 Test of the Phototoxicity of the 532 nm Excitation Laser . . . 63
3.2.8 Focus Drift . . . 63
3.3 Raman optimised Parameters . . . 64
3.3.1 Raman Single-Cell Spectroscopy . . . 64
3.3.2 Raman Mapping . . . 66
3.3.3 Raman Spectra from Standards . . . 66
3.3.4 Raman Spectra Pre-Processing . . . 67
3.3.5 General Pre-Processing Final Settings . . . 70
3.4 Raman Analysis . . . 71
3.5 Gaussian-Lorentzian Sum Fitting of Individual Single-Cell Spectra . . . 72
3.6 Cell Culturing . . . 74
3.6.1 Adherent cell lines . . . 74
3.6.2 Non-adherent cell lines . . . 75
3.6.3 Model System Cells Staining and Imaging . . . 75
3.6.4 Incubation with Photosensitisers . . . 76
3.7 Rapid Prototyping Facilities . . . 76
3.8 Microfluidic Designs and Fabrication . . . 77
3.8.1 Microfluidic Devices Fabrication . . . 77
3.8.2 Microfluidic Experimental Procedures . . . 81
3.8.3 PDMS Metal Coating Experiments . . . 81
3.8.4 Trap Dimensions Optimisation . . . 81
3.8.5 Takeuchi-like Trap Dimensions Optimisation . . . 83
3.8.6 Cell Traps Design and COMSOL Modelling . . . 85

III Results 90

4 Raman Spectroscopy of Live Single cells 91
4.1 Adenocarcinoma Model System . . . 91
4.1.1 Cell lines Imaging . . . 91
4.2 Raman on Live Single-Cells of Different Cell Lines . . . 93
4.2.1 Spectral Averaging . . . 94
4.2.2 Correlation Analysis . . . 98
4.3 Multivariate Analysis . . . 100
4.3.1 Principal Component Analysis . . . 100
4.3.2 Principal Component Analysis and Linear Discriminant Analysis . . . 101
4.3.3 Discrimination Trees . . . 103
4.3.4 Support Vector Machines . . . 104
4.3.5 Model Comparison and Final Performance Values . . . 104

4.3.5.1 Average Analysis of All Cell Lines . . . 105
4.3.5.2 Multivariate Analysis of All Cell Lines . . . 107
4.3.6 Partial Least Squares Regression . . . 108
4.4 Final Conclusions . . . 109

5 Raman for detection of photodynamic agents in single colorectal cancer cells 111
5.1 Classic Photodynamic Agent - Protoporphyrin IX . . . 111
5.1.1 Standards Raman Spectra . . . 113
5.1.2 Raman on PPIX Expressing Single Adenocarcinoma Cells . . . 118
5.1.3 Correlation Analysis for Separation of Cytochrome C and Protoporphyrin IX Contributions . . . 120
5.1.4 PCA Analysis of the 5-Aminolevulinic Acid Data . . . 120
5.1.5 Protoporphyrin IX Detection Using Fluorescence . . . 123
5.1.6 Photodynamic Activation of Protoporphyrin IX . . . 125
5.1.7 Conclusions and Challenges . . . 125
5.2 Novel Photodynamic Agent DC473 Accumulation . . . 127
5.2.1 DC473, EC23 and DC271 Raman Spectra . . . 129
5.2.2 Confocal Fluorescence and Single Point Raman Detection of DC473 within Fixed Single-Cells . . . 134
5.2.3 Raman Mapping and Confocal Fluorescence Detection of DC473 within Fixed Single-Cells . . . 134
5.3 Detection of DC473 in Live Single-Cells . . . 138
5.3.1 Principal Component Analysis of the DC473 Accumulation Data . . . 140
5.3.2 Fitting of the Main DC473 Bands . . . 142
5.4 Tracking Apoptosis after UV Light Activation of DC473 . . . 145
5.4.1 Activation of DC473 in solution . . . 145
5.4.2 Tracking Apoptosis after DC473 Activation . . . 146
5.4.3 Analysis of the Average and Difference Spectra . . . 148
5.4.4 Principal Component Analysis of the PDT Data . . . 150
5.4.5 PCA/LDA Analysis of the PDT Data . . . 153
5.4.6 Band Fitting of the PDT Data . . . 154
5.4.6.1 Viability Assays of Cells incubated with DC473 . . . 158
5.4.6.2 Conclusions and Future Experiments . . . 160

6 Raman and Microfluidics 161
6.1 What are the Chemical Components of PDMS? . . . 162
6.2 Characterization of the Substrate Background on a PDMS Device . . . 164
6.2.1 Metal Evaporation on PDMS Devices for Background Reduction . . . 164
6.2.2 Gold Nanocomposites for Background Reduction . . . 167
6.2.3 Glass SU8 Devices for Minimization of Background . . . 170
6.2.4 Confocality Reduction of PDMS Background with Channel Height . . . 171
6.2.5 Cell Raman Signal in a Microfluidic Channel . . . 172
6.3 Least Square Fitting for *on-chip* Spectra Background Subtraction . . . 177
6.4 Raman *on-chip* Results . . . 178
6.4.1 *On-chip* Raman for Tracking Accumulation of DC473 on Live Single SW480 Cells . . . 178
6.4.2 Other *on chip* Raman Projects . . . 186
6.4.2.1 Raman on Single *Escherichia coli on-chip* . . . 186
6.4.2.2 Conclusions and Future Challenges . . . 189

IV Conclusions and Future Work 191

7 Conclusions and Future Work 192

7.1 Chapter by Chapter Overview 192
7.2 A Wider Context: The MRC Project 195
7.3 Next Steps 196

References I

List of Figures

Figure 1.1: Pie charts of the approximate chemical composition by weight of a cell 5
Figure 1.2: Example of a cell Raman spectra reproduced from El-Mashtoly *et al.* 6
Figure 1.3: Schematics of the Duke's stages cell line model system of SW480/HT29/SW620 12
Figure 1.4: Chemical structures for 5-ALA, PPIX, heme and tertiary and secondary structure of cyt C 15
Figure 1.5: Chemical structure of the novel photosensitiser DC473 17
Figure 1.6: Backgrounds for PDMS, perspex, silicon, glass, quartz and water reproduced from Chrimes *et al.*. 19

Figure 2.1: Schematics of the polarizability changes with vibrations for the CO_2 molecule 27
Figure 2.2: Jablonski energy diagram comparing IR-absorption, Rayleight scattering, Stokes and anti-Stokes Raman scattering. 29
Figure 2.3: 2D example scatter plots to illustrate PCA 35
Figure 2.4: Schematics illustrating the data analysis steps for performing PCA onto spectroscopy data 36
Figure 2.5: Example scatter plots comparing PLSR and PCA 37
Figure 2.6: Schematics illustrating k-fold cross-validation 39
Figure 2.7: Example scatter plots illustrating LDA classification for 4 classes . . 40
Figure 2.8: Scatter plot and posterior probability plots illustrating LDA for 4 classes 41
Figure 2.9: Scatter plot, posterior probability plots and obtained DT classifier example for 4 classes 43
Figure 2.10: Jablonski energy diagram and schematics explaining type I, II and III photodynamic reactions 45
Figure 2.11: Schematics and images reproduced from Van Dyke 1987 illustrating laminar and turbulent flow 47

Figure 3.1: Raman microscope schematics 50
Figure 3.2: Raman microscope pictures 50
Figure 3.3: Example of the calibration peak obtained from the silicon sample . . 53
Figure 3.4: Raman signal of potential substrates 54
Figure 3.5: Raman signal of buffer 54
Figure 3.6: Axial confocality data obtained from depth scans onto a silicon sample using the 532 nm laser for different objectives, fitted with GLS curves 55
Figure 3.7: Bar plot with a summary of the intensity and the FWHM of the silicon peak when doing depth scans onto a silicon sample for different objectives. 56
Figure 3.8: Raw Raman spectra acquired with the 532 nm and the 785 nm excitation lasers on the same live single cell 57
Figure 3.9: Baseline subtracted Raman spectra acquired with the 532 nm and the 785 nm excitation lasers on the same live single cell 57
Figure 3.10: Artefacts caused by white light contribution with the shutter open or closed when using the 785 nm excitation laser 58

Figure 3.11: Comparison of detectors, gain and scan type when acquiring dark noise 60
Figure 3.12: Dark noise comparison barplot 60
Figure 3.13: Comparison of detectors, gain and scan type when acquiring the spectrum of the same live single cell 61
Figure 3.14: Barplots for each detector's cell signal intensity and noise intensity for different scan types and gain values 61
Figure 3.15: 532 nm laser spot for different laser intensities 62
Figure 3.16: Image of the 532 nm beam expanded laser spot (20 %) when measuring on an HT29 live cell 62
Figure 3.17: Test of the phototoxicity of the 532 nm excitation laser on an HL60 cell, showing spectra, PCA results and bright field images of the cell. 63
Figure 3.18: Focus drift characterisation when measuring on a silicon sample by tracking the intensity of the silicon peak at 520.5 cm^{-1} 64
Figure 3.19: Estimation of the expected focus drift over time 65
Figure 3.20: Example of a raw HL60 single cell spectrum excited with the 532 nm laser, to illustrate the posterior pre-processing steps 67
Figure 3.21: Example of interpolation and alignment steps performed on the spectrum of an HL60 cell. 68
Figure 3.22: Example of the quartz background subtraction step on a HL60 single cell spectrum 69
Figure 3.23: Example of a Savitzki-Golay smoothing step on a HL60 single cell spectrum 69
Figure 3.24: Example of baseline subtraction using the algorith from Koch *et al.* on a HL60 single cell spectrum 70
Figure 3.25: Example of removal of the silent region on a HL60 single cell spectrum 70
Figure 3.26: Example of a HL60 single cell spectrum normalised to the Amide I . 71
Figure 3.27: Pictures of the Rapid Prototyping facilities 77
Figure 3.28: Detailed protocol for SU8 and PDMS fabrication 78
Figure 3.29: Negative and positive photoresist schematics and examples of the main issues found in SU8 fabrication: delamination, ripples and underdeveloped structures 80
Figure 3.30: Schematics for the nomenclature for the Di Carlo-like traps design optimisation 82
Figure 3.31: Schematics for the nomenclature used for the Takeuchi-like traps design optimisation 84
Figure 3.32: Schematics for Takeuchi-like traps trapping and untrapping events . 84
Figure 3.33: Dimensions of an example of working Takeuchi-like traps design . . . 84
Figure 3.34: Bright field image of Takeuchi-like traps with SW480 trapped cells . 85
Figure 3.35: Example of an array of COMSOL results for Di Carlo-like traps, showing the velocity plots when varying channel width and length in traps designs 86
Figure 3.36: Velocity plots of the COMSOL simulations for different Di Carlo-like trap shapes 86
Figure 3.37: Example of HCT116 and SW620 cells trapped in Di Carlo-like trap structures 87
Figure 3.38: HUVECs trapped *on a chip* on Di Carlo-like optimised traps 87
Figure 3.39: Spheroids trapped *on a chip* on Di Carlo-like cell traps 88
Figure 3.40: Di Carlo-like traps efficiency vs. density comparison 88

Figure 4.1: Phase contrast image of the cell lines used: HL60, SW480, HCT116, HT29 and SW620. 92
Figure 4.2: Nile Red and Hoechst staining of the model system cells: SW480, HT29 and SW620. 92

Figure 4.3: Raman pre-processed data for all cell lines used: HL60, HCT116, SW480, HT29 and SW620. . . . 93
Figure 4.4: Peak labelling of the average live single-cell Raman spectrum 94
Figure 4.5: Comparison of the average spectra for SW480 and SW620 live single cells measurements 95
Figure 4.6: CH stretching, Amide III and Amide I fitting of average live single cell spectra for each cell line 97
Figure 4.7: Bar plots of the fittings for the Amide I and Amide III bands and CH stretching bands 98
Figure 4.8: Correlation matrix for all the live single cell spectra acquired 99
Figure 4.9: Correlation lines from the correlation matrix for selected values: 748, 810 and 1679 cm^{-1} peaks 100
Figure 4.10: PCA/LDA results on SW480 and SW620 cells, including loadings, 2D scatter plots of PC1 and 2 and histogram of LDA scores 101
Figure 4.11: Histogram of the scores from PCA for SW480 and SW620 cells . . . 102
Figure 4.12: Variance explained for the first 25 PCs and Performance of LDA with training set size 102
Figure 4.13: DT results on SW480 and SW620 cells, including example tree, performance with training set size, chosen peaks plot and 3D plot of three best single point biomarkers 104
Figure 4.14: Classifiers comparison (PCA/LDA, DT, DT C5.0 and SVM), LDA example scores 3D scatter plot for all cell lines. . . . 105
Figure 4.15: Raman average signal for each cell line used (HL60, HCT116, SW480, HT29 and SW620) 106
Figure 4.16: Boxplot and Beesawmplot for the scores of SW480/HT29/SW620 in the two selected coefficients from PLSR analysis 108
Figure 4.17: PLSR coefficients for SW480, HT29 and SW620 single cell data . . 109

Figure 5.1: Schematics of the PPIX synthesis from 5-ALA in a mammalian cell, adapted from Chiabrando et al. . . . 112
Figure 5.2: Raman 532 nm excitation spectrum of 5-ALA crystal 114
Figure 5.3: Raman 532 nm excitation spectrum of PPIX in solution 114
Figure 5.4: Raman 532 nm excitation spectrum of cyt C in solution 115
Figure 5.5: Comparison of the cyt C Raman 532 nm excitation standard spectrum and the cyt C assigned live single cell correlation series 117
Figure 5.6: Overlap of cyt C and PPIX Raman spectra to illustrate the signal expected from a live single cell 118
Figure 5.7: Average and Difference 532 nm excitation Raman spectra of the 5-ALA treated cells and controls 119
Figure 5.8: Live single cell Raman correlation series assigned as PPIX and cyt C 121
Figure 5.9: PCA variance explained changes with smoothing frame size for the 5-ALA Raman dataset 121
Figure 5.10: PCA loadings for 5-ALA and control samples in smoothed data . . . 122
Figure 5.11: Scores for the most relevant PCs from the 5-ALA Raman single cell dataset 122
Figure 5.12: Fluorescence confocal images of PPIX in the model system cells: SW480, HT29 and SW620. . . . 124
Figure 5.13: Image of SW480 cells death after PDT with PPIX imaged using the Trypan Blue membrane integrity assay 124
Figure 5.14: Picture of the PPIX-death patterning with UV light on HT29 cells imaged using the Trypan Blue membrane integrity assay 125

Figure 5.15: Pictures showing the effects of PPIX activation with UV light on SW480 cells, showing a drop in PPIX fluorescence and bright field images showing morphology changes . . . 126
Figure 5.16: Table of Contents figure for publication in The Analyst journal . . . 127
Figure 5.17: 785 nm excitation Raman spectrum of DMSO . . . 130
Figure 5.18: 785 nm excitation normalised Raman spectra of akyne novel compounds provided by LightOx: EC23, DC371 and DC473 . . . 132
Figure 5.19: Raman spectra of DC473 (1 mM) in DMSO solution when excited with the 785 nm laser . . . 133
Figure 5.20: Absorption and emission spectra of DC473 in different solvents . . . 133
Figure 5.21: Single points Raman spectra excited with the 532 nm and the 785 nm lasers on fixed SW480 cells following incubation with DC473 . . . 135
Figure 5.22: Intensity Raman mapping on a fixed SW480 cell incubated with DC473 and a control cell incubated with DMSO . . . 136
Figure 5.23: PCA on Raman mapping of a fixed SW480 cell excited with the 785 nm laser following incubation with DC473 . . . 137
Figure 5.24: Average spectra of adenocarcinoma live cells (SW480, HT29 and SW620) following incubation with DC473 . . . 139
Figure 5.25: Difference spectra of the average of adenocarcinoma live cells (SW480, HT29 and SW620) following incubation with DC473 or with DMSO . . . 139
Figure 5.26: DC473 1597 cm^{-1} correlation series the 532 nm excitation Raman spectra of all measured cells . . . 140
Figure 5.27: Loadings for PCA analysis of Raman spectra of adenocarcinoma cells following incubation with DC473. Boxplots for the PCA scores for the 532 nm excitation Raman spectra of adenocarcinoma cells. . . . 141
Figure 5.28: Examples of the GSL fittings on DC473 Raman peaks at 1597 and 2195 cm^{-1} on the single cell data for the adenocarcinoma cell lines when exciting with the 532 nm laser line . . . 143
Figure 5.29: Boxplot of the peak fitting amplitudes for the 1597 and the 2195 cm^{-1} DC473 Raman bands when exciting with the 532 nm laser line 143
Figure 5.30: Boxplot of the integrated intensity of the CH-stretching Raman bands (2800-3080 cm^{-1}) for each cell line following incubation with DC473 or DMSO . . . 144
Figure 5.31: Raman spectra of 100 μM DC473 in DMSO before and after UV light exposure, compared to the DMSO pure spectra . . . 145
Figure 5.32: Example of the Raman single-cell response to DC473 PDT activation with UV light . . . 147
Figure 5.33: Examples of Bright field images of cells before and after PDT following incubation with DC473 . . . 147
Figure 5.34: Average of the single cell Raman data before and after each DC473 PDT experiment for the adenocarcinoma cell lines . . . 148
Figure 5.35: Average difference spectra (after - before) of the DC473 PDT data for the adenocarcinoma cell lines . . . 149
Figure 5.36: PCA loadings for the first 5 PCs for the DC473 PDT Raman data . 150
Figure 5.37: PCA average score-time traces for the first 5 PCs for the DC473 PDT data for each adenocarcinoma cell line . . . 151
Figure 5.38: PCA traces for PC4 and PC2 for the DC473 PDT data for the adenocarcinoma cell lines . . . 153
Figure 5.39: Beeswarms plots of the LDA scores and LDA loading for the DC473 PDT Raman data . . . 154

Figure 5.40: Average time changes of the amplitude, position and width of the DC473 bands from the 532 nm excitation Raman spectra single cell GLS fittings . . . 155
Figure 5.41: Average time variation of the GLS Fitting results for the 747 cm^{-1} cyt C Raman band in the single cell adenocarcinoma cells DC473 PDT Raman data excited with the 532 nm laser: amplitude, position and width . . . 157
Figure 5.42: Average time dependence of the integrated intensity from 1218-1238 cm^{-1} (highlighted by the LDA loading) of the single cell adenocarcinoma Raman data excited with the 532 nm laser, following DC473 PDT. . . . 157
Figure 5.43: PI-FDA live-dead assay for adenocarcinoma cells following incubation with different concentrations of DC473 and exposure to 405 nm UV light . . . 159

Figure 6.1: Chemical formulas for the chemical components of PDMS base and crosslinker . . . 162
Figure 6.2: Raman spectrum of polymerized PDMS excited with the 532 nm laser 163
Figure 6.3: Raman depth scan on a microfluidic device using the 532 nm excitation line illustrated with the intensity of selected bands for PDMS, water, immersion oil and glass. . . . 164
Figure 6.4: Raman spectra for the Z-scan in a 25 μm microfluidic device at different depths, obtaining the signal of the PDMS, water, glass and immersion oil . . . 165
Figure 6.5: picture and schematics of a microfluidic PDMS device with an evaporated 50 nm layer of Aluminium. Reflective with light image of 532 nm laser damage on the Al layer. . . . 166
Figure 6.6: Raman spectrum of PDMS with and without an Al evaporated layer excited with the 532 nm laser . . . 166
Figure 6.7: Raman spectrum of PDMS with and without a Cr/Au evaporated layer excited with the 532 nm laser . . . 167
Figure 6.8: Schematics and pictures of the PDMS devices with evaporated Cr/Au. Bright field images of the device, showing trapping of HL60 cells and 532 nm laser effects on the cells and overheating causing water evaporation. . . . 168
Figure 6.9: Raman spectrum of trapped HL60 in Cr/Au PDMS device when exciting with the 532 nm laser . . . 169
Figure 6.10: Raman Silicon spectrum when shielded with a gold nanocomposite compared to signal from the silicon sample exciting with the 532 nm laser . . . 169
Figure 6.11: Pictures and reflective white light images from the glass-SU8 device fabrication results . . . 170
Figure 6.12: Raman PDMS background contribution by the glass coverslip in a microchannel – changes with channel height when exciting with the 532 nm laser . . . 171
Figure 6.13: Raman signal of an HL60 single cell in a 129 μm deep microfluidic chamber excited with the 532 nm laser . . . 172
Figure 6.14: Raman single-event of HL60 cell crossing the laser scanning point on a microfluidic channel, detected in the CH stretching region (exciting with the 532 nm laser) . . . 173
Figure 6.15: Raman mapping of a trapped cell in a PDMS microfluidic device exciting with the 532 nm laser . . . 173

Figure 6.16: Loadings for PCA on Raman mapping of a trapped cell in a PDMS trap exciting with the 532 nm laser . . . 174
Figure 6.17: Average Raman signal of a cell and a trap obtained from the thresholded PCA scores from Raman mapping on a trapped cell using the 532 nm laser . . . 175
Figure 6.18: Raman map of a trapped HCT116 cell exciting with the 532 nm laser, including the PCA loadings and false colour images of the PCA scores 176
Figure 6.19: PDMS background fitting and subtraction example from a SW480 single cell Raman spectrum when exciting with the 532 nm laser . . . 178
Figure 6.20: Takeuchi-like traps design for Raman on chip, including bright field images of trapped SW480 cells . . . 179
Figure 6.21: Raw spectra before and after incubation with DC473 *on chip* exciting with the 532 nm laser . . . 180
Figure 6.22: Confocal fluorescence images of the *on-chip* SW480 cells after DC473 incubation . . . 180
Figure 6.23: Average Raman spectra of SW480 cells before and after incubation with DC473 *on chip excited with the 532 nm laser* . . . 181
Figure 6.24: Cell index dependence of the Raman intensity of the band at 878 cm^{-1}, excited with the 532 nm laser . . . 181
Figure 6.25: Boxplots and beeswarm plots for the Raman integrated intensities of selected bands from the spectra excited with the 532 nm laser . . . 182
Figure 6.26: PCA scores and loadings for the *on-chip* Raman DC473 uptake SW480 cells spectra excited with the 532 nm laser . . . 183
Figure 6.27: PC1 vs PC3 scores scatter plot for the Raman *on chip* SW480 results before and after incubation with DC473 (ex. 532 nm) . . . 184
Figure 6.28: PCA/LDA scores and loading for prediction of low and high SW480 uptakers for DC473 *on chip* Raman data (ex. 532 nm) . . . 185
Figure 6.29: Raw and pre-processed Raman data for single *E. coli on chip* (ex. 532 nm) . . . 188
Figure 6.30: Average of *on chip* Raman spectra (ex. 532 nm) of single *E. coli* . . . 188
Figure 6.31: Device for measuring bacteria *on chip* using Raman spectroscopy . . . 189

List of Tables

Table 1.1: Composition of a cell in weight percentage and number of molecules . 5
Table 1.2: TNM and modified Dukes system stages 11

Table 2.1: Example of a confusion matrix, showing the number of observations in each class. 44

Table 3.1: Available objectives and characteristics 51
Table 3.2: Intensity for DAPI filter settings 52

Table 4.1: Tentative peak assignment for cell Raman spectrum 95
Table 4.2: Confusion Matrix of PCA/LDA 107

Table 5.1: Peak labelling of cyt C and PPIX standard spectra 116
Table 5.2: Peak labelling of DMSO 130
Table 5.3: DC473, DC271 and EC23 bands positions, intensities and assignments 131
Table 5.4: Spectroscopy values for DC473 in different solvents 133
Table 5.5: Summary of alkyne fitting parameters for the accumulation of DC473 in each adenocarcinoma cell line 144
Table 5.6: Time dependence of DC473 fitting parameters 156

Table 6.1: PDMS bands positions, intensities and assignments 163

Table 7.1: Summary of mechanical phenotyping on HL60, SW480, HT29 and SW620 196

Part I

Introduction and Theory

Chapter 1

Introduction

This PhD forms part of a joint project between the University of Leeds and Leeds Teaching Hospitals NHS Trust in order to develop a microfluidic platform that will integrate different phenotyping techniques that are first to be developed separately. The main assays proposed were Deformation Cytometry as a mechano-phenotyping assay [3]; Surface Acoustic Wave-induced dielectro-spectroscopy as an electro-phenotyping assay [5], and Single-Cell Raman spectroscopy as a chemical phenotyping assay [6]. The purpose of the platform was to be able to distinguish between healthy and diseased cells in a population in a non-destructive manner. By characterizing them mechanically and chemically, a deeper understanding of the changing properties of diseased cells can be found. This PhD contributed by developing the Raman expertise of the group on measuring live single-cells and optimizing the Raman acquisition on a microfluidic device. Furthermore, Raman was used to detect the accumulation of drugs in live single cells. For this, the detection of photodynamic compounds was proposed as a model system.

1.1 Background to Single Cell Analysis: Combining Raman and Microfluidics

The mapping of tumours, from sub-cellular to whole organ length scales represents a major challenge in cancer research for understanding how biological changes relate to pathology. Most of the biological knowledge between genotype and phenotype of cells come from average measurements of a cell population.[7] Even though they can be useful, this sort of measurements cannot assess the individual response of single cells within a heterogeneous population. For example, an average study might see a gradient transition of a concentration of a certain factor while single cell analysis might show an *all-or-nothing* switch. Only by understanding the single components of the system, can the whole system be understood.

Single-cell techniques are more time consuming and require the analysis of large numbers of cells in order to be statistically significant. Cell cytometry is a high-throughput technique but requires fluorescent labelling.[8] In the last decade, high-content screening has been key in drug discovery, often using fluorescence reporters.[9] Cellomics reffers to the combination of automated microscopy with image analysis applied to drug discovery and cell biology,

both HTS[1] and HCS[2] [10], and can also be combined with microfluidics.[11] However, often HTS techniques are based in fluorescent labels, which can be cytotoxic, do photobleach, only allow a limited number of stains, due to spectral overlap, and require extensive and expensive preparation steps.[10] Some other single-cell analytical techniques are available such as single cell chemical cytometry techniques but are destructive and very low-throughput. A high-throughput and high-content label-free non-destructive technique to determine the chemical composition of the cell is not yet available [12].

RS[3] is an analytical technique based on inelastic scattering of light when it interacts with matter. This effect was first predicted by Smekal in 1923 [13] and reported in 1928 by C.V. Raman and K. S. Krishnan in their Nature Letter [14]. Raman scattering is inherently weak and was not widely used until the late 1960s when high power lasers were introduced.[15] RS and microscopy have since been used in a variety of systems, including for biological research.[15] RS offers a label-free, technique for studying biological systems. Another big advantage of RS is that it is a high-content technique, obtaining multiple information from a single measurement, both when doing spectroscopy and microscopy. Importantly, the technique can be used to probe living systems, providing biochemical information with sub-cellular and cellular spatial resolution on live cells.[8, 16–19] It allows the discrimination between cell types, at the single-cell level, and thus has a potential for application in studying cell heterogeneity, differential response to drugs, automatic mapping of tissue samples and microfluidic-based identification of cancers.[20]

In order to interrogate single cells in a high throughput fashion, it is necessary to be able to manipulate small volumes of fluid in a precise way and immobilize the cells while the measurement is taking place.[21] In microfluidics, fluid-based processes are reduced to the microscale, with channels in the tens of microns in height and depth fabricated in a solid substrate, making devices just some centimetres long where volumes of the order of nanolitres or picolitres are manipulated.[22] Microfluidic devices have several advantages over conventional large volume, bench top bioanalytical techniques: low reagent and power consumption, low cost, the potential for mass production and massively parallel scale analysis [12]. When using single cells, microfluidics would provide the perfect platform to push RS to higher throughput, and thus the compatibility of Raman with microfluidics should be explored. One possible application of this technique would be the RACS[4], a promising label-free technology that could provide information on the chemical composition of the cell at high throughput [23].

1.2 Raman Spectroscopy

RS[5] is a vibrational spectroscopic technique in which laser light interacts with a sample and a spectrum is obtained. When photons from a laser light source interact with a vibrating molecule, most of them retain their energy and undergo elastic scattering also known as Rayleigh scattering. However, approximately one in a million photons changes its frequency

[1] High-throughput screening
[2] High-content screening
[3] Raman spectroscopy
[4] Raman activated cell sorter
[5] Raman spectroscopy

in what is known as Raman scattering.[24–27] The photon may either lose (Stokes) or gain (anti-Stokes) energy, causing the frequency to drop or increase. The energy gained or lost by the photon is associated with a change in the vibrational state of the molecule.[28] Linear RS is sometimes referred to as spontaneous RS in this thesis. Here, the technique used is Stokes Raman confocal microspectroscopy, where the scattered light goes through a slit to reduce out of focus contributions to the spectrum. Section 2.1 (page 24) has an in-depth description of the theory of RS and section 3.1 (page 49) has a description of the system used.

1.2.1 Raman Spectra Analysis

Cells have a rich spectral content, which provides Raman with great potential as a diagnostic tool [29]. Tentative assignments of the main spectral features usually observed in cells have been reported in the literature [17, 30, 31]. Some reviews have collected the main relevant bands for Raman single cell analysis [32, 33]. However, as the cell spectrum is very complex and multiple molecules contribute to the same regions, it is a difficult task. Also, peak labelling is useful for the biochemical understanding of different cell states but is poor for population discrimination.

When the aim is to classify different cell lines, the differences between cell types are usually subtle. This, coupled with the need to sample large numbers of cells, means that multivariate analysis for discrimination between cell types or states is required.[34] Various chemometric methods have been employed in the literature to identify the main spectral variations in pre-processed data.[34, 35]

Section 2.2 (page 32) describes in more detail the theory behind the main multivariate methods used in this thesis. The most common method for dimensionality reduction is PCA[1]. A derivation of PCA is given in subsection 2.2.1 (page 33). Additionally, more complex multivariate methods can be used to improve the classification of the chemometric data. PLSR[2] is described and compared to PCA (page 37). Supervised multivariate methods are described in more depth in subsection 2.2.2 (page 38), where LDA[3], DT[4] and SVM[5] are introduced in pages 38, 41 and 42 respectively.

1.3 Raman Spectroscopy of Biological Systems

1.3.1 Raman Spectroscopy of Single Cells

Cells are composed of thousands of molecules of different types and concentrations, all of which simultaneously contribute to the Raman spectrum of the cell [37]. For a certain molecule with N atoms, the number of possible bands would be equal to the number of

[1] Principal Component Analysis
[2] Partial Least Squares Regression
[3] Linear Discriminant Analysis
[4] Discrimination Trees
[5] Support Vector Machines

Table 1.1: Composition of a cell in weight percentage and number of molecules adapted from Alberts 2002 [36].

	Weight (cell %)	Number of types
Water	70	1
Macromolecules (proteins, nucleic acids, and polysaccharides)	26	3000
Inorganic ions	1	20
Sugars and precursors	1	250
Fatty acids and precursors	1	50
Amino acids and precursors	0.4	100
Nucleotides and precursors	0.4	100
Other small molecules	0.2	300

vibrational degrees of freedom $3N - 6$ (for a non-linear molecule). However, not all of them will be Raman active [38]. Also, even though the number of different molecules in a cell is very high, their abundance might not be (see figure 1.1 and table 1.1). Around 70% of the weight of a cell is water, and the remaining weight is mainly macromolecules (proteins, NA[1], and polysaccharides). The limit of detection of RS[2] depends on the cross section of the molecule of choice.[15] UV-RS can detect proteins at a concentration of 1 mg/mL, [39] and thus RS proves the overall ratio of macromolecules within every single cell.

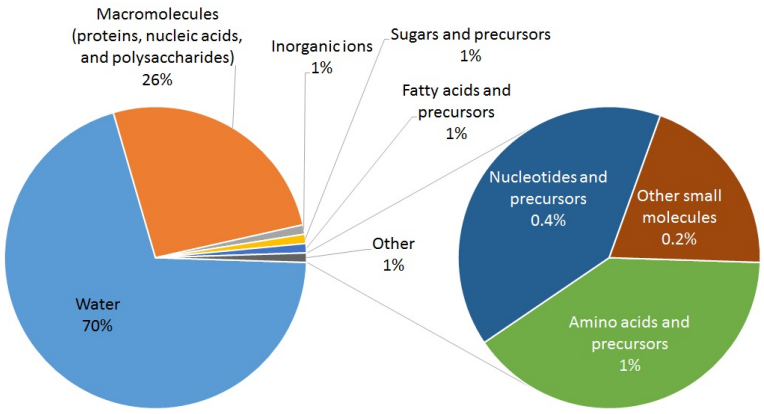


Figure 1.1: Pie charts of the approximate chemical composition by weight of a bacterial cell, based on the reported values from Alberts *et al.* 2002 [36].

An example of a labelled cell spectrum is shown in figure 1.2A. Two main regions can be identified: the region below 1800 cm^{-1}, also known as the fingerprint region, with multiple small peaks, and the region above 2800 cm^{-1}, with a very strong and broad band. Figure 1.2B shows the average spectra of the cell membrane, the cytoplasm, the nucleus and a lipid droplet, reproduced with permission from El-Mashtoly *et al.*[40].

The $C - H$ bonds give rise to strong Raman signals in the $C - H$ stretching region around 3000 cm^{-1} where the sp^3 carbons have a contribution below 3000 cm^{-1} and the sp and sp^2 carbons above 3000 cm^{-1}. Bending modes have contributions around 720, 1460 and 1375 cm^{-1} [38]. Because $C - H$ bonds are very frequent in lipids - even though they are also present in multiple other molecules - the strong band around 3000 cm^{-1} is usually labelled as the main lipid band or CH-stretching band. Lipid droplets show a strong band in this region, as shown in figure 1.2B.

[1] Nucleic acids
[2] Raman spectroscopy

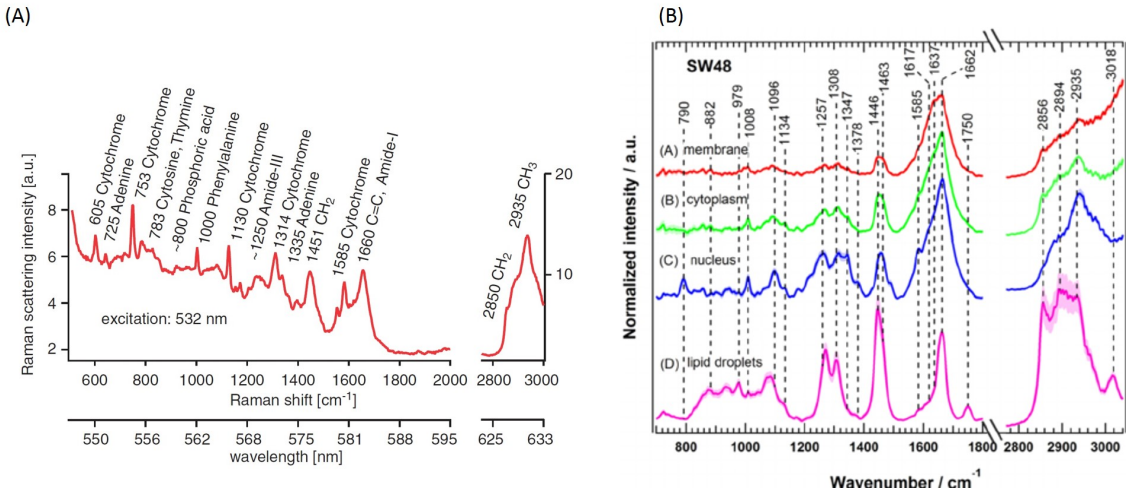


Figure 1.2: (A) Example of the Raman spectra of the cytosol of a HeLa cell using a 532 nm laser reproduced from Palonpon *et al.* 2013 [17] with permission from Elsevier. The main relevant peaks have been tentatively labelled by the authors. (B) Raman spectra from cluster analysis of the cell membrane, the cytoplasm, the nucleus and lipid droplets obtained with a 532 nm laser from a SW48 CRC cell, adapted from El-Mashtoly *et al.* 2015 [40]. Copyright 2019 American Chemical Society.

Another very relevant contribution to the spectra is due to proteins, specifically due to the amide bond ($-NH-CR-COOH-$). The $C=O$ stretch of this bond gives a broad peak near 1650 cm^{-1} called the Amide I band. $N-H$ bend and $C-N$ stretch contribute to the Amide II band near 1550 cm^{-1}, and $N-H$ stretch and $C-N$ bend cause the Amide III band near 1300 cm^{-1}. However, Amide II band is very weak in near-IR and can not be seen in the absence of resonant excitation.[41] The shape and intensity of the Amide I and III bands can give information about the secondary structure of proteins.[42, 43] Other relevant contributions to the protein spectra are the disulphide bond stretching ($S-S$) with a band located between 500-550 cm^{-1} and aromatic amino acids. Phenylalanine shows a very sharp peak around 1002 cm^{-1}, whereas the peaks of tryptophan (1010, 1340 and 1360 cm^{-1}) and Tyrosine (830 and 850 cm^{-1}) are usually weaker [44].

NA also have a very important contribution to the spectra, with the thymine, cytosine, uracil contribution around 780 cm^{-1}, the adenine contribution around 726 cm^{-1} or the phosphate group contribution around 800 cm^{-1} among others.

Some useful reviews with RS tables for the tentative labelling of peaks of biological samples are Talari *et al.* [45] or Movasagui *et al.* [32].

1.3.2 Raman Spectroscopy for Medical Diagnosis

RS is a sensitive and non-destructive technique suitable for analysis of living cells. However, it faces a number of challenges when applied to diagnosis: firstly, cells are inhomogeneous, so an average spectrum of the whole cell needs to be obtained [25]; secondly, long acquisition times – of the order of tens of seconds per cell – are usually needed as the signal is weak and high laser powers could compromise the integrity of the sample [46, 47], and thirdly, Raman signals are easily overwhelmed by other signals such as autofluorescence of the sample, that shows as a baseline [25], or stronger Raman signals from the substrate. Kamemoto *et al.*

2010 did a study of the background of glass slides in near-IR[1] Raman microscopy, concluding that substrates such as quartz, MgF or CaF_2 should be used to minimize the background signal [48] (also mentioned in Downes *et al.* 2010 [46]).

There are some alternative Raman-based techniques designed to increase the signal intensity such as RR[2] or SERS[3]. In RR the excitation source matches the resonance band of the molecule of interest [25, 49] – it occurs when the virtual state lies close to an electronic excited state [27, 43] An example is the RR at 532 nm of cyt C[4] [1, 50] or PPIX[5] [51]. Subsection 2.1.5 in page 32 has a more in-depth description of RR. In SERS, the signal is amplified by the large electric fields excited at metal surfaces of nanoparticles.[23] There are numerous examples of Raman microscopy in cells using SERS in the literature [52, 53], some of them reviewed in Cialla *et al.* 2012 [54]. However, can either be used as a label, targeting a particular molecule in the cell, or can depend on the cell uptake of nanoparticles, that will often accumulate in an inhomogeneous manner within the cell.[55] Additionally, the nanoparticles' accumulation can cause cytotoxicity.[55]

There are many good reviews in the literature about the multiple biomedical applications of Raman microscopy. A general review comparing the biological applications of Raman and IR microscopy in cell biology can be found in Matthäus *et al.* [16] and more recently in Krafft *et al.* [25], from monitoring overall cell composition to classification of cell types, imaging of mitochondria, tracking of labelled liposome cell uptake or bacterial cell classification. The book from Rehman *et al.* *Vibrational Spectroscopy for Tissue Analysis*, has an extensive review of vibrational spectroscopy covering applications to Cancer for tissue samples [33]. A review on Raman microscopy as an imaging tool can be found in Palonpon *et al.* [17]. For example, Klein *et al.* used Raman microscopy in conjunction with fluorescent tags to obtain the spectra of each organelle within a cell [56], and Swain *et al.* used Raman to study the chemical composition of a cell line in different cell cycle stages.[57] Another good example is Smith *et al.* that studied single exosomes using RS and optical trapping, finding four types of exosomes common to multiple different cell lines.[58] In diagnosis, Raman studies have been mostly focused on cancer. Raman in breasts cancer has been studied by multiple groups [31, 47, 59–61] for breast cancer cell lines, where lipid peaks and high protein content were fundamental features of cancer cells. Kamemoto *et al.* 2010 [48] studied cervical cancer with near-IR Raman and saw differences mainly in collagen-related peaks, and Oshima *et al.* 2010 discriminated lung cancer cell lines from normal cell lines [62], among other examples [63–65].

1.3.3 Live versus Fixed Cells

When doing RS on single-cells, long acquisition times are needed for high-quality spectra –of between tens and hundreds of seconds per cell–, which can limit the number of cells measured in each experiment. For example, Swain *et al.*[57] needed 40 s per spectra averaging in 4

[1] Infrared
[2] Resonant Raman
[3] Surface enhanced Raman spectroscopy
[4] Cytochrome C
[5] Protoporphyrin IX

different spots of the cells, and this limited the population size to be analysed. In another example, Krishna *et al.*[30] used 90 s acquisition spectrum averaging 3 times. Live mammalian cells degrade quickly, and when kept starved in a buffer like DPBS[1] and at room temperature, they will only last viable for some hours depending on the cell line. For this reason, finding ways of freezing the metabolic state of the cell in a long-lasting way is very attractive for RS studies. Fixation methods effectively kill the cell in a controlled long-lasting manner, allowing the use of higher laser intensities without risking the integrity of the cell. Additionally, after measuring, the sample can be stained and the spectra can be correlated with the staining information.

Tissue studies When working with tissue, multiple tissue donation programs with fixed samples are available (e.g. the Leeds GIFT 2 Research Tissue Project [66]). These fixed samples have been stored for years, and the outcome from the patients is already known. Two main procedures can be used in order to fix the tissue: the tissue can be flash-frozen, cryosectioned, thawed and dried; alternatively, it can be formalin fixed and paraffin-embedded, followed by sectioning of the block and optional deparaffination [35, 67]. Flash-freezing is generally preferred as it avoids the use of organic solvents, and is the most similar to fresh samples [67]. Less known alternative methods are ethanol fixation followed by paraffin embedding which has been reported to keep more proteins than formalin-fixed samples [67].

The validity of these fixation methods for Raman studies is controversial in the literature. On the one side, it was noted by Diem *et al.*[35] that there are publications that showed similar images and conclusions using both frozen and paraffin-embedded tissues. On the other side, other publications [61, 67–69] and reviews [70] have shown that formalin-fixed samples showed a decrease in lipid and protein content, an overall weaker cell signal, and new peaks due to the treatment, as well as shifts in some bands. When working with fixed and then paraffin embedded tissue, the de-waxing process can sometimes leave wax behind, that is detected by RS.[71] Another issue related with formalin fixation is that while fixing, cells on the surface of the tissue are exposed to formalin for longer than the ones inside it, causing a spatial heterogeneity. Alternatively, non-deparaffinized samples have been successfully used for Raman mapping [67], but the visualization of unstained tissue becomes more difficult, and the paraffin signal needed to be subtracted. Formalin-fixed samples have been widely used in the literature [31, 35, 72, 73]. Multiple comparative studies have been done between fresh, flash-frozen and paraffin-embedded tissue, showing that fresh tissue always has a greater signal, followed by flash-frozen tissue [67].

Single-cell studies For single-cell studies, the standard is paraformaldehyde fixation, that polymerizes the proteins within the cell. Meade *et al.* 2010 [74] studied the effects of chemical fixation in different cell lines, showing that all fixation methods affected the vibrations of lipids, proteins, NA and carbohydrates. They concluded that formalin fixation gave the most similar spectrum to live cells. However, the use of fixed cells not only affects the signal obtained from the cells but also limits the type of studies that can be undertaken. When using a low enough dose, RS is non-phototoxic, and thus can be used to track changes in

[1] Dulbecco's Phosphate-Buffered Saline

a cell over time. This opens the door to time-dependent studies on single cells, such as prediction of drug response or metabolism of a drug over time inside a cell. Raman on live cells would also allow using the cells later with a different technique that may require the cell to be alive, such as drug resistance measurements on cancer cells, migration studies or cell to cell interactions. For diagnostic proposes, working with live cells is closer to endoscopic Raman probes application, where Raman can be done in situ during an endoscopy, allowing diagnosis before the extirpation of the abnormal mass such as the SERS-fluorescence endoscopic probe used by Jeong *et al.* (2015) [75] or the CARS[1] flexible endoscope by Lombardini *et al.* (2018) [76] among others [77–79].

Among the live cell applications, RS has previously been used on live cancer cell lines. Oshima *et al.* did a RS study of one healthy and four cancer lung cell lines, using a 532 nm laser excitation, showing that cyt C was the major source of variability between cells. However, only five measurements were taken from each cell culture disk and with a 1 μm laser spot, obtaining very specific signals from the cells instead of an average. Pully *et al.* [80] used RS for time-lapse imaging of human lymphocytes from healthy donors using a 647.1 nm laser. After 10 images of 2 minutes acquisition each, the cell showed blebbing and loss of cytoplasmic and nuclear material, likely caused by both photodamage and the acquisition conditions, done in a buffer and at room temperature. Authors observed a drop in 1524 and 1154 cm^{-1} bands that they labelled as carotenoids, but no change in protein or NA bands. LTRS[2], where the cell is optically trapped and Raman is acquired, often uses live cells, like in Harvey *et al.* [81], where they classified two cell lines of bladder and prostate cancer, acquiring spectra of around 20-30 cells from each cell line. Another example is Casabella *et al.* [82], where they used laser trapping RS in a PMMA[3] microfluidic device to analyse live prostate cancer cells and lymphocytes from cell lines. Their microfluidic approach allowed them to analyse hundreds of cells.

1.3.4 Colorectal Cancer as a Model System

CRC[4] is the third most common cancer worldwide and second in mortality, with over 1.8 million new cases diagnosed in 2018, and more than 550k million deaths [83] and it is expected to increase to more than 2.2 million cases and 1.1 million deaths by 2030 [84]. Approximately half of the patients develop metastasis to the liver [85] and resistance to standard chemotherapy [86]. CRC has estimated mortality of 56% (2012) and around 20% of diagnosed patients already have metastases at diagnosis.[87] CRC is a disease associated with developed countries, with almost 70% of all cases and about 60% of all deaths occurring in countries with high human development index.[84] The risk of developing it is related to bad alimentary habits, smoking, intestinal inflammatory disease, polyps, genetic factors and ageing - 90% of patients diagnosed are older than 50.[88] CRC is also one of the most preventable cancers because it (almost) always arises from benign neoplasms, which evolve into CRC over many years, offering an opportunity to detect and remove the polyps before

[1] Coherent anti-Stokes Raman spectroscopy
[2] Laser tweezers Raman spectroscopy
[3] Poly(methyl methacrylate)
[4] Colorectal Cancer

they undergo malignant transformation.[89] Isolating the chemical fingerprint of metastatic colorectal cells will aid tissue and single-cell studies on the effectiveness of pre-operative treatments and tumour identification, and can lead to the development of in situ diagnosis of polyps during a colonoscopy.

There are multiple methods to detect CRC, such as the guaiac test, immunochemical test of stools, DNA stool test, sigmoidoscopy, colonoscopy and barium enema.[88] It usually presents with symptoms such as abdominal pain, alterations of chronic bowel habits and changes in bowel movements, involuntary weight loss, nausea, vomit, malaise, anorexia and abdominal distension. Distal cancers cause rectal bleeding, whereas proximal cancers can give mixed blood with stool.[88]

The carcinogenic model of CRC include the suppressor or classic pathway – also known as the CI[1] pathway – and the mutator or alternative pathway – also known as the MSI[2] pathway, usually with hereditary deficiencies in the DNA mismatch-repair system.[88, 90] Additionally, cancers with the CIMP[3] phenotype exhibit aberrant DNA methylation.[90] The CI pathway is mainly sporadic (80%-85% of sporadic CRCs) with typically aneuploid malignant cells whereas the MSI pathway is mainly hereditary (80% of hereditary CRC), and presents a wide range of mutations [88, 90] that dysregulate cell cycle progression and metastasis, resulting in aberrant Wnt/β-catenin signalling - usually due to mutation of APC[4] - occurring in 80% of CRC, gain in function mutations in K-ras and PI3K[5]/Akt, COX2[6] over-expression (mutated in 80-90% of CRC), reduction in the activation of PPARγ[7] and loss of p53 function (mutated in 50% of CRC). Consequently, cells increase proliferation and invasion and decrease cell to cell interaction and differentiation.[85]

CRC usually starts as a polyp - a pendunculated structure projecting from the intestine mucosa, but can also start as an initially benign lesion called adenoma (adenomatous polyps).[88] 40% of the Western population will develop adenomas, however, only 3% will suffer from CRC, meaning only a small proportion of adenomas will progress to malignancy, [91] and 24% of patients with polyps will develop CRC [88]. The adenoma-carcinoma sequence, also known as polyp-carcinoma progression sequence [89], is the widely accepted mechanism that explains the process by which most CRC arise.[91] It includes genetic mutations, epigenetic alterations and local inflammatory changes.[89] The sequence is heterogeneous and involves multiple different pathways to CRC. Normal glandular epithelial cells accumulate genetic and epigenetic changes and transform into invasive adenocarcinomas.[89] In the original Fearon and Vogelstein model, from the healthy colon epithelium, cells transform into a benign neoplasm, in the form of an adenoma. A next step promotes the progression to a more histologically advanced neoplasm, that then transforms into an invasive carcinoma.[89]

CRC is currently classified following the Astler-Coller-Dukes system (modified Dukes system)

[1] Chromosomal instability
[2] Microsatellite instability
[3] CpG island methylator phenotype
[4] Adenomatous polyposis coli
[5] Phosphatidylinositol-3-kinase
[6] Cyclooxygenase-2
[7] Peroxisome proliferator-activated receptor γ

Table 1.2: Equivalences and description of the TNM and modified Dukes system stages, adapted from Singla *et al.* [92]

TNM stage	Modified Dukes stage	Description
T1N0M0	A	Limited to submucosa
T2N0M0	B1	Limited to muscularis propia
T3N0M0	B2	Transmural extension
T2N1M0	C1	T2, enlarged mesenteric nodes
T3N1M0	C2	T3, enlarged mesenteric nodes
T4	C2	Invasion of adjacent organs
Any T, M1	D	Distant metastases present

or TNM system.[88] The Astler-Coller-Dukes system uses stages A to D, where A corresponds to a malignancy limited to the mucosa, B when it extends into the muscle layer, C when lymph nodes are involved and D with distant metastatic spread. Alternatively, the TNM system classifies according to tumour size (T), nodes involved (N) and a number of distant metastasis (M). A table of equivalences between these two systems is shown in table 1.2, adapted from Singla *et al.* [92].

1.3.5 Model System

For this study, the main cell lines chosen were SW480, derived from a primary Duke's stage B adenocarcinoma (equivalent to a T2-3 N0 M0) and SW620, derived from a secondary tumour in a lymph node from the same patient (Duke's stage C, equivalent to a T2-4 N1 M0) [93–95]. These cells were first isolated and described by Leibovitz *et al.* [95] in 1976 from a respectively 50 and 51 years old while male patient (blood type A, Rh+) from the colon and a lymph node one year later. Using these cell lines can help isolate metastasis variability from the person-to-person variation. Previous reports of vibrational spectroscopy on the SW620/SW480 model system at the single-cell level have been undertaken using synchrotron FTIR[1] microspectroscopy on live cells [96] and Raman spectroscopy of a small number of fixed cells combined with SRS[2] [97].

In addition, HL60, HCT116 and HT29 cells were added to the model. HL60 is a non-adherent blood cell line derived from human promyelocytic leukaemia and was used to prove the ability of Raman to differentiate between cell lines with very different origins. Both HT29 and HCT116 cell lines are colorectal. HCT116 cells are derived from human colon carcinoma Duke's D from the primary site (equivalent to T1-4 N1+ M1+), so are expected to show similarities with primary colon cancer cell lines, particularly HT29, and will test the ability of the system to separate between different cancer types from the same tissue. HT29 cells are derived from Duke's C stage human colon adenocarcinoma (equivalent to a T2-3N1M0) from a 44 year old white female patient (blood type A, Rh+) isolated in 1964 by J. Fogh [98] and are thus expected to show similarities with the SW480 cell line that is human colon adenocarcinoma Duke's stage B, challenging the system to differentiate between different stages of the same disease. Ahmed *et al.* [90] has a comprehensive analysis of CRC cell lines genetic and epigenetic properties. A schematic outlining the Duke's stages of the colorectal adenocarcinoma model system chosen is shown in figure 1.3.

[1] Fourier-transform infrared
[2] Stimulated Raman scattering

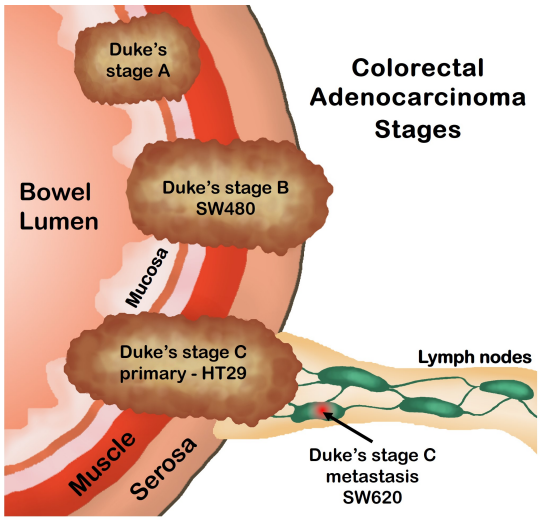


Figure 1.3: Schematics of the positions within the vowel linning or lymph nodes of each of the three adenocarcinoma stages chosen as the model system: SW480, HT29 and SW620.

Previous studies have done bulk Raman measurements in HL60 cell pellets [99, 100], on single-nuclei of HL60 cells [101] and on fixed HT29 cells [102]. Single-cell live label-free RS of these cell lines has been previously done comparing HCT116 cells with HT29 cells [103], studying apoptosis induction on HCT116 cells [104, 105], studying proliferation effects caused by co-culture of HL60 cells with mesenchymal stem cells [106] and comparing HL60 cells with peripheral blood mononuclear cells [107], but the number of cells analysed in these studies was always below 30.

1.4 Raman Spectroscopic Single-Cell Detection of Drugs

Drug efficacy or resistance in vitro is often tracked using single-protein assays, but this may miss the big picture and lead to responses that are not consistent with *in vivo* results.[108] RS[1] detects chemical functional groups [109, 110] in a label-free and non-destructive manner, with spatial resolution in the micrometer scale.[10, 18, 19, 111, 112] RS has been extensively applied on pharmacology characterization.[58] If the concentration is high enough, it can detect a compound within a cell in a semi-quantitative manner and obtain information on the response of the cell to the compound.[29, 113, 114]

However, drug peaks often appear in the fingerprint region, merging with bands from cell compounds such as proteins, lipids and NA[2].[115] For tracking a drug within a cell, alternative approaches can be used, such as RR[3] or the use of isotopologues such as deuterium labelling [116, 117] or biorthogonal Raman tagging such as alkyne tagging.[115, 118–120] Biorthogonal Raman tagging uses molecules with vibrational bands in the cell Raman silent region ($1800 - 2800\ cm^{-1}$) such as alkynes, enabling more accurate detection of compounds in single-cells.[2, 118–125]

RS has been used to observe both uptake, metabolism and response of cells to drugs.[58] For example, there is a line of research with multiple fixed cells microspectroscopic studies on the response of cancer cells to Tyr-kinase inhibitors for lung cancer cell resistance studies

[1] Raman spectroscopy
[2] Nucleic acids
[3] Resonant Raman

[108]; additionally, they detected metabolism of the drug within the cell [126, 127]. First (erlotinib) and second (neratinib) generation Tyr-kinase inhibitors used in these studies had a $-C \equiv C-$ or $-C \equiv N$ group with a unique signature in the Raman silent region, aiding in the detection of the drug within the cell using RS.[108, 126] They also detected changes induced by erlotinib [128] and sorafenib [129] in different organelles of fixed cells from the CRC[1] model system SW48, SW480 and HT29, where mutations in KRAS[2] and BRAF[3] in the latter two conferred resistance to erlotinib but not sorafenib, and effects of panitumumab on fixed SW48 (susceptible) and SW480 (resistant) cells [40]. Fu *et al.* also imaged Tyr-K inhibitors imatinib, nilotinib and GNF-2/-5 uptake and localization but on live chronic myelogenous leukaemia murine BaF3 single cells using SRS[4], showing that combinational therapy with chloroquine had synergistic effects by decreasing lysosomal trapping.[130] RS not only allows detecting the drug within a cell but also its chemical state, as shown by Yamakoshi *et al.* that used RS to image different chemical forms of a mitochondrial uncoupler with a nitrile Raman tag on live HeLa cells.[120]

A series of Spectroscopic studies have detected cisplatin binding to DNA in nuclei [131] and cell membrane and cytoplasm [132] of the lung adenocarcinoma cell lines A543, and doxorubicin accumulation and cell response at different incubation times for both A549 [133] and Calu-1 [134] on fixed cells. Mouras *et al.* used a multimodal approach to image changes induced by doxorubicin in live MCF7 cells and a doxorubicin-resistant MCF7 clone, using CARS[5] to image the cells and two-photon excitation fluorescence to image the accumulation of the drug.[135]

Internalization of biodegradable biocarrier systems such as PLGA[6] with different surface properties has been extensively studied with Raman microspectroscopy.[58] Previous work based at the University of Twente studied the mechanisms of degradation of PLGA microspheres in fixed RAW 264.7 macrophages using Raman microspectroscopy.[136] For instance, Xu *et al.* used CARS to image PVA[7] stabilized PLGA NP[8]s interaction with live KB cells, confirming that the NPs released the load but did not enter the cells [137] whereas both Romero *et al.* [138] and Chernenko *et al.* [139] saw uptake of PLGA NPs when stabilized with PEI[9] or BSA[10] [138] and pluronic acid [139] in fixed HepG2 and HeLa cells respectively. Chernenko *et al.* also used Raman microspectroscopy to show the targeting effect of EGFR[11] peptide-modified PLGA NPs encapsulating deuterated ceramide on fixed SKOV3 cells, showing faster uptake and release.[140] They additionally studied deuterated liposomes both modified with different cationic moieties [141] and with a cell-penetrating peptide [16] for Raman confocal imaging of uptake on fixed single HeLa and MCF7 cells respectively. See Matthäus *et al.* for a more detailed review on research on drug delivery systems.[6] Garrett *et al.* showed how

[1] Colorectal Cancer
[2] Kirsten RAt Sarcoma virus oncogene
[3] v-Raf murine sarcoma viral oncogene homolog B
[4] Stimulated Raman scattering
[5] Coherent anti-Stokes Raman spectroscopy
[6] Poly(lactide-co-glycolide)
[7] Polyvinyl alcohol
[8] Nanoparticle
[9] Polyethylenimine
[10] Bovine serum albumin
[11] Endothelial growth factor receptor

deuterated GCPQ[1] polymeric NPs could be measured in fixed tissue from a mouse model using CARS, showing its ability to study drug distribution and penetration of biological barriers such as the blood-brain barrier or the gut mucosa.[142] Confocal RS has been used *in vivo* for penetration of topical formulations in skin.[58]

While microspectroscopy studies give insights on the responses to the drugs of interest, mapping is time-consuming and limiting the number of cells imaged, that is usually in the tens. Additionally, using fixed cells does not allow for same single-cell response spectroscopy studies. SRS is an emerging technique that allows quick mapping on live cells, but its technical requirements make their translation to clinical and biological application challenging.[25]

1.5 Photodynamic Therapy

PDT[2] is a treatment method where a PS[3] drug is administered and preferentially accumulates in the target area - i.e. tumour cells - and is activated by the application of light of a certain wavelength. The light is absorbed by the PS that becomes activated, and will then react with other molecules, ultimately causing cell death through oxidative damage.[28, 143] PDT selectively destroys tumours using three mechanisms: direct cytotoxicity, indirect cytotoxicity due to vascular damage and immunomodulatory, activating an immune response against the tumour.[144] A description of the main physical events that occur during PDT is given in section 2.3 (page 45).

CRC[4] are prone to showing drug resistance, with almost all patients with CRC showing resistance to therapy. However, resistance to PDT is uncommon, although it has been induced in HT29 cells with different PS when using consecutive treatments.[143] PDT is also a minimally invasive technique and its localized treatment allows minimizing side effects. For this reason, PDT is gaining attention as a therapy to use in combination with other chemotherapy drugs to overcome resistance.[145]

PS subcellular location determines its effectiveness, mitochondria being one of the most desired targets.[145] PS location is usually determined using fluorescence, however, this can cause photodynamic activity that could alter the system and cause relocalization of the sensitiser. Sensitisers aggregation in cells can lead to a reduction or elimination of fluorescence.[28] For this reason, vibrational spectroscopy can detect both the free and the aggregated PS. Additionally, when doing PDT *in vivo*, the PS activating light and the Raman excitation light could be matched to be able to record both the drug contents and the phenotype of the cells in the region being treated, holding potential as a technique for simultaneous tumour margin identification [146] and treatment.

A review on PDT in CRC can be found in Kawczyk-Krupka *et al.* [143] and in Hodgkinson *et al.* [145]. Most PDT studies in the literature are done *in vitro* using human cell lines with porphyrins, chlorins, bacteriochlorins and phthalocyanines as PSs.[143]

[1] Quaternary ammonium palmitoyl glycol chitosan
[2] Photodynamic Therapy
[3] Photosensitiser
[4] Colorectal Cancer

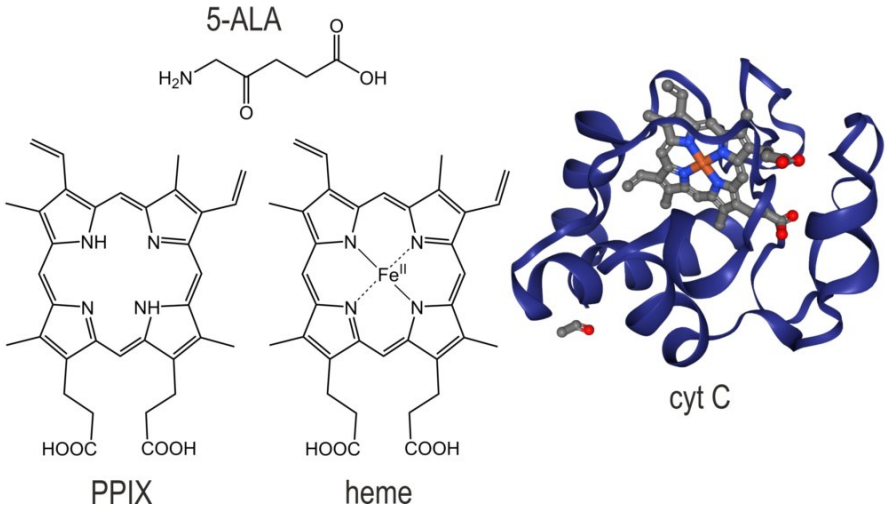


Figure 1.4: Chemical structures for 5-ALA, PPIX, heme and 3D structure of cyt C obtained from the Protein Data Bank [149].

1.5.1 Porphyrins

PPIX[1] is a very conserved molecule between all living organisms and the precursor of the heme group. It is a very hydrophobic molecule that exhibits photodynamic activity.[147] It is present with chelated metallic ions in metalloporphyrins such as the heme group in haemoglobin, cyt C[2] or cytochrome P450. Figure 1.4 shows PPIX and heme chemical structure, and the 3D structure of cyt C. Within cells, it should be present at low concentrations, with a very efficient conversion to heme by the ferrochelatase enzyme. Accumulation of PPIX can lead to skin photosensitivity, biliary stones, hepatobiliary damage and even liver failure.[147] PPIX occurs naturally in CRC, and both tumours and invaded lymph nodes in patients have been shown to exhibit red fluorescence with a 62% of sensitivity and 78% of specificity.[148]

PPIX synthesis occurs partially in mitochondria and cytoplasm with two possible final outcomes: either chelation with iron and conversion to heme or excretion via ABCG2[3] transporter.[147] The main point of regulation is ALAS[4], that synthesises 5-ALA[5]. High concentrations of heme can bond and destabilize the ALAS1 mRNA and downregulate ALAS1 import to mitochondria among other iron-related regulations.[147, 150] Genetic alteration of enzymes involved in its synthesis can lead to diseases such as porphyria, where a gain in function of ALAS2 in red blood cells leads to PPIX over accumulation.[147]

For this reason, when 5-ALA is present, the cell uptakes it and the main regulation point for porphyrins synthesis is bypassed, with the accumulation of all intermediates including PPIX.[147] figure 1.4 shows the chemical structure for 5-ALA. When 5-ALA is metabolised, porphyrins are produced and excreted from cells.[151] Cancerous cells have increased production of porphyrins [151, 152] that can be detected in the urine of patients with colorectal cancer when administering 5-ALA.[151] This is due to reduced activity of ferrochelatase and

[1] Protoporphyrin IX
[2] Cytochrome C
[3] ATP-binding cassette super-family G member 2
[4] Aminolevulinic acid synthase
[5] 5-aminolevulinic acid

increased activity of deaminase.[147, 150, 152, 153] Concentrations of 5-ALA found in the literature are usually of 1-2 mM, with incubation times between 3-4h.[143]

CRC cells naturally express low amounts of PPIX, showing considerably higher red fluorescence in tumour and higher fluorescence in metastasized lymph nodes compared to normal tissue.[148] Krieg *et al.* studied PPIX accumulation in CRC cell lines, with three cancerous lines (HT29, SW480 and CaCo2) and one stromal one (CCD18). 5-ALA uptake normalized to protein signal was found similar for all the cells, and the PPIX accumulation normalized to proteins was quantified. They saw that when incubating with 2 mM 5-ALA, subconfluent cultures accumulated less PPIX than confluent ones, with more accumulation of HT29 cells than of SW480 cells.[153]

Krupka *et al.* saw that when exposed to 5-ALA PDT (1-1.5 mM) and light (10, 30 and 60 J/cm^2), SW620 showed lower proliferation (MTT assay) and more cytotoxicity (LDH[1] assay) than SW480 cells, meaning that SW620 cells were more susceptible to PDT than their primary counterparts. Additionally, the amount of PPIX is not the only factor affecting the effectiveness of PDT, with evidence that increased PUFA[2]s increase the toxicity of PDT.[154] There is a concern that PDT in sublethal doses or under hypoxic - conditions that could be found inside a tumour - could enhance tumour vascularization and growth, and multiple groups have studied how the secretion of growth factors changes after sublethal doses of PDT. Krupka *et al.* observed that sub-lethal doses of PDT (1 mM of 5-ALA and 10 mJ/cm^2) caused a decrease in VEGF[3] secretion in SW620 cells - that naturally secrete more VEGF than SW480 cells, but no significant difference in SW480 cells.[155] In hypoxic conditions, PPIX production from 5-ALA was reduced, but also both cell lines were seen to reduce their secretion of FGF[4] and G-CSF[5], increasing their secretion of GM-CSF or CSF2[6], with a stronger response in SW620 cells, with an overall antitumorigenic response.[144] Previous experiments with SW480 and HT29 cells showed PPIX accumulations of the order of 16-20 ng of PPIX per mg of protein after 3-3.5 h in 5-ALA concentrations of 0.12-2 mM.[154, 156]

PPIX has also been described to target tumours directly, independently of light activation. Sznarkowska *et al.* showed that PPIX targets antiapoptotic factor p53 and p73 and leads to their accumulation in HCT116 cells, causing apoptosis. Deficiency on DNA mismatch repair did not confer HCT116 cells resistance to PDT.[157] PDT has also been described to induce autophagy, an undesired effect that protects from apoptosis.[143]

5-ALA was approved by the F.D.A.[7] in 1999 under the name *Ameluz* for PDT in a topical formulation, and has very recently been approved by the F.D.A. for oral administration as an intraoperative optical imaging agent in glioma surgery under the name *Gleolan*.[158]

[1] Lactate dehydrogenase
[2] Polyunsaturated fatty acid
[3] Vascular endothelial growth factor
[4] Fibroblast growth factor
[5] Granulocyte colony stimulating factor
[6] Granulocyte-Macrophage colony stimulating factor
[7] Food and Drug Administration

1.5.2 Novel Photosensitisers

Researchers at the University of Durham have developed a new low molecular weight, small photosensitiser, DC473, that elicits cell death through the stimulated production of ROS[1] when activated by UV-A, violet or corresponding near-IR[2] two-photon wavelengths.[159, 160] It was developed by David Chisholm during his PhD as part of a library of highly fluorescent synthetic retinoids.[160, 161] However, DC473 does not bind nuclear retinoid receptors and thus is not an active retinoid and was pinpointed as a promising photosensitiser. The chemical structure of DC473 is shown in figure 1.5. DC473 is a substituted diphenylacetylene with a highly lipophilic, electron rich tetrahydroquinoline donor and a methyl cinnamate acceptor.

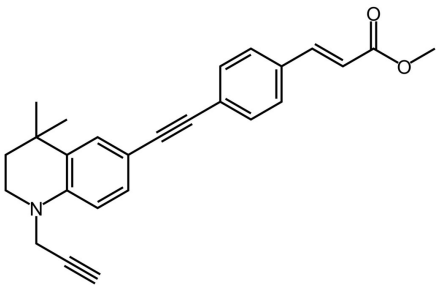


Figure 1.5: Chemical structure of novel photosensitiser DC473.

DC473 exhibits strong solvatochromatic fluorescence and possesses two alkynes ($-C \equiv C-$) with Raman vibrational peaks in the silent region of the spectrum, ideal for RS[3] of single cells. It was anticipated that this unique combination of structural and photophysical properties would enable a range of novel tandem fluorescence and Raman (fluoRaman) imaging methodologies and experiments. DC473 is commercialized by LightOx Ltd, a spin-out company from the University of Durham.

1.6 Raman Spectroscopy and Photodynamic Therapy

One of the first reports of Raman detection of PS[4] within cells is the publication by Freeman *et al.* in 1997, detecting zinc phthalocyanines in a fixed human endothelial hybridoma cell line.[162] Abramczyk *et al.* used Raman microspectroscopy to study the location of aluminium phthalocyanine, a PS, in non-cancerous and cancerous breast tissue, seeing preferential accumulation in the latter.[163] They also found that the dynamics of the photosensitiser in normal tissue was faster than in cancerous tissue.[164] From the same research group, Brozek-Pluska *et al.* used Raman microspectroscopy and fluorescence to study the distribution of Hematoporphyrin in healthy and cancerous breast tissue, finding lower accumulation in healthy tissue. Raman was used to provide a map of lipids and proteins, but could not be used to detect the hematoporphyrin, which was detected by using fluorescence.[165] Similarly, Brozek-Pluska *et al.* analysed the distribution of Phthalocyanine in cancer and healthy breast tissue by tracking its Q-band at ca. 5000 cm^{-1}.[61] *In vivo*, Bhattacharjee *et al.*

[1] Reactive oxygen species
[2] Infrared
[3] Raman spectroscopy
[4] Photosensitiser

did RS[1] measurements in rats with breast cancer models pre- and post- PDT[2] with chlorin photodithazine, showing detection of both PS accumulation and PDT effects.[166]

Detection of PS using SERS[3] has been previously shown, for example, Veloso *et al.* used SERS to study PDT with the PS chloroaluminumphthalocyanine on a mouse cancer cell line by depositing cell extract onto a silver film. [167] SERS detection of PS has been more frequently done by including metal nanoparticles in theranostic constructs with the PS. An example is the work from Fales *et al.*, that created gold nanostars labelled with Raman dye *p*MBA coated with silver nitrate and a porous silica shell, loaded with the PS PPIX[4], and with either a cell penetrating peptide or folic acid conjugation for targeting, showing detection and targeting in single cells.[168, 169] Farhadi *et al.* coupled palladium porphyrin with plasmonic nanoparticles to allow SERS detection of the PS, allowing the use of the same wavelength to excite Raman scattering and to stimulate PDT. [170] Multiple other groups have detected photosensitisers using SERS constructs.[171–177] In this thesis, we will focus on detection using label-free RS. To our knowledge, there have been no reports in the detection of PS accumulation in live single cells using RS.

1.7 Raman on a chip

Microfluidic platforms allow precise manipulation of small volumes of fluids, which can be a major advantage for single-cell analysis. For instance, microfluidic traps allow studying cells exposed to changing environmental conditions, like drug effects on single-cells.[20] Other techniques often implemented in microfluidics are FTIR[5] or ATR-IR[6], that are useful for analysis of homogeneous liquid samples but impractical for small volumes like the ones required in single-cell analysis. IR[7] spectroscopy is also strongly affected by water vibrations, which can complicate the analysis of water-based systems.[20] Raman spectroscopy has high resolution and is not affected by water background, and is thus the ideal vibrational spectroscopy technique for microfluidic integration.[20]

Current RACS[8] often suffer from low throughput due to the intrinsically weak signal of RS[9] and are often manually operated.[23] Given the long acquisition times needed in RS, the first generation of RACSs require of some sort of trapping, and it is common to find in the literature the use of LTRS[10].[23] An alternative approach is the RACE[11], where cells are investigated on a slide immobilized into a light absorbing layer, and a laser is used to eject the cells of interest exploiting the mechanism of laser induced forward transfer, so they can be collected for further investigation.[23] The main challenges are improving the quality of the acquisition, the cell isolation and the alignment of the cell with the laser spot.[23]

[1] Raman spectroscopy
[2] Photodynamic Therapy
[3] Surface enhanced Raman spectroscopy
[4] Protoporphyrin IX
[5] Fourier-transform infrared
[6] Attenuated total reflection infrared
[7] Infrared
[8] Raman activated cell sorter
[9] Raman spectroscopy
[10] Laser tweezers Raman spectroscopy
[11] Raman activated cell ejection

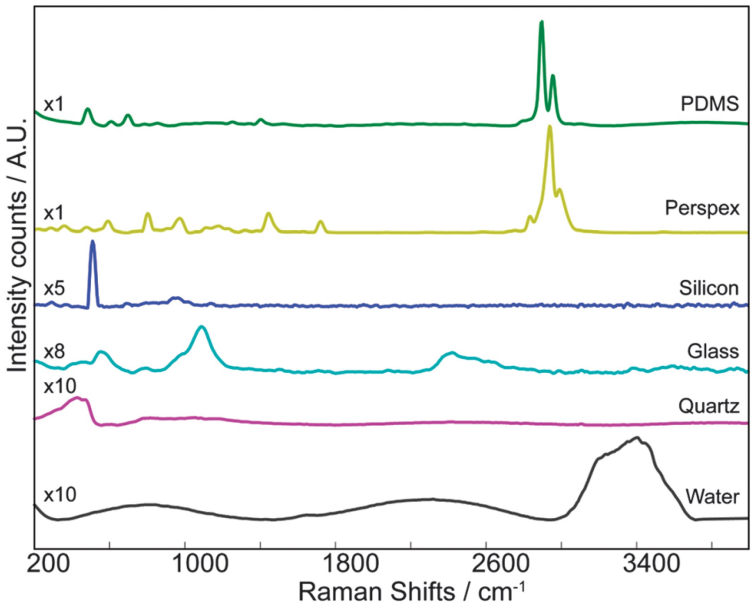


Figure 1.6: Normalised backgrounds for PDMS, perspex, silicon, glass, quartz and water, adapted from Chrimes *et al.* 2013 [20] with permission from the Royal Society of Chemistry.

For microfluidic fabrication, silicon and glass are often chosen when durable and chemical resistant materials are needed, but fabrication is costly and requires dangerous chemicals. Quartz is the ideal material for background reduction in microfluidic devices but can be costly. Alternatively, materials such as PDMS[1] and Perspex are cheap and can be fabricated rapidly, but require a rigid structure and have a strong Raman background signal. For instance, Krafft *et al.* claimed that although polymer microfluidic devices can be fabricated low cost, they are incompatible with RS due to intense *out-of-focus* polymer Raman signal, and thus chips are often made in quartz and glass [25], and Yuan *et al.* were unable to use the single bacteria spectra obtained in shallow PDMS chambers due to the strong background [178]. Figure 1.6 adapted from Chrimes *et al.* shows the main features of different materials often used in microfluidics when exciting with a 532 nm visible laser. This can be overcome by adding a Raman transparent window into the microdevice.[20] Some other issues when integrating Raman and microfluidics are reviewed in Chrimes *et al.* One common concern is the so-called *memory effect* of PDMS, where analyte adsorption and absorption can give a constant Raman background. Integration of RS with droplet microfluidics, where small fluid droplets in an immiscible phase (often oils) are used as micro-reactors, has other considerations. For instance, the droplet shape can cause a lensing effect, causing a distortion of the acquisition volume if the light does not enter through a flat surface, and the carrying oil may have contributions to the Raman signal.[20]

The first application in the literature of *on-chip* RS is an article in Chinese from 2007 from Huang *et al.* where they used LTRS on RBC[2]s in a glass-silica device.[179] Since then, more microfluidic applications of RS on single-cells have been done. An early application of *on-chip* LTRS is Lao *et al.* in 2008 with a PDMS multichannel approach, where a cell was optically trapped in one channel and manually moved to a side stream for measurement with a visible 660 nm laser and sorting. Their PDMS channel was 500 μm high, and they reduced the contributions of PDMS to the cell spectra by using a high confocality system. They obtained a good classification of two live leukaemia cell types – JM-1 and Molt-4 – but with a low

[1] Polydimethylsiloxane
[2] Red blood cell

throughput of 20 cells per cell type.[180] Another early application was Huang *et al.* that used LTRS in a glass capillary to immobilize and sort single bacteria that had utilized a ^{13}C source, that were then genetically analysed by cleaving the capillary. Bolwien *et al.* developed a bacterial contamination analysis system based on 785 nm excitation RS, where a solution was filtered in a microfluidic device onto a glass coverslip, and image analysis was used to optically trap the filtered particles and obtain their Raman spectrum.[181] Multifoci LTRS allows acquiring up to 8 cell spectra simultaneously.[182] Additionally, the Rayleigh peak can be used as a trigger as it increases when a cell is trapped due to changes in the refractive index.[182]

In another example, Dochow *et al.* did a proof-of-principle study with both a quartz capillary and a near-IR laser, and with a glass device and a visible laser using LTRS, starting with a model system of fixed patient RBCs and WBC[1]s and cancer cell lines OCI-AML3, BT-20 and MCF7.[183, 184] LDA analysis obtained 90% accuracy both using the near-IR laser and the visible laser and 10 s/cell. However, numbers of cells were between 17-99 per sample type, and the visible laser data did not include RBCs as they could not be trapped in the glass device. Manual flow control allowed sorting each cell after measurement.[183, 184] Following work of this group in the same platform increased the number of fixed cells analysed to 100 cells per cell type, obtaining an accuracy of 98% after removing outliers.[185] Additional improvement of the microfluidic system was reported in a further publication, with a quartz-glass device with integrated optical fibres, eliminating the requirement of a microscope.[186] Further work of this group looked into using micro-hole arrays of silicon nitride – a low RS signal material – on a silicon wafer and bright field microscopy to preselect cells of interest, that were then scanned with a 785 nm laser, on a model system with healthy leukocytes, breast cancer cells and leukaemia cells. They used a novel integrating system to acquire an average spectrum of the cell, superior to training by mapping.[187]

Casabella *et al.* made use of a LTRS in a 278 μm glass-PMMA device, with a continuous spectra acquisition every 1s with a 532 nm laser, obtaining a background spectrum per cell and subtracting it. They used the system to compare classification efficiency on two live cell lines: PC-3 prostate cells and Jurkat lymphocytes, in a single channel device. They compared LDA classification accuracy when using 2s and 5s integrated signal, obtaining accuracies $> 90\%$. Additionally, they used a two channel device to compare the viability of cells after 3.5h storage at 4°C by tracking the drop of cyt C[2] RR[3] signal.[82]

Knoepp *et al.* performed an integrated Raman and patch clamp study on lung muscle single cell evolution when changing from normoxic to hypoxic conditions and under exposure to ROS[4]. Their microfluidic system consisted in an airtight polycarbonate cover of a glass Petri dish – whose contributions overlapped with the fingerprint region –, and only 4-11 cells were analysed per group, with no significant conclusions, but this was the first demonstration of *on chip* Raman tracking of biological processes.[188]

Other groups have performed Raman *on-chip* using a RR model system for increased signal

[1] White blood cell
[2] Cytochrome C
[3] Resonant Raman
[4] Reactive oxygen species

to allow quicker acquisition. Zhang *et al.* made a ITO[1]-glass-PDMS microfluidic device based on positive DEP[2] and a suction valve for pause-and-sort RACS with a 532 nm laser. As a *proof-of-concept* they sorted *Saccharomyces cerevisiae* based on carotenoids contents in an automated manner, where the cells were measured in deionized water, with the lowest acquisition time of 90 ms. The strong scattering of carotenoids allowed them to overcome the PDMS signal.[189] The same research group used a similar system for screening and sorting of an astaxanthin containing *Haematococcus pluvialis* strains in water-in-oil droplets, lowering the acquisition time to 30 ms [190, 191] when using a 100 mW 532 nm laser. They sorted 260 cells per min, and showed comparable results on CaF2 and *on chip* [191]. They also measured the well size dependence of carotenoids accumulation of microalgae [192]. Ramser *et al.*, Eriksson *et al.* and Gollner *et al.* used PDMS microfluidics to measure RBCs oxygenation by using photodissociation and the RR signal of haemoglobin, both using LTRS [193–195] or microfluidic trapping [196], and obtaining additional imaging information such as RBC membrane flickering [196]. Ramser *et al.* studied the RR of different heme-containing proteins in single RBCs or *E. coli* cells over-expressing neurogenin *on chip* in a subsequent publication.[197] Verma *et al.* used an optical trapping system to focus RBCs into a line flowing towards the 514 nm measuring spot, obtaining about 1300 high-quality RR spectra at a flow rate of 500 cells per hour.[198] Li *et al.* and McIlvenna *et al.* also made use of RR enhancement when *in-flow* measuring the carotenoids contents of cyanobacteria incubated with $^{12}CO_2$ compared to $^{13}CO_2$.[199, 200] RS has also been integrated with droplet microfluidics by Kim *et al.*, where they used Fluorinert PC-40 – a fluorocarbon-based oil – as the carrier oil, with 250 μm diameter droplets with 10-15 microalgae cells in each, measuring the RR of single cell carotenoids accumulation with a 532 nm laser. They report a high reduction of the PDMS peaks when increasing confocality, and no interference of the oil signal.[201]

An alternative approach is to either accept the presence of the polymer peaks and focus on the areas not affected by them or use chemometrics to subtract the polymer contribution to the cell spectrum. For example, Feng *et al.* did Raman on *Pseudomonas aeruginosa* forming biofilms *on-chip* with a PDMS device, analysing only the peaks that do not overlap with the PDMS background.[202] Another example is Liszka *et al.*, that did Raman microspectroscopy *on-chip* on PDMS-CaF_2 devices on adhered NCI-H1650 cells, and minimized the PDMS contribution by using high confocality and background removal from the cell spectra using chemometrics.[203]

Other groups, in order to increase the signal intensity of the band of interest for *on chip* applications, used non-linear RS techniques, such as SRS[3] or CARS[4]. For example, Cao *et al.* use SRS on a PDMS chip for high throughput single cell lipid droplet imaging after FA[5] stimuli. Their chip could culture different cell lines for up to 3 days and prepare different dilutions of FA from a stock in a controlled manner before adding them to the cells, and probed

[1] Indium tin oxide
[2] Dielectrophoresis
[3] Stimulated Raman scattering
[4] Coherent anti-Stokes Raman spectroscopy
[5] Fatty acid

the CH_2 and CH_3 vibrations at 2850 and 2950 cm^{-1} respectively, performing background removal on the images.[204]

More recently, the work from Goda's group using dual-comb CARS obtained up to 1200 polystyrene particles single-spectrum per second in an acoustofluidic integrated microfluidic device, where they also recorded ultrafast imaging of the passing particles.[205] They further improved this system with the fastest current Raman based cell cytometer, with a proof of concept study of the astaxanthin production of *Haematococcus lacustris* of fingerprint region spectrum acquired at a rate of 2000 events per second, limited only by the microfluidic design.[206]

In the present thesis, label-free RS *on chip* on glass-PDMS microfluidic devices will be used. Different method will be tested to remove or reduce the PDMS contribution, including background subtraction. Cell trapping in PDMS microfluidic devices will be used to track the accumulation in the same live cell of the photosensitiser DC473. To our knowledge, Raman microfluidic studies of drugs accumulation in the same live single cell have not been previously done. Here, we used single-cell trapping to immobilize cells and be able to record their spectra before and following incubation with DC473, increasing our understanding of the biomarkers that predict the accumulation of the compound.

1.8 Thesis Overview

This PhD optimized the Raman acquisition on live single-cells both for classification of CRC[1] disease stages, for detection of drug accumulation and activation in cells and for optimization of the measuring parameters on microfluidic devices. The next chapters of this thesis constitute a theoretical background chapter (**Chapter 2**, page 24) and the methodology chapter (**Chapter 3**, page 49), including the first experiments in the optimization of the system acquisition parameters and the experimental methods used in the following results chapters. This thesis reports the results on RS[2] on live single CRC cells, divided into three results chapters:

Chapter 4 (page 91) is the first results chapter and introduces the first single cell RS results obtained from different cell lines: HL60, HCT116, SW480, HT29 and SW620. This represents the first approach to multivariate methods analysis (introduced in Chapter 2, pages 32 onwards) and choice of the best multivariate methods to classify the single cell data from different cell lines. It also explores the main cell peaks assignments, performing correlation analysis, and identifies the peaks with the most relevant contributions to within-population variability. It obtains potential RS biomarkers for metastasis based on the adenocarcinoma progression model system SW480/HT29/SW620.

[1] Colorectal Cancer
[2] Raman spectroscopy

Chapter 5 (page 111) focuses on PS[1] detection on single cells using RS or RR[2]. First part works on cells following incubation with 5-ALA[3] for the accumulation of PPIX[4], identifying the best peaks for PPIX detection and its overlap with the RR signal of cyt C[5]. The second part focuses on the novel photosensitiser DC473, studying its location within fixed single SW480 cells. Furthermore, the accumulation of DC473 in live single cells is detected using RS, obtaining the ratio between DC473 accumulation relative to Amide I band intensity. Finally, the activation of DC473 in live cells using UV light irradiation is tracked with RS, observing the molecule partial reaction and the DC473 bands' intensity reduction, with additional effects showing band shifting and broadening in the following minutes.

Chapter 6 (page 161) reports the optimization of the Raman *on chip* acquisition in PDMS[6] microfluidic devices, showing the effects of metal evaporations for background reduction and of confocality. It subsequently reports first Raman on *on chip* results, measuring before and following incubation with DC473 on the same SW480 single cells, and obtaining Raman biomarkers to predict the accumulation of DC473 from the untreated cells. It also reports our first efforts on measuring bacteria single cells *on chip*.

Finally, **Chapter 7** (page 192) summarizes the conclusions with a chapter by chapter overview, comparing the Raman findings to the findings obtained in the same model system using a mechanical phenotyping technique: Deformation Cytometry. It also discusses the next steps and future work.

[1] Photosensitiser
[2] Resonant Raman
[3] 5-aminolevulinic acid
[4] Protoporphyrin IX
[5] Cytochrome C
[6] Polydimethylsiloxane

Chapter 2

Background Theory

In the present thesis, Raman spectroscopic and chemometric analysis were used for single cell analysis and for detection of both photodynamic drugs and their effects, with the aim of integrating this into a microfluidic setup. The previous chapter constituted a general introduction to RS[1] for single-cell analysis, including a literature review of the use of RS for drug detection in single cells and tissue, of the detection of photodynamic agents and of previous reports on Raman *on chip*.

In this chapter, the background theory behind RS and the multivariate methods used in this report and the main principles and equations related to PDT[2] and microfluidics will be described.

2.1 Theory of Raman spectroscopy

When photons from an incident light source interact with an atom or a molecule, most of them retain their energy in an elastic scattering also known as Rayleigh scattering. However, approximately one in a million photons changes its energy in a process known as Raman scattering. [24–27] In RS[3], the photon may either lose (Stokes) or gain (anti-Stokes) energy, causing its frequency to decrease or increase - consequently leading to a red or blue shift of the scattered photons. On the other hand, the molecule or group of atoms the photon interacted with gains or loses vibrational energy, making the probed vibration increase or decrease amplitude. [28] The Raman effect was first theoretically predicted by Smekal (1923) followed by quantum mechanical descriptions by Kramers and Heisenberg (1925) and Dirac (1927), and first observed by Sir C.V. Raman and Krishnan in 1928 [14], after whom it was named. Raman's experiment was deceptively simple: using sunlight, a violet filter and a green filter he showed that violet light scattering on different liquid samples all showed a green component of light [26, 207]

The Raman effect can be explained by two different approaches: classical theory or quantum

[1] Raman spectroscopy
[2] Photodynamic Therapy
[3] Raman spectroscopy

theory, each bearing their own merits in aiding our understanding of the processes involved. Here, both of them will be briefly explained.

2.1.1 Classical Theory

The classical theory describes the molecule as a group of bonded atoms that vibrate at a characteristic frequency, performing anharmonic oscillations. The electron clouds around the atoms may shift with these vibrations. If a molecule is in the presence of an electric field - such as the one from the incident light - it displaces the electron cloud inducing a dipole moment μ, that has the same direction as the electric field and is proportional to it for small electric fields E following equation 2.1, where α is defined as the polarisability. [15, 26, 28, 49, 208]

$$\mu = \alpha \cdot E \tag{2.1}$$

The polarisability measures how easily the electron cloud associated with a molecular bond may be distorted, and will depend on the orientation of the molecule respect to the electric field, making α a tensor with components α_{ij} as shown in equation 2.2. However, for this simplified description, the geometrical aspects of the interaction will be ignored.

$$\alpha = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} \tag{2.2}$$

Light is an electromagnetic wave that interacts with matter predominantly via its electric field component, $\vec{E}$, that at a certain position can be represented in equation 2.3, when the phase is zero and the frequency is ν_0.

$$E = E_0 cos(2\pi\nu_0 t) \tag{2.3}$$

If the polarisability of a bond or a group of atoms changes as the molecule vibrates, this change can be described as a Taylor polynomial series respect to equilibrium, where for small vibrations it can be approximated by only keeping the first order change of α with the displacement of the bond from equilibrium q_v along the vibration axis, as shown in equation 2.4. Note that the partial derivative over the vibration coordinate is taken at the equilibrium position. This is illustrated in figure 2.1 for the CO_2 molecule.

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial q_v}\right)_0 \cdot q_v \tag{2.4}$$

Even though the vibration is anharmonic, for simplicity it can be approximated as a harmonic vibration over time with a frequency ν_v, as shown in equation 2.5, where q_0 is the vibration amplitude.

$$q_v = q_0 cos(2\pi\nu_v t) \tag{2.5}$$

By substituting equations 2.4, 2.3 and 2.5 into equation 2.1 and expression for the change of the dipole moment μ with the vibration as a function of time is obtained:

$$\mu(t) = \alpha_0 E_0 cos(2\pi\nu_0 t) + \left(\frac{\partial\alpha}{\partial q_v}\right)_0 \cdot E_0 q_0 \ cos(2\pi\nu_0 t) cos(2\pi\nu_v t) \tag{2.6}$$

By using the trigonometric identity $2cos(A)cos(B) = cos(A+B) + cos(A-B)$, equation 2.7 for the dipole moment is obtained: [15, 28, 49].

$$\mu = \underbrace{\alpha_0 E_0 cos(2\pi\nu_0 t)}_{(1)} + \left(\frac{\partial\alpha}{\partial q_v}\right)_0 \cdot \frac{q_0 E_0}{2} [\underbrace{cos((\nu_0 - \nu_v)2\pi t)}_{(2)} + \underbrace{cos((\nu_0 + \nu_v)2\pi t)}_{(3)}] \tag{2.7}$$

from where it becomes clear that the interaction of light with an oscillating bond leads to three different components of the dipole moment [15, 26, 207]. According to the laws of classical electrodynamics, and oscillating electric dipole emits radiation of its oscillating frequency. Therefore, from equation 2.7, the interaction of light with this oscillating dipole will result in three different emitted frequencies corresponding to (1) Rayleigh scattering, of the same frequency as the incident wave ν_0; (2) Stokes scattering, with a frequency smaller than the incident wave $\nu_0 - \nu_v$, and (3) anti-Stokes scattering, with a frequency greater than the incident wave $\nu_0 + \nu_v$ [15, 24, 26–28, 49, 207]. From equation 2.7 the main condition for Raman scattering can also be obtained: the variation of the polarisability with the displacement in the equilibrium position with respect to the vibration axis must be different to zero:

$$\left(\frac{\partial\alpha}{\partial q_v}\right)_0 \neq 0 \tag{2.8}$$

Since the change of polarisability over the vibration is usually very much less than the polarisability itself, Raman scattering is a very weak signal compared to Rayleigh scattering.

A schematic for the possible vibrations of the CO_2 molecule and the change of polarisability is shown in figure 2.1. The change of polarisability with the vibration coordinate is plotted for each vibration. Only those vibrations where the condition given in equation 2.8 holds true are "Raman active".

To predict the frequency of a particular vibration, a molecule can be modelled by using balls and springs. For two atoms bonded by a spring, the frequency of harmonic oscillations obeys equation 2.9, where k is the spring force constant and μ_m is the reduced mass, with m_1 and m_2 the masses of the atoms. This shows that each vibration has a unique frequency depending on both the bond and the mass of the atom. Heavier atoms will have a higher reduced mass and thus lower vibrational frequencies, and stronger bonds such as double or triple bonds will have a higher spring constant k with higher frequencies. [38]

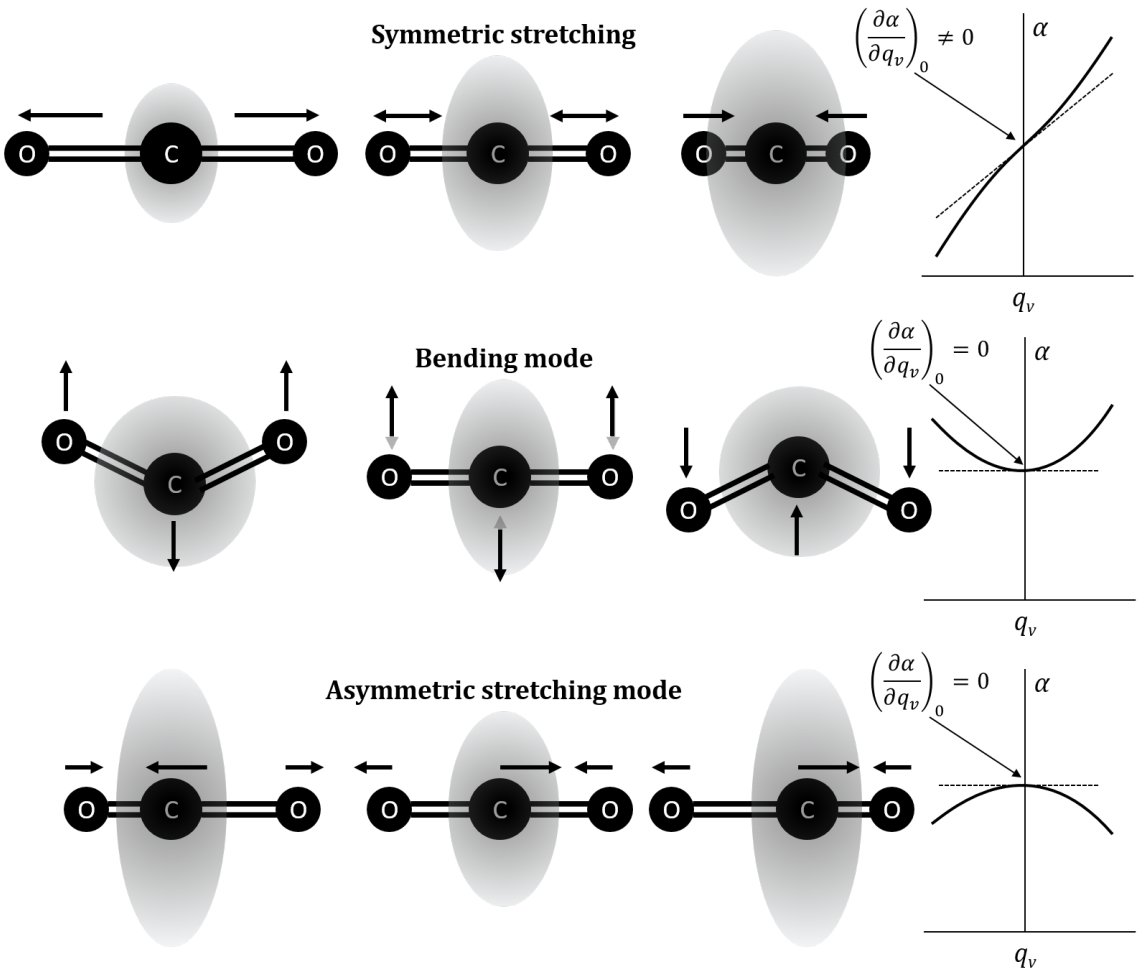


Figure 2.1: Schematic of the possible vibrations for the CO_2 molecule, where the polarizability is shown as a sphere and the changes of the polarizability with the vibration coordinate are plotted. Only the symmetric stretching holds the Raman condition of equation 2.8 and will be Raman active.

$$\nu_v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu_m}} \qquad \mu_m = \frac{m_1 m_2}{m_1 + m_2} \tag{2.9}$$

Some Raman active bonds – whose change in polarisability with vibration is strong – are C=C, C–S, S–S or aromatics [33]. Heavier atoms such as C–C have in general shorter shifts than lighter atoms such as C–H. Polarized bonds such as N–O or O–H have a weaker Raman signal than neutral bonds such as C–C or C–H, as these ones will have a greater change in α [209]. Raman spectroscopy is a powerful technique for studying organic compounds enriched with carbon and hydrogen atoms, making it ideal for biochemical applications [49].

The classical description helps us understand some of the basic characteristics of the Raman effects. However, it is important to bear in mind that a number of approximations were taken. For instance, molecular bonds display anharmonicity which will lead to non-equally spaced overtones (slightly lower than multiples of the main frequency of the vibration), generally much weaker than the fundamentals. We also considered only the first order of the Taylor polynomial for the changes in polarisability with vibration, and considering more terms leads to non-linear Raman scattering [15, 210] such as Stimulated Raman Scattering, Hyper Raman,

Stimulated Raman Gain, Inverse Raman, Coherent anti-Stokes Raman or Coherent Stokes Raman scattering [210].

2.1.2 Quantum Theory

Quantum theory assumes that the energy of a molecule is quantised, and thus a molecule can only be in certain energy levels. There are electric, vibrational and rotational components of the energy of a molecule: $E = E_e + E_v + E_r$. Energy levels are thus electric, vibrational and rotational. The spacing between rotational levels is much smaller than that of the vibrational levels, that is much smaller than that of electronic levels. Vibrational energy levels can be approximated to follow a harmonic relation as stated in equation 2.10.

$$E_v = h\nu_v\left(v + \frac{1}{2}\right) \qquad v = 0, 1, 2, 3, \ldots \tag{2.10}$$

where ν is the frequency of vibration, v is the vibrational number and h is Planck's constant. In Raman scattering, for the incident photon to interact with a molecule, the molecule needs to be raised to a higher energy state known as a "virtual" energy state. This state is classically forbidden. However, in order to explain scattering processes using this scientific model, it was argued that the transition can occur with a very low probability as long as it does not violate Heisenberg's uncertainty principle (equation 2.11), meaning that the time for a Raman transition is usually less than 1 ps. In reality, this "virtual" energy state does not exist.

$$\Delta t \Delta E \geq \hbar \tag{2.11}$$

Figure 2.2 shows a comparative Jablonski diagram of IR[1]-absorption, Raman scattering and fluorescence emission events. From a Quantum Mechanics point of view, scattering events always leave the material in the same electronic state where it started, usually the ground state, but sometimes in different vibrational energy levels [26, 27]. When the molecule interacts with the photon, it gets excited to a virtual energy state, and then returns to a different vibrational state, therefore with a shift of energy. That is, the two photons involved in the scattering are connected by a single quantum mechanical process, a coherent event [208].

At room temperature, most molecular vibrations are in the ground level – populations of each level follow a Boltzmann distribution as shown in equation 2.12, where N is the total number of molecules, n_j is the population of a certain level, ν_j is the frequency of that level and k is Boltzmann's constant. For this reason, anti-Stokes transitions are less likely to occur than the Stokes, giving a weaker anti-Stokes signal. This is why Stokes transitions are usually widely studied experimentally [26, 27].

$$\frac{n_j}{N} = \frac{e^{-h\nu_j/kT}}{\sum_i e^{-h\nu_i/kT}} \qquad N = \sum_i n_i \tag{2.12}$$

[1] Infrared

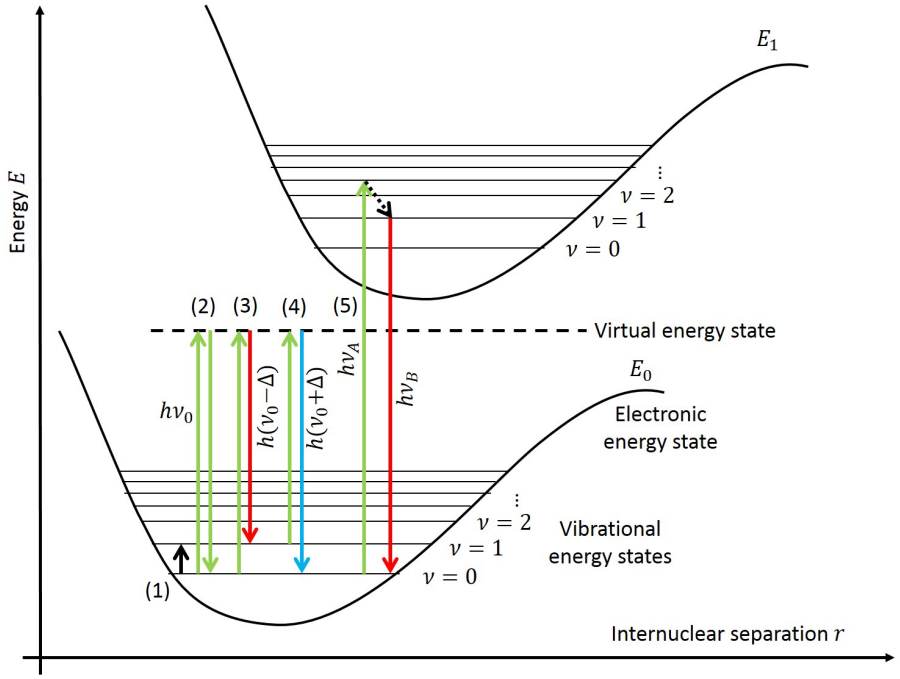


Figure 2.2: Jablonski Energy diagram showing the difference between (1) IR absorption, (2) Rayleigh scattering, (3) Stokes and (4) anti-Stokes Raman scattering, and (5) Fluorescence. The energy states E_0 and E_1 show different vibrational states for different values of v. A virtual state has been indicated with a discontinuous line. Colors were used to indicate the shift (either to red or to blue) of the light (represented as green).

Raman scattering, IR absorption and fluorescence are intrinsically different. Scattering processes undergo a transition into a virtual state whereas IR absorption is a transition between vibrational states. In fluorescence, the incident light is completely absorbed and the molecule transients into an excited electronic state, followed by a non-radiative transition to a different energy state and emission of a photon, making the process much slower [26, 27, 33].

For a mode to be Raman active, the polarisability needs to change with the vibration, but for it to be IR active it is the dipole moment that needs to change with the vibration. RS and IR spectroscopies are often seen as complementary [26]. A practical review of the differences in spectra from IR and RS can be found in the Course Notes from Mayo *et al.*[38].

2.1.3 Intensity of Raman scattering and Raman cross-section

The intensity of the Raman spectral bands is proportional to the fourth power of the frequency of the band, and is given by equation 2.13, where I is the intensity of the scattered light, with frequency $(\nu_0+\Delta\nu)$, I_0 is the incident intensity with frequency ν_0, α is the polarisability, q_v is the vibrational coordinate and K is a proportionality constant [211]:

$$I = K I_0 (\nu_0 - \Delta\nu)^4 \left(\frac{d\alpha}{dq_v}\right) \tag{2.13}$$

This means that choosing a higher wavelength laser, such as a near-IR laser instead of a

visible one, will incur in lower intensity Raman signal with a dependence of intensity with wavelength of λ^{-4}.[211]

For quantitative considerations, light absorption follows the Bouguer-Beer-Lambert law – often refer to as the Beer-Lambert law – where the intensity after the sample is proportional to the incident intensity as stated in equation 2.14, where I is the intensity, a is the absorptivity at a certain frequency ν and l is the sample thickness [211]:

$$I(\nu) = I_0(\nu)e^{-a(\nu)l} \tag{2.14}$$

For Raman scattering, the ratio of the total scattered power P and the power of the incident light per unit area (Irradiance, J) is defined as the scattering cross-section σ, with area dimensions.[211] Molecules with higher cross-sections will scatter more light at the same concentration than molecules with lower cross-sections.

$$P = \sigma J \tag{2.15}$$

Similar to the Beer-Lambert law, Raman scattering intensity shows linearity with concentration. However, light losses due to reflections in the sample interfaces and differences in the detector's sensitivity to different frequencies require calibration for each Raman instrument and sample setup before any quantitative analysis can be done. Nevertheless, Raman can be used semi-quantitatively to indicate relative changes in chemical composition.[211]

2.1.4 Bandwidths

In vibrational spectroscopy, attention is traditionally concentrated on the position and intensity of spectral bands, but lots of insight can be obtained from their shapes, that contain information about the dynamics of the system [211]. According to quantum theory, each vibration would have a well-defined frequency, and thus a spectrum would have infinitely sharp bands, only broadened by our instrument.[211] However, the vibrating molecule (frequency ω) will have energy losses described by a damping constant γ, with often $\gamma \ll \omega$ so that the amplitude of the oscillation $x(t)$ is exponentially damped over time. The amplitude in the frequency space corresponds to a Lorentzian profile, where the decay time is $\approx 1/\gamma$. This is known as *Natural line broadening* [212]

However, molecules exist in a bath of surrounding molecules with which they interact, causing perturbations to this frequency.[211] An example of these perturbations is the so-called *Doppler width*, caused by thermal motion in gases at low pressure. At thermal equilibrium, the molecules of a gas follow a Maxwellian velocity distribution, with density states following a Gaussian. As the emitting or absorbing power is proportional to the density of molecules, the bands will follow a profile between Gaussian and Lorentzian, but closer to a Gaussian distribution: a Voigt profile, convolution of these two – see equation 2.18.[212] Another relevant effect is collisional broadening, that can cause pressure broadening and shifts in gasses, or *Dicke narrowing*, where collisions cause narrowing of linewidths in the IR.[212]

In liquids and solids, molecules are closer together. In liquids, the distance between molecules shows random fluctuations, and inelastic collisions may cause radiationless transitions that shorten the lifetime, causing a collisional line broadening. This often results in broadening larger than the separation of spectral lines, obtaining a broad continuum. In solids, very different effects are seen depending on whether if the solid is in a crystalline lattice, following a Lorentzian profile, or amorphous, with a situation analogous to Doppler broadening in gases, with linewidths broadened by orders of magnitude. This is a simplified view on bandwidths, and for more on this topic, the book *Laser Spectroscopy* from Demtroder (2003) is recommended.[212]

When the band responds to an intermediate shape between a Gaussian and a Lorentzian peak, three options are regularly considered for fitting: the Voigt function, which is the convolution of a Gaussian with a Lorentzian function (see equation 2.18); the GLS[1] and the GLP[2].[213, 214] Following a similar notation to the one used in Jain et al. 2014, a Gaussian and a Lorentzian equation are defined in equations (2.16) and (2.17). The GLS function is defined in equation (2.19) as the sum of both using a weighting parameter m between 0 and 1. The GLP function is defined in equation (2.20) as the product of a Gaussian and Lorentzian functions, where the mixing parameter m also weights the contributions of each, with $0 < m < 1$.

$$G(x, w, x_0, A) = A \cdot e^{-4ln(2)\frac{(x-x_0)^2}{w^2}} \tag{2.16}$$

$$L(x, w, x_0, A) = \frac{A}{1 + 4 \cdot \frac{(x-x_0)^2}{w^2}} \tag{2.17}$$

$$V(x, w, x_0, A) = \int_{-\infty}^{\infty} G(x', w, x_0, A) L(x - x', w, x_0, A) dx' \tag{2.18}$$

$$GLS(x, w, x_0, A, m) = (1 - m) \cdot G(x, w, x_0, A) + m \cdot L(x, w, x_0, A) \tag{2.19}$$

$$GLP(x, w, x_0, A, m) = h \cdot A \cdot e^{-4ln(2)(1-m)\frac{(x-x_0)^2}{w^2}} \cdot \frac{A}{1 + 4m \cdot \frac{(x-x_0)^2}{w^2}} \tag{2.20}$$

The GLS function shows the closest shape to the Voigt profile, as the GLP function loses the Lorentzian *wings* when they are multiplied by the close to zero Gaussian *wings*.[213, 214] For simplicity and seeing that the GLS is easier to implement than the Voigt function, with similar results, in this thesis Raman peaks in liquid samples are fitted to either Gaussian or to a GLS function with mixing factor m =0.5.

[1] Gaussian-Lorentzian sum
[2] Gaussian-Lorentzian product

2.1.5 Resonant Raman

RR[1] scattering occurs when the energy of the incident photon is close to the energy of an electronic transition due to a coupling of electronic and vibrational transitions [25, 49, 211] when the virtual state lies close to an electronic excited state [27]. This leads to a resonance that can lead to an increased probability of the transition, with an enhancement of up to 10^6 [28, 211]. Most vibrational modes strong in the RR spectrum have a geometrical change associated with the electronic transition that leads to a change in the polarizability of the molecule. The bonds more affected by the tuned electronic transition experience a stronger change in bond length or force constant than other bonds in the molecule, that, when leading to an increased change in polarizability with vibration, get an enhanced Raman intensity. This means that sometimes RR can reveal features not observed with nonresonant excitations. When the excitation source is resonant, symmetric vibrations mimicking the geometrical changes that accompany electronic excitation dominate the Raman spectra. In some cases, non-totally symmetric vibrations are also enhanced. Fluorescence emission must be considered when doing RR as the electronic excitation of the sample could excite intrinsic fluorescence of the chromophore [211].

Higher RR intensity will be obtained when matching the first or second electronic excited state. This resonance will be stronger when the energy of the electronic level is perfectly matched, but great enhancements can also be obtained when the match is not exact. This can be achieved by using a tunable laser or it can "accidentally" occur when using certain visible lasers in biological samples. The enhanced detection of RR can allow detection of compounds at concentrations as low as 10 nM, compared to usual Raman that generally requires concentrations around 100 mM, presenting an increased sensitivity and selectivity [211].

RR depends on the wavelength, and is common for molecules with chromophores in their structure, such as proteins with the heme group: haemoglobin, myoglobin, cytochromes or porphyrins among others. For example, the use of the 532 nm laser excites RR of cyt C[2], allowing detecting mitochondria-related processes. This can have applications in reproductive medicine [24] or in early apoptosis detection [50].

2.2 Theory of Chemometric Analysis

Chemometrics is the use of mathematics and statistics to understand chemical information, extracting information from chemical systems by data-driven means. Spectroscopy methods such as Raman obtain large amounts of data in the form of spectra with hundreds of different observables (each wavenumber). Statistical methods used to deal with a big number of variables are known as multivariate methods and can be "supervised" or "unsupervised". When there is only a discrete number of possible outcomes (categories, ie. a diagnosis) the process becomes one of "classification", whereas a continuous output requires the use

[1] Resonant Raman
[2] Cytochrome C

of "regression" techniques. For classification, the goal is always to reduce the number of misclassifications and increase noise-robustness. [215]

In supervised methods, a group of observations with known expected output are used to train the system. Starting from this information a classifier is trained so as to maximise the separation between different outputs. However, the system needs to be tested with samples that have not been used in the training to avoid overfitting, which causes "overlearning". An overfitted model can no longer generalise effectively, with a high predictive error for new inputs outside from the training group. Unsupervised learning does not rely on previous training, and it is useful to see if a group of observations shows any initial clustering according to their intrinsic variability. [215] However, there might be sources of variability that are not of interest such as background contribution or measuring artefacts that unsupervised methods will consider equally valid for classification.

Models can be linear or non-linear. Linear models are in general easier to relate back to the data to understand the underlying biology, but they may perform worse when classifying non-linear data. Increased complexity can increase performance but complicate the understanding of the model. The general equation of a linear model is shown in equation 2.21, where the output y is calculated from an initial coordinate to centre the data y_0 and a linear combination of the inputs $\{x_i\}$ with coefficients $\{m_i\}$.

$$y = y_0 + \sum_{i}^{p} m_i x_i \tag{2.21}$$

An example of a widely used linear unsupervised method is PCA[1]. Examples of supervised models are LDA[2] or CA[3]. Other data mining techniques, such as SVM[4], genetic algorithms, DT[5]s or ANN[6] can be very powerful for class separation but are more difficult to relate to the underlying biology. [216] In the present thesis, a simple approach on these models will be explained for chemometric applications for classification and analysis of both single cell and Raman mapping data that will be used in the Results chapters (Part III, page 91).

2.2.1 Unsupervised Multivariate Methods

In this subsection, PCA and PLSR[7] will be described as examples of unsupervised multivariate methods used in this thesis.

2.2.1.1 Theory of Principal Component Analysis

When trying to classify a group of observations, relevant questions could be: which bands are increased in this class but not in the others? which bands vary together - ie. are correlated?

[1] Principal Component Analysis
[2] Linear Discriminant Analysis
[3] Cluster analysis
[4] Support Vector Machines
[5] Discrimination Trees
[6] Artificial neural networks
[7] Partial Least Squares Regression

If the number of variables p is very large, it can be overwhelming to try to look at individual correlations and covariances. PCA was firstly described by Pearson 1901 [217] and Hotelling in 1933 [218] and described in depth in Wold 1987 [219], and it looks for a few derived variables which preserve most of the information of the system

For p variables (each wavenumber) $\{x_i\}$ there are N observations (each cell), so the data matrix X is a $N \times p$ matrix. The covariance definition is shown in equation 2.22, where x_1 and x_2 would be two different wavenumbers, and $(x_1 1, ..., x_1 N)$ and $(x_2 1, ..., x_2 N)$ would be their observations for each single cell. Note than when $x_1 = x_2$, the variance is obtained. [220]

$$cov(x_1, x_2) = \frac{1}{N-1} \sum_{i=1}^{N} (x_{1i} - \bar{x_1})(x_{2i} - \bar{x_2}) \qquad cov(x_1, x_1) = var(x_1) \tag{2.22}$$

For the covariance matrix when there are more than two variables, each element ij is the covariance of the variable i and j, and the diagonal elements are the variances, with the trace of the matrix – the sum of diagonal elements – being the total variance of the system. The covariance matrix is thus a square matrix of size $p \times p$. [220]

PCA searches for a linear transformation β (of dimensions $p \times p$) that maximizes the covariance, that is, that maximizes the variability of the system. [220] β will contain all the new variables of the system $(\beta_1, ..., \beta_p)$ (coefficients, PC[1]s), each a combination of the previous variables x. The aim is to obtain a first coefficient with maximum covariance, with the following coefficients both with maximum covariance and perpendicular to each other. By following this method, the first few PCs contain most of the variability of the system, and the last ones have little to none relevant information. [220]

To obtain β_1 (PC1), $\beta_1^T x$ needs to be calculated, then maximise its covariance imposing that β_1 is normalised, where x contains our initial variables. Based on equation 2.22 it can be shown that $cov(\beta_1^T x) = \beta_1^T cov(x) \beta_1$ – where T denotes the transpose. The normalisation constraint takes the form of $\beta_1^T \beta_1 = 1$. Finding the maximum can be done by using the Lagrange multipliers technique – a strategy to find local maxima that satisfy equality conditions – where λ_1 is a Lagrange multiplier:

$$max(\beta_1^T cov(x) \beta_1 - \lambda_1(\beta_1^T \beta_1 - 1)) \xrightarrow{\frac{\partial}{\partial \beta_1}(...)=0} cov(x)\beta_1 = \lambda_1 \beta_1 \tag{2.23}$$

This final expression means that the Lagrange multiplier is just the eigenvalue when diagonalizing the covariance matrix, where β_1 is one of the eigenvectors. The eigenvector corresponding to β_1 will have to maximise the first part of equation 2.23, and thus must be the one with the highest eigenvalue. [220]

To obtain β_2 (PC2) the derivation is similar, now with both a normalisation condition $\beta_2^T \beta_2 = 1$ and a perpendicularity condition $\beta_1^T \beta_2 = 0$. This means maximising the following equation, with λ_2 and ϕ Lagrange multipliers:

[1] Principal coefficient

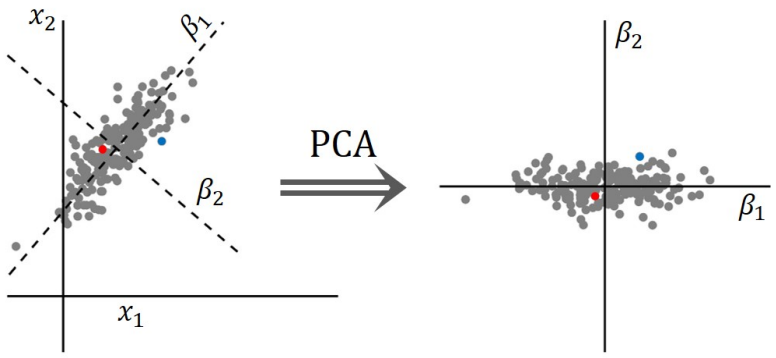


Figure 2.3: 2D example scatter plots to illustrate the effects of PCA on a linear combination of a random "normal population" overlapped with "normal noise". PCA identifies the axis of maximum variability β_1 (PC1) and its perpendicular β_2 (PC2). The data is then projected onto them, obtaining the scores. Note that the original data was not centred, but the projected data is centred to the population mean, and that the sign of each coefficient was chosen to keep the maximum score positive. Two points are shown in different colours (red and blue) to ease the visualization of the linear transformation.

$$max((\beta_2^T cov(x)\beta_2 - \lambda_2(\beta_2^T\beta_2 - 1) - \phi\beta_2^T\beta_1) \xrightarrow{\frac{\partial}{\partial\beta_2}(\ldots)=0} cov(x)\beta_2 - \frac{\phi}{2}\beta_1 = \lambda_2\beta_2 \quad (2.24)$$

By multiplying by β_1^T (equation 2.25), as $\beta_1^T cov(x)\beta_2 = \lambda_1\beta_1^T\beta_2 0$ and $\lambda_2\beta_1^T\beta_2 = 0$, it is concluded that $\phi = 0$, arriving to a similar condition of β_2 as we obtained with β_1, where its eigenvalue will be the second biggest.

$$\beta_1^T cov(x)\beta_2 - \frac{\phi}{2}\beta_1^T\beta_1 = \lambda_2\beta_1^T\beta_2 \rightarrow \phi = 0 \rightarrow cov(x)\beta_2 = \lambda_2\beta_2 \quad (2.25)$$

This derivation can be done for the remaining coefficients obtaining a general expression for PCA stated in equation 2.26, proving that by simply diagonalising the covariance matrix and ordering the obtained eigenvector by eigenvalue size, the PCs are obtained. [220]

$$cov(x)\beta_k = \lambda_k\beta_k \quad (2.26)$$

Placing the PCs obtained in a matrix $A = \{\beta_1, \beta_2, ..., \beta_p\}$ yields the linear transformation to the PC space, so $A^T cov(x)A = \Lambda$, where Λ is a diagonal matrix whose elements are the ordered eigenvalues. The data matrix X is projected onto the PC space as XA, obtaining the scores of each single cell in each PC. PCA allows keeping only a number of components, which allows reducing the dimensionality of high-dimensional data [35, 46, 57, 72]. An example of a simple application of PCA for 2D data ($p = 2$) is shown in figure 2.3.

For the present experimental problem, the covariance matrix is unknown, and the covariance matrix of our variables can be approximated with the covariance of our sample when taking enough observations ($N \gg$). By diagonalising the sample covariance matrix and ordering the eigenvectors by eigenvalue from largest to lowest, A is obtained and can be used to project the observations onto the PC space. [220]

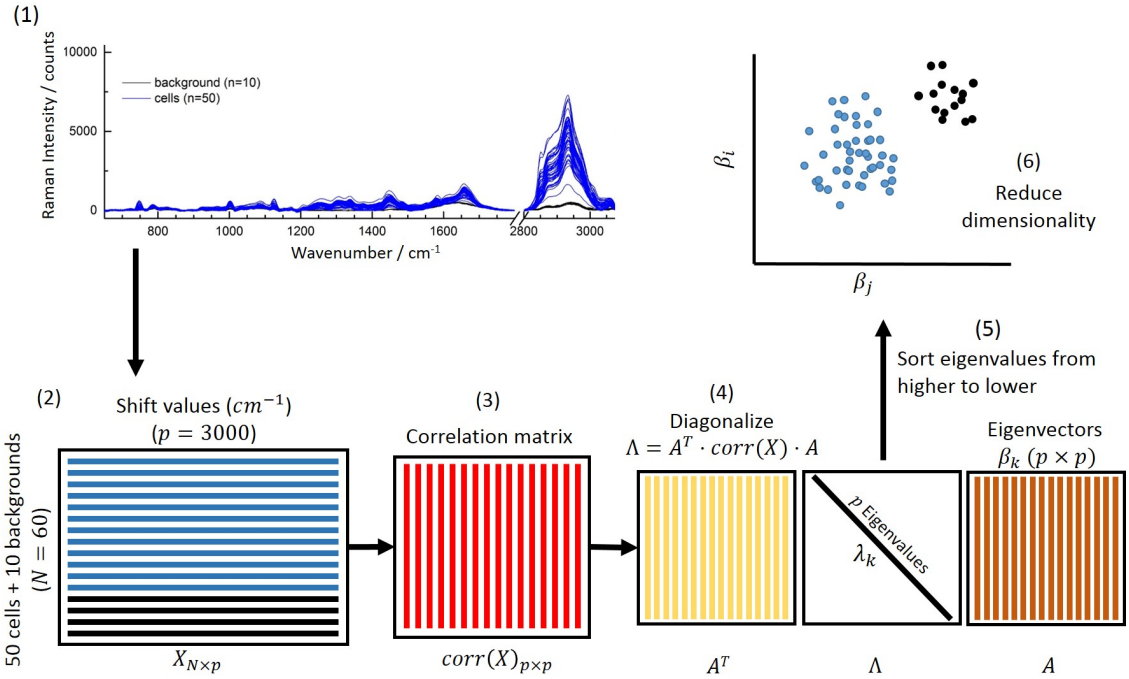


Figure 2.4: **(1)** Example of the sort of data obtained from an experiment. **(2)** This data can be ordered in a matrix, where 3000 shift-intensity pairs constitute each spectrum. **(3)** The data can be standardized to an average of 0 and a variance of 1 and the covariance matrix can be calculated. **(4)** The covariance matrix is diagonalized obtaining 3000 eigenvectors with their eigenvalues. **(5)** Sorting the eigenvalues from higher to lower allows obtaining the eigenvectors that explain the higher variance of the system. **(6)** A subspace of these eigenvectors that explains most of the variance of the system can be chosen, reducing the dimensionality. A typical plot of the data is a 2D plot of the scores of each spectrum from two different eigenvectors, usually 1 and 2.

In fact, it is more common to use standardised variables x^*, which can be done by doing PCA on the correlation matrix instead of the covariance matrix. The correlation matrix is defined in a similar fashion to the covariance matrix, but each ij element is the covariance of variables i and j normalised by the square root of their variances. For $i = j$, the correlation is 1. The equation for each of the elements of the correlation matrix is shown in equation 2.27.

$$corr(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{var(x_1)}\sqrt{var(x_2)}} \tag{2.27}$$

The covariance matrix had the variances as the trace, whereas the correlation matrix has a trace of p. Doing PCA in the covariance matrix and of the correlation matrix gives usually different coefficients, giving different information. PCA based on covariance matrices can have a strong weight if some variables are stronger than other - e.g. some Raman peaks have stronger intensity than others. [220] For the present report, PCA is done on the correlation matrix by using the function `pca` from Matlab that centres the data by default.

A schematic showing the work-flow for doing PCA of single-cell Raman spectroscopy data is shown in figure 2.4, with a quick example with a few cell spectra and background spectra. In this example, 3000 different shift values were acquired and used, from which the correlation

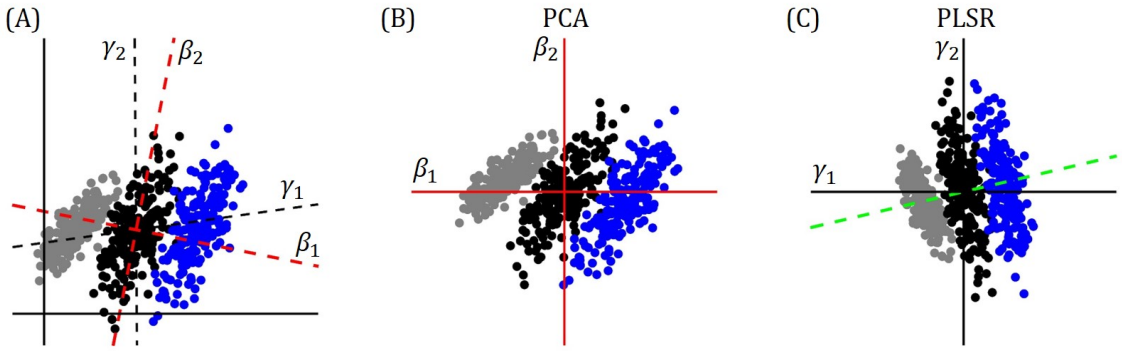


Figure 2.5: **(A)** 3 groups of normal random data were created and labelled as three different classes, with responses $0, 1, 2$ respectively. PCA was done on the data, with coefficients indicated with β_1 and β_2 in dashed red, maximizing the system variability. On the other hand, PLSR was done of the data using the given responses, obtaining the coefficients γ_1 and γ_2 indicated in dashed black, as a tradeoff between maximizing variability and fitting the responses. **(B)** Shows the data projected onto the coefficients for PCA and **(C)** for PLSR. The green dashed axis shows the responses loading.

matrix was calculated and diagonalised. The eigenvectors were then sorted and the scores of the discriminating eigenvalues were obtained and plotted.

PCA is widely used in the literature [57, 59, 62]. For example, Pallaoro *et al.* 2015 [52] use PCA to obtain the main components of the PDMS[1] background of their devices, cancer cells and healthy cells and Kein *et al.* 2012 [56] used PCA for analysis Raman images.

2.2.1.2 Partial Least Squares Regression

PLSR is a dimensionality reduction method alternative to PCA that allows assigning expected outputs to each of the groups, finding the components that correlate with a particular characteristic of the classes [221]. For example, they can be used for observations of a process over time or for observations of a process at different concentrations of a particular substrate if we expect the observations to change linearly with the changed parameter. In chemometrics, it has previously been used to correlate metastasis potential with metabolic data. [222]

While PCA can still be helpful with these type of datasets, PLSR can have the benefit of picking up the variability of interest that correlates with our known experimental parameters. In this sense, PLSR is in between unsupervised and supervised multivariate methods.

PLSR works by projecting the observations X and the expected outputs Y so that the projected matrices X_S and Y_S have maximum covariance, where E and F are matrices with random errors assumed to be normal and identically distributed and X_L and Y_L are the loadings for X and Y respectively. PLSR chooses directions that have high variance and high correlation with the response.

$$X = X_S X_L^T + E \qquad Y = Y_S Y_L^T + F \qquad max[cov(X_S, Y_S)] \tag{2.28}$$

The derivation of the equations for PLSR is tedious and is beyond the scope of this thesis.

[1] Polydimethylsiloxane

As an illustration of PLSR, both PCA and PLSR were done in the same datasets as shown in figure 2.5, where the expected outputs of the three classes were $\{0, 1, 2\}$ respectively. PCA and PLSR obtain different coefficients or loadings when fitting the data, as PCA maximizes the variability of the system and PLSR additionally performs the fitting with the given responses. Note that the PLSR coefficients are not perpendicular but PCA coefficients are. In this report, PLSR will be used to find a relation between cancer progression and Raman biomarkers.

2.2.2 Supervised Multivariate Methods

In this subsection, we will talk about training and validation as the main method to evaluate the performance of a supervised method, and we will talk about LDA, DTs and SVMs as the main supervised methods used in this thesis.

2.2.2.1 Training and Validation

When a dataset is large enough, it is good practice to split it into training and test sets to both avoid overfitting and also estimate the performance of the classification on new data.[215] Both datasets must reflect the original distribution and the original dataset must be randomly shuffled before the split phase to avoid a correlation between consequent elements. [215] For the splitting it is common to use a k-fold, where the data is split in k roughly equal-sized parts, where $(k-1)$ are used for training and the last one is used for testing. By shuffling the parts we can repeat this process k times obtaining an estimate of the performance of our model with the current dataset. This process is called *Cross-validation*. Typical choices of k are 5 or 10. [223]

Alternatively, the *bootstrap* method could be used, where a subsample of the data is randomly chosen from the data in every repetition e.g. the data is split in two randomly, using half for training and the other half for validation. A combined approach is also possible, where the data can first be randomly shuffled and then divided 5-fold and doing cross-validation, with 10 repetitions. This will generate 50 models and allows estimating the performance and its error as shown in figure 2.6.

2.2.2.2 Linear Discriminant Analysis

LDA is a linear supervised method and thus is trained with a subset of the sample whose expected outputs are already known. [34, 223] A predictor $G(x)$ is obtained and can be applied onto unknown data, returning a probability of each of the observations belonging to each of the classes. LDA is a linear decision boundary classifier, meaning that it finds the linear hyperplane that best separates each of the classes from the remaining. That is, the input space is divided into regions of constant classification with piecewise hyperplanar decision boundaries. It predicts the output Y using a linear model as shown in equation (2.29), where $\vec{\beta}$ are the coefficients of the model and $\vec{x}$ the input variables, where we include 1 as the first element of $\vec{x}$ for convenience, and we denote as x_j the elements of this vector.[223]

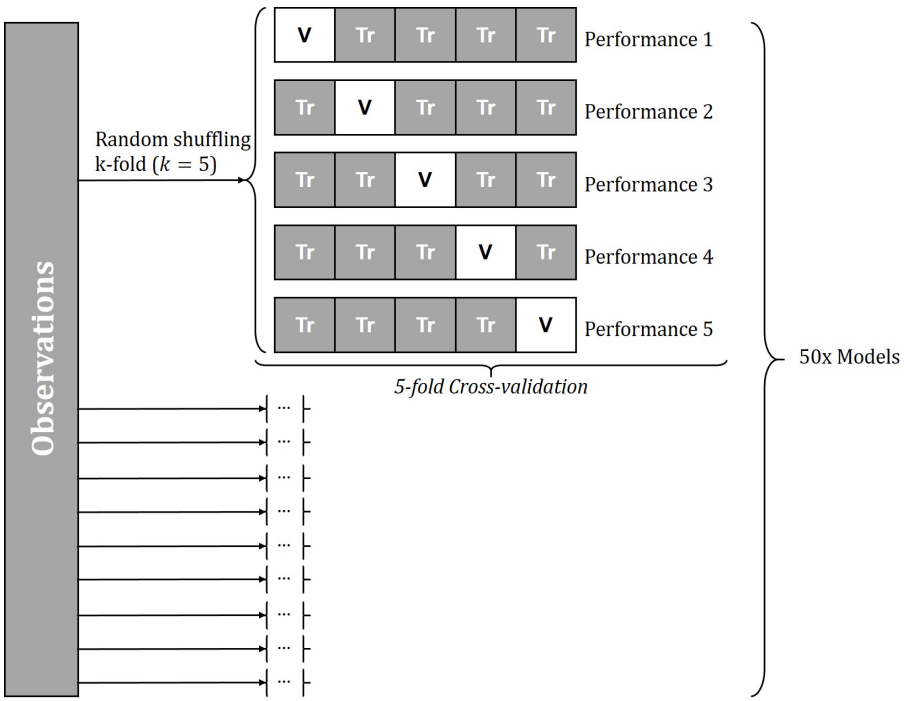


Figure 2.6: Schematics illustrating 10 repeats of k-fold cross validation when $k = 5$, where random sampling is done every repeat, and then the data is divided in 5 subsets, where one is chosen as the validation group (V) and the rest are used as training (Tr). This provides 50 Models with different performances that allow estimating a final performance for the classification of the data.

$$\vec{y} = l_0 + \sum_{j=1}^{p} x_j l_j = \vec{x}^T \cdot \vec{\beta} \tag{2.29}$$

To fit the model, a least squares method can be used, minimising the residual sum of squares of the difference between the expected output on the training data $\vec{y}$ and the output of the model with the current β onto the variables of the training set X, a matrix where each row is one observation, so the expression for the *Residual Sum of Squares* is:

$$RSS(\vec{\beta}) = \sum_{i=1}^{N} (y_i - x_i^T \vec{\beta})^2 = (\vec{y} - X\vec{\beta})^T (\vec{y} - X\vec{\beta}) \tag{2.30}$$

The minimum is obtained by differentiating with respect to $\vec{\beta}$ and equating to 0, obtaining the following equation – assuming that $X^T X$ is non-singular and thus can be inverted.

$$X^T(y - X\vec{\beta}) = 0 \quad \rightarrow \quad \vec{\beta} = (X^T X)^{-1} X^T \vec{y} \tag{2.31}$$

and thus, by calculating the value of β from the training set, the output of any point $y_i = x_i^T \vec{\beta}$ can be predicted just with a linear equation with as many parameters as input variables.

An example of the output of a linear model when 4 classes are available is shown in figure 2.7, made using the function `fitcdiscr` from Matlab with a linear model. In this example,

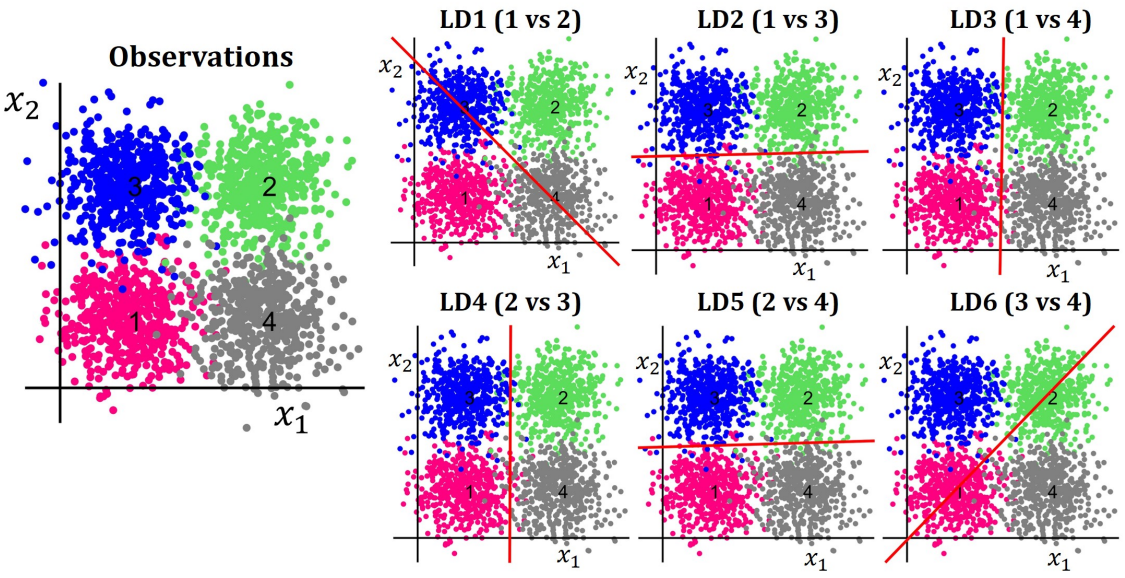


Figure 2.7: Example of an LDA model ran on 2D data from 4 different classes. Classes 1-4 have been labelled and are shown in magenta, green, blue and gray respectively. In a 2D system, each LD can be plotted as a straight line. When running LDA, 6 LDs are obtained corresponding to the lines of best separation between each 6 pairs of classes.

the model outputs 6 different LD[1]s corresponding to the 6 possible pairs of different classes – for N classes, $N!/[2!(N-2)!]$ discriminants are output. [224] However, how are these discriminants used to classify new data?

For this, the posterior probabilities should be calculated. In Bayesian statistics, the posterior probability is the probability of a certain thing when taking into account its context – ie. the probability of an observation x belongs to a certain class G: $P(x|G)$. Bayes theorem gives equation (2.32). [224]

To put some context into this, imagine that a new observation for the dataset in figure 2.7 is made. The probability of the new observation belonging to class 3 for instance ($P(G=3|x)$) is the probability of those coordinates belonging to class 3 based on the training we did, and we want to calculate the posterior probability ($P(x|G=3)$), with $P(x)$ the probability of those coordinates being observed at all, and $P(G=3)$ the probability of something belonging to class 3 instead of all other classes. In general, we can have uniform prior probability $P(G=3)=1/m$ where m is all the number of classes $m=4$, or an empirical probability that is the frequency of class 3 observations in the training model. $P(x)$ is usually calculated as a normalization constant to ensure that $\sum P(x|G=3)=1$. [224]

$$P(x|G=3)=\frac{P(G=3|x)\cdot P(x)}{P(G=3)} \tag{2.32}$$

Now to calculate the $P(G=3|x)$ in a model, the class conditional densities $f_k(x)$ are used, that are in general multivariate Gaussians for linear models as shown in equation (2.33), where Σ_k is the covariance matrix of each class and $|\Sigma_k|$ is its determinant, and $\bar{x_k}$ is the mean for each class. [224]

[1] Linear Discriminant

$$f_k(x) = \frac{1}{(2\pi)^{p/2} \cdot |\Sigma_k|^{1/2}} \cdot e^{-\frac{1}{2}(x-\bar{x_k})^T \Sigma_k^{-1} (x-\bar{x_k})} \tag{2.33}$$

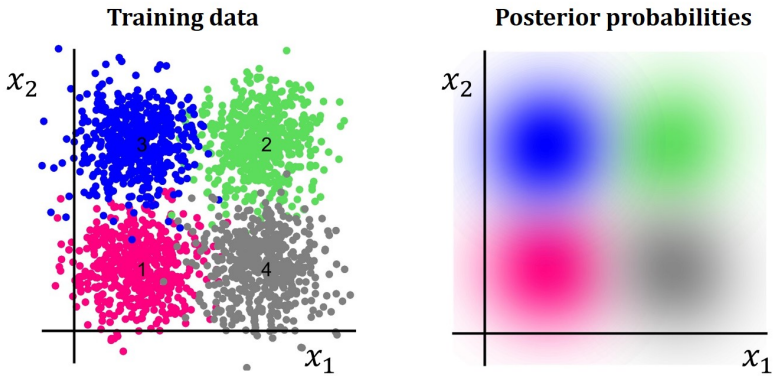


Figure 2.8: *(left)* Data from figure 2.7 and *(right)* Posterior probability plot from the trained LDA model ran on that training data.

Returning to the example, figure 2.8 shows the data from figure 2.7 and the calculated posterior probabilities from the fitted LDA model using equation 2.33. For each new observation, the probability for each class is calculated according to the model and the predicted classification is obtained. In this thesis, LDA will be used to classify different single cell measurements into different cell types in a similar fashion as in this example. However, the data will comprise observations in thousands of different wavenumbers, adding complexity to this 2D exercise. For this reason, a more powerful approach is to use PCA in conjunction with other multivariate methods such as LDA [57] or CA [46, 56, 74], so a first dimensionality reduction is done and only the relevant PCs are used to train the LDA model. Common dimensionality reductions keep the first PCs that explain up to 75% or 95% of the variability of the system.[225] Here, PCA/LDA will be used, where only the first PCs that explain 95% of the total variability of the system are considered.

2.2.2.3 Discrimination Trees

Tree classifiers, also known as DTs, have been less widely used [226, 227] and even though they are sometimes less powerful than the previously mentioned classifiers, their output is easier to relate to the original spectral features and they can also capture non-linear relationships within the data.

Tree-based methods partition the feature space into a set of rectangles and then fit a simple model in each one. They are simple to understand yet powerful models. A very popular method is CART, with C4.5 being a major competitor. Partitions will be of the shape $x_k = c$, in a particular variable. For chemometric data, this means the system chooses a particular wavenumber for the partition, highlighting wavenumbers relevant to the classification. Binary trees will partition at a particular point $x_1 = t_1$ and only consider two options: either $x_1 < t_1$ or $x_1 \geq t_1$. This two spaces can then be partitioned again, in a fashion of *"If $x_1 > t_1$ then consider the partition $x_2 = t_2$"* and so on. [223] This has been illustrated in figure 2.9,

where similar data to the one used in figure 2.7 was used to train a DT. This stratification is popular among medical scientist and has straight-forward interpretation.

The final regions will be defined by the density function $f(x)$ from equation (2.34), where R_m are the M different regions, c_m are the classes, and x the inputs.

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m) \tag{2.34}$$

In order to choose the right partitions, the optimum variable j and the optimum point t need to be chosen for each binary partition so that our 2 classes of interest show the best separation. There are different approaches for the minimization, where we can use misclassification error, Gini diversity index or Cross-entropy deviance among others to penalize the presence of members of other classes in a region. For example, when using the Gini index – default for the Matlab algorithm for `fitctree` – for a certain node:

$$GI = 1 - \sum_{i} p^2(i) \tag{2.35}$$

where we sum over the number of classes that reach the node, and $p(i)$ is the fraction of classes with members of class i that reach the node. If a node only has members of one class, then its fraction is 1, and $GI = 0$: a *pure* node. The higher the value of GI, the higher the impurity of the node.

As shown with the example data from figure 2.9, it is very easy to overtrain a tree. One alternative to avoid this is to first overtrain it, and the prune it, removing the splits that contribute less to the separation. This will depend on the data of interest.

Another popular methodology is ID3, whose latest version is C4.5 and C5.0 [228]. The main difference with CART is that after they build the tree, the algorithm simplifies it by removing some splits that don't change the final outcome. The final set of rules not always follow a tree structure, but their simplicity can be attractive. [224]

2.2.2.4 Support Vector Machines

Support Vector classifiers or SVMs are very powerful classification methods [229, 230], but it is sometimes difficult to extract useful knowledge from the trained models. In linear SVM, the best hyperplane to separate classes is found in a similar fashion to LDA, but allowing for some margin width where the classes are a bit mixed. For this reason, points too close to the boundary do not play a big role in shaping the boundary. SVM can have different kernels, including nonlinear kernels, although linear kernels tend to achieve good class separation. Three popular choices for the kernel in SVM literature are polynomial, radial (exponential transformation) or neural network (hyperbolic tangent transformation). [223]

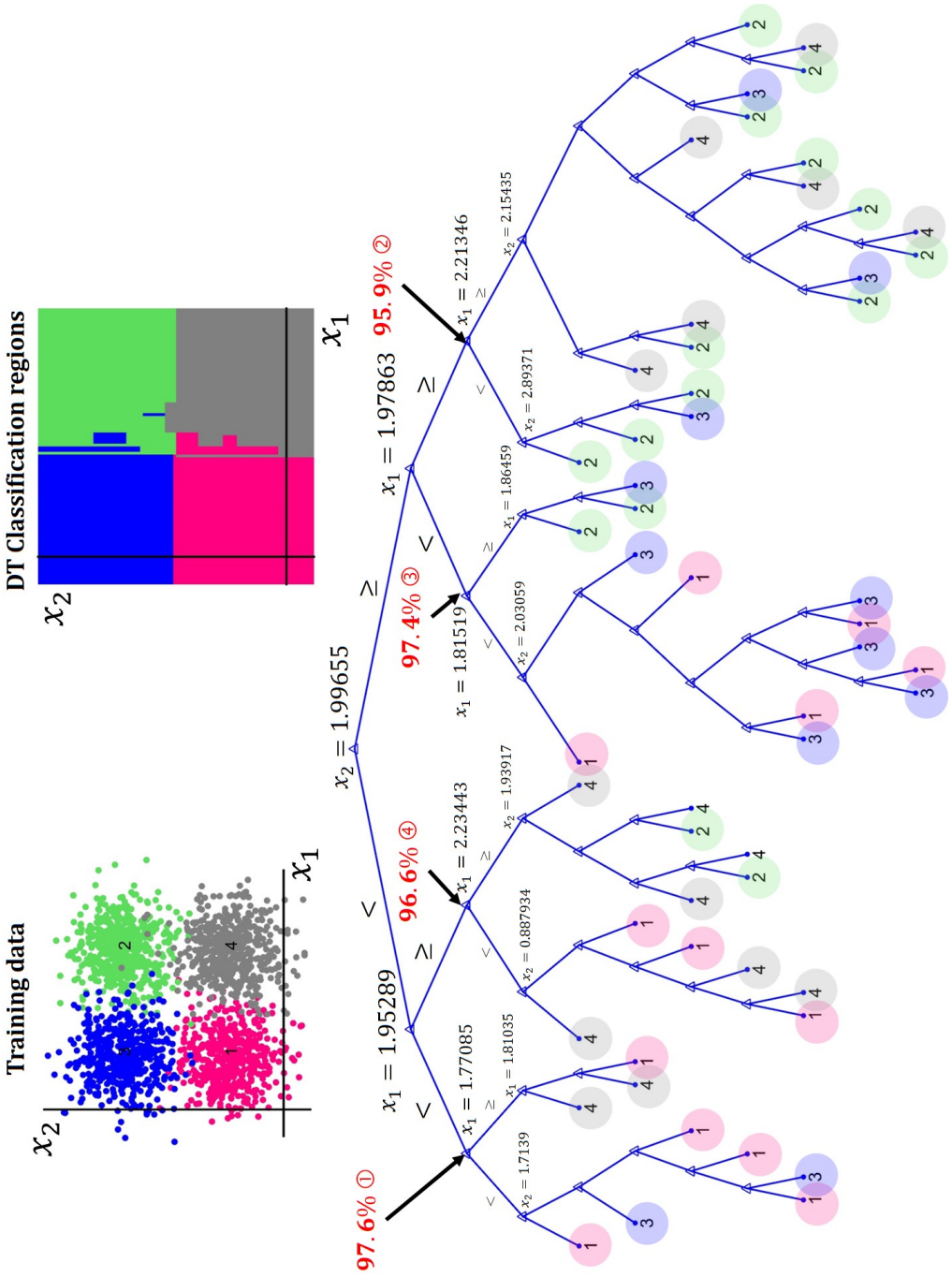


Figure 2.9: Using similar data to the one in figure 2.7 we trained a DT, obtaining the rectangular regions for classification as shown. The binary tree is also shown in a node graph mode, where the first nodes have been labelled. Note that even though the model can be trained to obtain a perfect classification of the training data, the first 3 levels of the tree already achieve > 95% classification and the remaining partitions correspond to overtraining of the model.

2.2.2.5 Evaluation of a Classifier and Confusion Matrices

In a two-class classification problem, the performance of the classifier can be evaluated by comparing the observations correctly and incorrectly classified. In a Positive/Negative problem, for instance, *True Positives* (TP) are those positive observations classified as positives, and *True Negatives* (TN) are the negative observations classified as negatives. This is useful in a positive/negative diagnostic test for a disease. For instance, a screening program will always try to have the lowest possible false negatives, usually increasing the false positives. A way of measuring the ratio of false positives and negatives for a certain classifier is to calculate the *sensitivity*, *specificity* and *accuracy* of the test:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2.36}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2.37}$$

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{2.38}$$

Sensitivity is defined as the ratio of positives that are true positives (equation (2.36)), specificity is defined as the ratio of negatives that are true negatives (equation (2.37)) and accuracy or correct rate is the percentage of all observations that are classified correctly (equation (2.38)).

Table 2.1: Example of a confusion matrix, showing the number of observations in each class.

		Output		
		Class 1	Class 2	Class 3
Observations	Cass 1 (N=100)	80	17	3
	Cass 2 (N=100)	0	100	0
	Cass 3 (N=100)	0	50	50

When obtaining a classifier for a multi-class problem, a lot of information can be obtained about its performance, and a confusion matrix is better used. An example of a confusion matrix is shown in table 2.1. In this example, 100 observations were made of each class. However, the classifier did not classify them all the same way. For example, when observing the Class 2 row, all members of Class 2 were correctly classified as Class 2 with zero false negatives, giving a very good sensitivity. However, when looking at all the observations that the model classified as members of Class 2, 17 members of Class 1 and 50 members of Class 3 are misclassified as members of Class 2, giving very bad specificity.

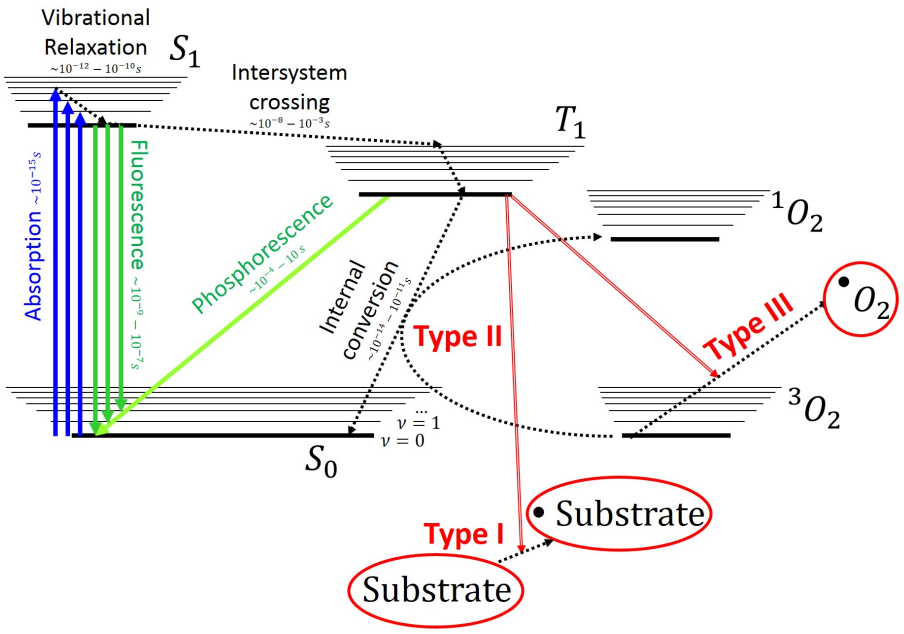


Figure 2.10: Jablonski diagram for the main processes involved in PDT, where S_0 is the first single state of the PS, S_1 is the excited singlet state and T_1 is the triplet state. The PS[1] can absorb a photon and transition into S_1, where it will likely undergo vibrational relaxation onto the ground vibrations state of S_1 and either: transition onto S_0 and emit a photon (Fluorescence); transition onto S_0 in a nonradiative way, or undergo intersystem crossing onto T_1. From T_1 it can radiatively transition onto S_0 (Phosphorescence); nonradiatevely transition onto S_0 transferring the energy to oxygen tiplet $^3O_2$ (Type II), that transition onto its single state $^1O_2$, or oxidise, donating an electron to either $^3O_2$ (Type III) or to an alternative molecular substrate (Type I).

2.3 Theory of Photodynamic therapy

PDT[2] is a cancer treatment method that involves the administration of a PS that preferentially accumulates in the tumour, and the application of light of a certain wavelength. The PS can be administered orally, intravenously or topically within an ointment. The absorption of the light induces a photochemical reaction, which usually involves oxygen present in the tumour, and selectively kills the tumour cells sparing the surrounding healthy tissue. [28]

The light used in PDT can be a laser but also incoherent light, chosen to correspond with a strong absorption band of the PS. However, it is important to consider the penetration depth of the chosen light, higher for higher wavelengths. For 460 nm light, the penetration depth in homogeneous human skin is around 0.4 mm, and for 630 nm light, it is around 3.6 mm. However, it also varies with tissue type, where blood content lowers the penetration depth. [28]

When a PS molecule absorbs a photon, it transitions from the ground singlet state (S_0) to an excited singlet state (S_1) due to the excitation of an electron. The excited molecule can lose its excess energy in a variety of ways, both radiative or non-radiative. It can return to the ground state by emitting a photon (fluorescence); transfer energy to a high vibrational level of the ground state (internal conversion) and then undergo vibrational relaxation, or transfer energy to the triplet state T_1 (intersystem crossing). T_1 can transition back to S_0

[2] Photodynamic Therapy

(phosphorescence or intersystem crossing). However, this is of low probability as the spin quantum numbers are different, and thus the lifetime of T_1 is longer. Alternatively, it can undergo a photochemical reaction.

Photochemical reactions can occur with different mechanisms, illustrated in figure 2.10:

- Type I reaction: T_1 reacts with surrounding molecules, by hydrogen or electron transfer, producing transient free radicals. This radicals may then react with oxygen, producing superoxide radical anions ($O_2^{\bullet -}$) which are ROS[1], toxic for the cells.
- Type II reaction: T_1 transfers energy into $^3O_2$ that forms singlet oxygen $^1O_2$ which is also a ROS.
- Type III reaction: T_1 can also transfer an electron to oxygen, forming $O_2^{\bullet -}$. The PS is then oxydised.

The contribution of each process is variant, but it is generally accepted that type II reactions are dominant in PDT. ROS interact with amino acids, lipids and guanine residues in DNA among others. Due to its short lifetime ($\approx 4\mu s$ in H_2O) the range of reactivity of $^1O_2$ in water is of $\approx 125nm$, shortened by reaction with cellular molecules.[231] In a mammalian cell, measured values range between 10-100 nm,[231] therefore the type of damage is highly dependent on the sensitiser localisation both within the cell and within the tumour. For instance, sensitisers can accumulate in cancerous cells or in the vasculature, both causing tumour death but with different mechanisms. [28]

2.4 Microfluidic Theory

The first miniaturised gas chromatography was created in 1975, but it was not until 1991 that microfluidics was popularised for electrophoretic separation systems, electro-osmotic pumping systems, micromixers or cytometers among other [232]. At the microscale, different forces become dominant over those experienced at the macroscale, such as laminar flow, diffusion, fluidic resistance, SAV[2] and surface tension [233].

The Reynold's number (Re[3]) of a fluid describes the ratio of inertial forces to viscous forces. Its definition for a rectangular channel is shown in equation (2.39), where ρ is the fluid density, v is the characteristic velocity of the fluid, μ is the viscosity and D_h is the hydraulic diameter, that depends on the width (w) and height (h) of the channel. For $Re < 2300$ viscous forces are dominant and flow is laminar, whereas above 2300 the flow shows a turbulent behaviour [233]. Figure 2.11 shows the differences between laminar and turbulent flow.

$$Re = \frac{\text{innertial forces}}{\text{viscous forces}} = \frac{\rho v D_h}{\mu} = \frac{\rho v w h}{\mu(w+h)} \tag{2.39}$$

[1] Reactive oxygen species
[2] Surface area to volume ratio
[3] Reynold's number

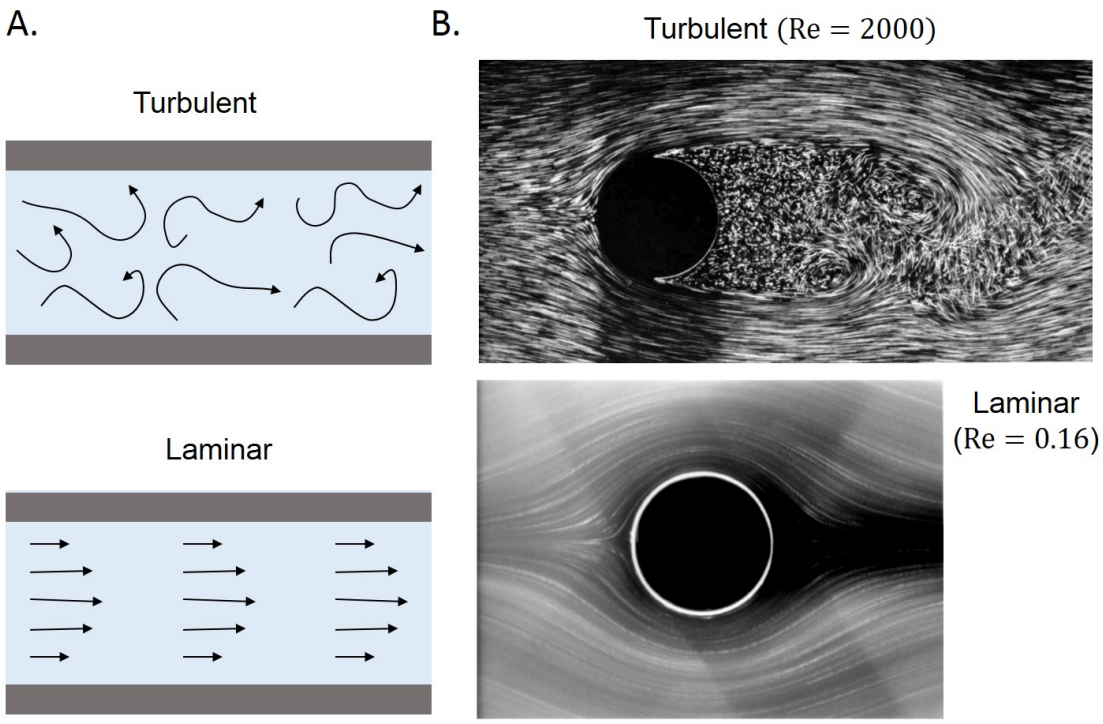


Figure 2.11: (A) Schematics of the difference between turbulent and laminar flow. **(B)** Images of turbulent and laminar flow around a 1 *cm* cylinder visualized by Van Dyke in 1982 using aluminium powder in suspension, and reproduced with permission from his book *An Album of Fluid Motion*.[4]

Laminar flow systems can be counter-intuitive as they follow different physics than the macroscopic systems of our everyday life. In a laminar flow regime, one or more streams flowing in contact do not mix except for diffusion. Another difference of microfluidic systems to macroscopic systems is the SAV, that is drastically increased by orders of magnitude. A Petri dish full of water has a SAV of around 5 cm^{-1} whereas a microchannel 50×50 μm^2 and 33 mm long has a SAV of 800 cm^{-1}. A large SAV allows macromolecules to quickly diffuse and adsorb to channel surfaces. Surface tension also becomes very significant at the microscale. Capillarity is inversely proportional to the radius of the capillary, and thus the lengths that liquids will travel in microchannels just due to capillarity are larger.[233]

Laminar flow simplifies fluid calculations as it makes flow controllable. For instance, fluidic resistance in a microchannel has a very well known equation, defined as the ratio between the pressure drop ΔP and the flux Q.[233, 234] Equation (2.40) shows this relation for a rectangular channel, where C is the perimeter of the channel in the flow direction, A is the area of the channel, L is the length of the channel, w is the width, h is the height, μ is the fluid viscosity and β is the aspect ratio of the channel defined as $\min[w/h, h/w]$ so $0 < \beta < 1$.[235]

$$R_h = \frac{\Delta P}{Q} = \frac{f(\beta)\mu C^2 L}{32A^3} = \frac{f(\beta)\mu(2w+2h)^2 L}{32(wh)^3}$$
$$f(\beta) = 96(1 - 1.3553\beta + 1.9467\beta^2 - 1.7012\beta^3 + 0.9564\beta^4 - 0.2537\beta^5) \quad (2.40)$$

Resistance calculations for microfluidic channels are used in Chapter 3 for optimization of microfluidic traps designs in section 3.8.4 (page 81), where the different types of microfluidic traps are introduced.

Part II

Methods

Chapter 3

Methods

The first chapter introduced the thesis objectives and presented a literature review on the main topics regarding this thesis. The second chapter presented a theoretical introduction to Raman spectroscopy, multivariate methods, photodynamic therapy and microfluidics. In this chapter, the facilities available for both the Raman spectroscopy and the microfluidic fabrication are described and the main data obtained for the optimisation of the system is shown. This is followed by a description of the main preprocessing steps used and their effect on improving the data quality for the analysis. Then, the sample preparation and the fabrication of the two used platforms are described: the *off-chip* setup and the *on-chip* setup. This includes the cell culturing methods, the *off-chip* setup preparation, the soft lithography techniques used for microfluidic fabrication and the methods used for any simulation.

3.1 Microscopy and Spectroscopy Facilities

Figure 3.1 shows the main components of a Raman microscope. Figure 3.2 shows the instrument used, with some of the components labelled. The instrument used for this thesis is a Renishaw inVia Spectrometer System integrated with a Leica DMi8/SP8 laser scanning confocal microscope. A diagram of the Raman spectrometer optics is shown in figure 3.1. Two lasers are available for excitation: a 532 nm DPSS Diode visible laser (50 mW, 22mW at the sample) and a 785 nm Diode near-IR[1] laser (100 mW, 45 mW at the sample). Both lasers are mounted with kinematic laser baseplates, and the 532 nm laser has a plasma filter. A pinhole is available to clean the laser shape. Motorized ND[2] filters are available from 0.00005 to 100% with 16 different power levels. A beam expander is available, with settings from 0-100% laser expansion. Increasing the size of the laser by defocusing the beam allows obtaining the average signal of a bigger region. The light then goes through a motorised shutter and a system of mirrors, through the chosen objective and to the sample in an inverted configuration. Objectives available are summarised in Table 3.1. The sample sits on in-house designed 3D printed holders with magnets to reduce any movement of the sample during acquisition. The holders sit on an automated XYZ stage, although the Z movement of the stage is not used – instead, the objective is adjusted for the Z movement.

[1] Infrared
[2] Neutral density (filter)

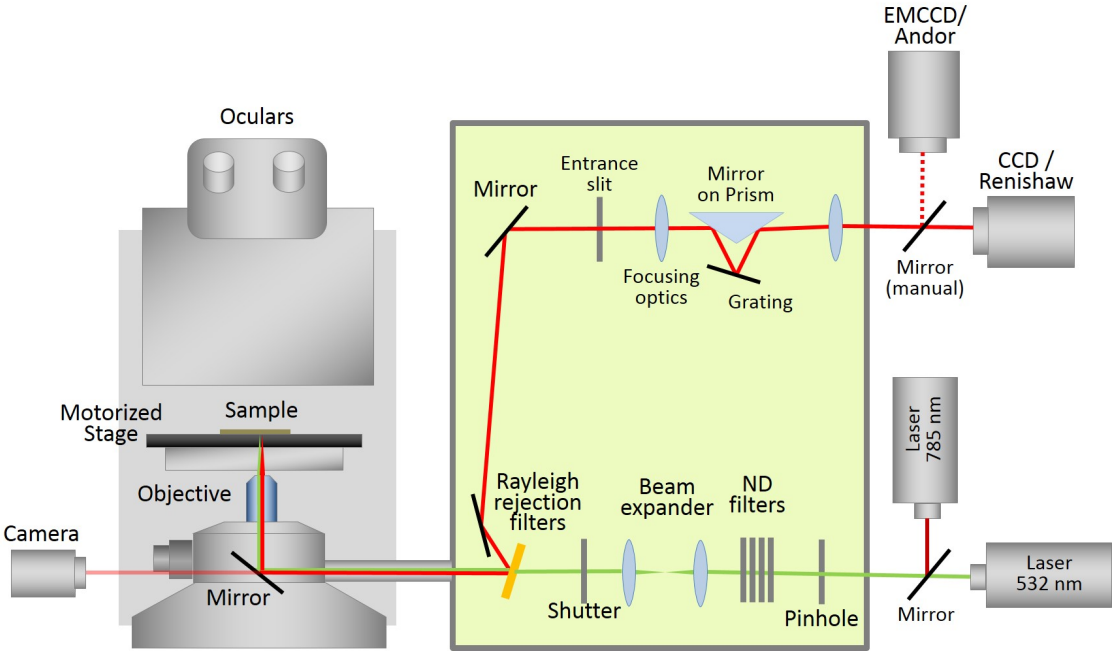


Figure 3.1: Main features of the optical configuration of the Renishaw inVia Raman microscope. This is a simplification for clarity - see figure 3.2 for details about the actual setup.

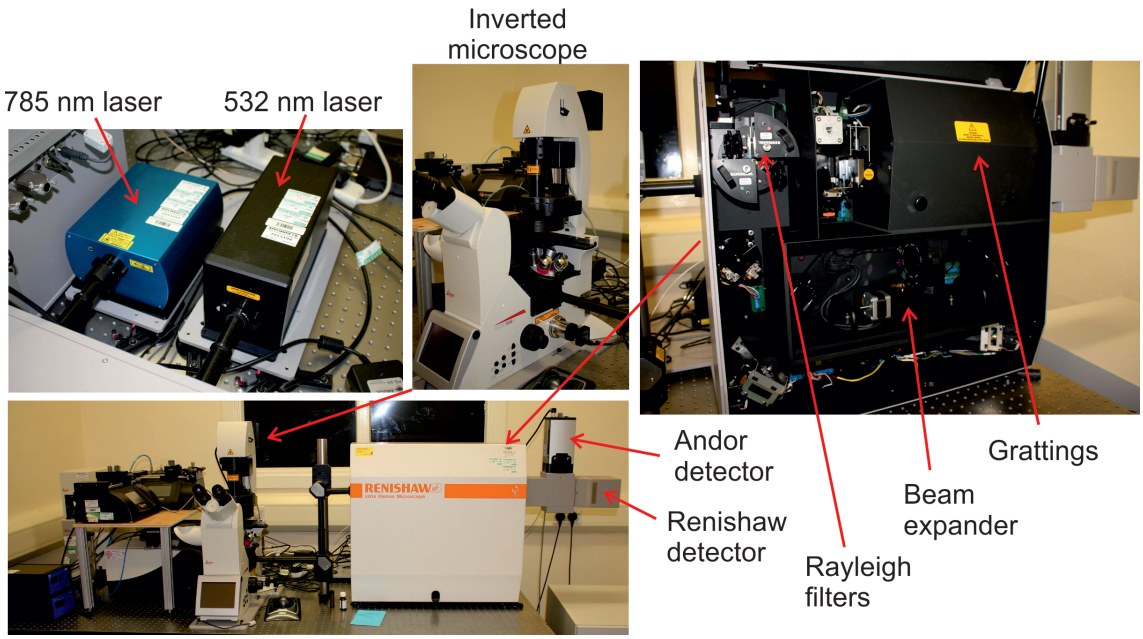


Figure 3.2: Pictures of the microscope used, where some of the main components have been labelled.

The stage is a HSES[1] with a step size of 0.1 μm. Wire software – by Renishaw – was used to control all the acquisition settings for the system, except for change in the detector, that required manual changing by the user, and changes in the optical path of the microscope, that are controlled by the fluorescence confocal software (LAS Leika) or manually by the user through the microscope touch screen.

Table 3.1: Summary of all the available objectives in the Raman-Confocal microscope and their resolution, where FWD stands for Free Working distance and N.A. for Numerical Aperture. The size of the Airy disk was used as an estimation of the lateral resolution.

		FWD (mm)	**N.A.**	**Airy disk** (μm)
10x	HC PL FLUOTAR	inf	0.3	2.2
20x	HC PL APO CS2	0.62	0.75	0.9
40x	HCX PL APO CORR CS	0.21	0.85	0.8
63xair	HC PL FLUOTAR CORR	2.6-1.8	0.7	0.9
63xoil	HC PL APO CS2	0.14	1.4	0.5
100xoil	HC PL APO CS2	0.13	1.4	0.2

The scattered light is partly reflected into a camera to observe the laser spot and for calibration reasons, but it is mostly reflected onto Rayleigh reflection filters (for either laser) that reject the intensity under 100 cm^{-1}. The entrance slit shapes the laser beam that will be focused in the grating with two default setting of 65 μm for standard confocality mode and 20 μm for high confocality mode. Then a grating separates the different wavelengths and spreads them across a CCD[2] detector. Two gratings are available mounted back to back: a 1800 lines/mm and a 1200 lines/mm for the 532 and 785 nm excitations respectively. Spectral resolution can be changed using a CCD binning control. A continuous movement of the grating allows obtaining each wavelength with high resolution in a continuous spectrum (synchro-scan), whereas a step motion allows obtaining different spectral ranges of the spectrum that can later be overlapped in a full spectrum (step scan). The grating will determine the maximum size of these windows.

Two different detectors are available: a Renishaw CCD array detector near infra-red enhanced, deep depletion (1024×256 px), Peltier cooled to -70°C, and an Andor EMCCD[3] high sensitive back-illuminated detector for UV/VIS measurements (1600×200 px) with TE[4] cooling to -100°C. These two detectors will be referred to as the Renishaw and Andor detectors respectively.

Additionally, the microscope has a fluorescence light source EL6000 (LQ-HXP 120 LEJ) which is a mercury short-arc reflector lamp (120 W / 45 mm focal length) with a wavelength range of 380-680 nm. Filter cubes are available for different standard fluorescence settings, with BP[5] or LP[6] filters. On a BP filter, only the light around a certain wavelength is transmitted, where the settings are indicated as centre/bandwidth. In a LP filter, all the light above a certain wavelength is transmitted. The filter cubes available are for DAPI

[1] High Speed Encoded Stage
[2] Charge-coupled device
[3] Electron multiplying charged-coupled device
[4] Thermoelectric
[5] Band pass (filter)
[6] Long pass (filter)

(exc.: BP of 360/40 nm, em.: LP 425), FITC (exc.: BP 470/40, em.: LP 515), RHOD (exc.: BP 540/45, em.: LP 590) and TXR (BP 560/40, em.: BP 630/75). There is also an LED[1] bright field light source with TTL shutter available for transmitted light.

For the PDT[2] experiments, the intensity of the light source was measured using a ThorLabs Laser power meter PM100D for the DAPI filtered light for different objectives. The instrument was equipped with a sensor SM1CP1. The wavelength was set as 360 nm (central wavelength for DAPI) and all measurements were taken to 3 significant digits. For the 1° aperture – all other apertures covered the whole field of view, – the image of the aperture was taken to measure the total area exposed for the 100x oil objective to be 288.6 μm^2, for a total power density on the sample of $\approx 2370\ mW/cm^2$ allowing a minimum power density when using the 10% filter of 237 mW/cm^2.

Table 3.2: Intensity for DAPI filter settings when using different objectives and different apertures, all using the circular shape aperture.

	Objective	10x	10x	40x	100x oil
	Intensity filters	100%	10%	100%	100%
Aperture (degrees)	**Shape**	**Power** (mW)	**Power** (mW)	**Power** (mW)	**Power** (mW)
6	*Circle*	10.83	1.04	5.82	1.28
5	*Circle*	9.34	0.907	4.93	1.04
4	*Circle*	8.36	0.816	4.36	0.902
3	*Circle*	4.58	0.458	2.33	0.457
2	*Circle*	1.19	0.123	0.61	0.114
1	*Circle*	0.701	0.00747	0.0375	0.00684

3.2 Raman System Characterisation and Experimental Design

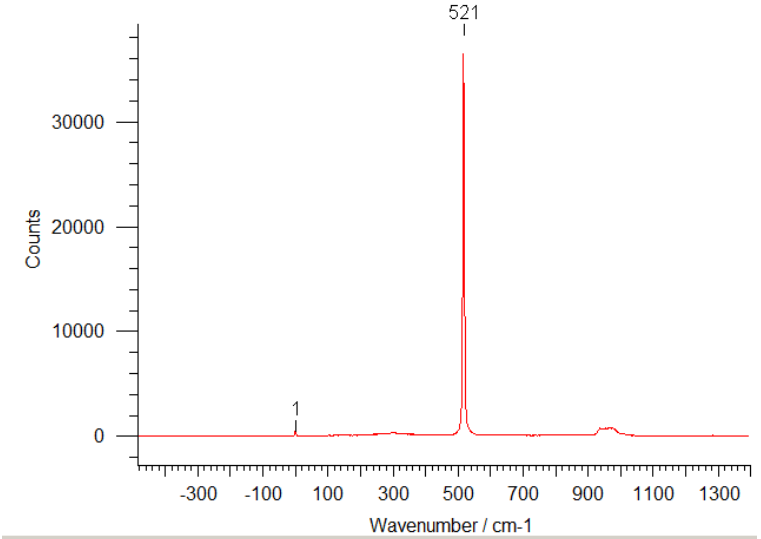
Raman spectroscopy is a sensitive and non-destructive technique suitable for analysis of living cells. However, it faces a number of challenges when applied to diagnosis: firstly, cells are inhomogeneous, so an average spectrum of the whole cell needs to be obtained [25]; secondly, long acquisition times are usually needed as the signal is weak and high laser powers could compromise the integrity of the sample [46, 47], and thirdly, Raman signals are easily overwhelmed by other signals such as autofluorescence of the sample, that shows as a baseline [25], or stronger Raman signals from the substrate. This section will show the system calibration steps, characterise the background of all the available materials, characterise the system confocality with the available settings and objectives, characterise the Raman signal for each of the available lasers and detectors and describe the beam expander settings to obtain a single cell average signal for single cell RS[3].

3.2.1 Raman System Calibration to the Silicon Peak

Motors can drift while the system is switched off or while in use for long periods of time, meaning that the references for the motors controlling the gratings can change, causing a drift

[1] Light Emitting Diode
[2] Photodynamic Therapy
[3] Raman spectroscopy

of the spectrum to lower or higher wavenumbers. To ensure reproducibility of the acquired spectra, it is essential to calibrate the system before every use. A calibration method often used and recommended by Renishaw is alignment to the sharp peak of a Silicon sample. A screenshot of a typical silicon spectrum when acquiring one window centred at 520 cm^{-1} is shown in figure 3.3 (40× objective, 532 nm excitation laser, 1 s acquisition time, 50% laser power, standard confocality). Calibration is specific for the used objective, laser/grating pair and detector, and thus one must think of the settings of the experiment before calibrating the system. When acquiring the calibration spectrum, the Wire software allows finding the peak and obtaining the position of the band. The microscope was calibrated prior to every experiment so the silicon band would be between 520.3-520.7 cm^{-1}, aiming for 520.5 cm^{-1}. The calibration files were saved for tracking any variations on the silicon peak intensities over time and for further alignment of the spectra to the exact silicon peak position, to improve reproducibility.



Peak no.	Centre	Height	Width	Area	Absolute intensity	Low edge
1	0.740042	444.162	4.455	2297.51	445.94	-29.9727
2	520.627	38361.4	5.19522	279881	38449.3	464.064

Figure 3.3: Example of the calibration silicon peak obtained from the silicon sample (40× objective, 532 nm excitation laser, 1 s acquisition time, 50% laser power, standard confocality) as a screenshot from the Wire software, with the obtained peak positions (520.627 cm^{-1}).

3.2.2 Raman Signal of Potential Substrates

The background contribution to the Raman spectra needs to be minimised so it does not obfuscate the spectral features of biological samples. The main factors that determine the background contribution are the substrate, the objective, the excitation wavelength and the confocality of the system.

Quartz and glass were initially compared to determine the best substrate for the following measurements, using Alfa Aesar fused quartz slides (1 mm thick) and Alfa Aesar fused quartz coverslips (150-250 μm) and conventional soda-lime extra white glass slides (Thermo Scientific, 1 - 1.2 mm) and borosilicate glass coverslips (Fisherbrand, 0.13 to 0.17 mm). In both cases, 50 μm PET spacers cut from a film (Goodfellow, UK) were used to create the water/buffer chamber. Figure 3.4 shows the spectra of quartz and glass setups when using the two available excitation wavelengths – 532 nm and 785 nm. Glass showed strong fluorescence when using the 785 nm excitation wavelength, known to be caused by transition

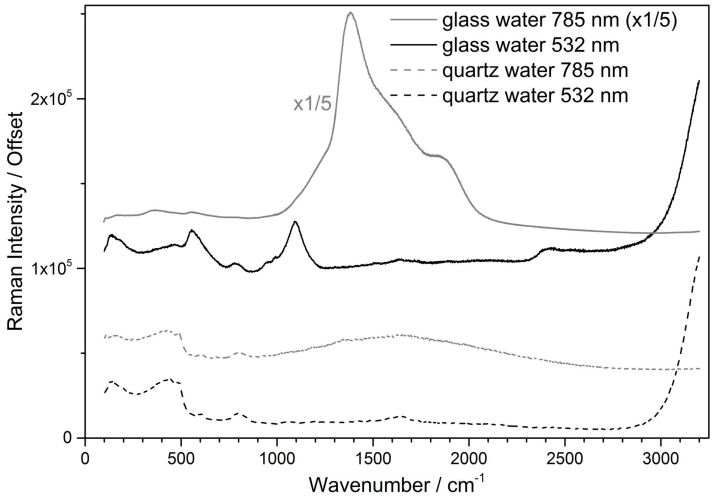


Figure 3.4: Raman signal of a glass (solid line) or quartz (dashed line) setup when exciting with the 532 nm laser (black) or the 785 nm laser (grey). The glass signal when exciting with the 785 nm laser is shown multiplied by a factor of 1/5.

elements (Cu, Mn, Cr, V, Ti), several rare elements and uranium impurities in the glass [236]. Results showed a broad band at 1386 cm^{-1} with a shoulder at 1860 cm^{-1}. The glass spectrum when using the 532 nm laser showed some strong features – bands at 552, 1100 and 2396 cm^{-1} – in the fingerprint region, that were minimal when using quartz. For the 532 nm spectra, the side of the broad water band was present above 3000 cm^{-1}, that did not show when using the 785 nm laser due to the weaker scattering at higher wavenumbers.

Quartz spectra had a lower contribution for both lasers with only a plateau until $\approx 488\ cm^{-1}$. The small glass contribution seen when using the 785 nm laser is likely due to reflections in the glass of the objective lens (20×). However, quartz coverslips are costly and fragile. For this reason, when possible, quartz was used for the experiments, piranha cleaned and reused to reduce costs. For microfluidic experiments, where reusing the coverslip is difficult, glass was used and Raman was acquired exciting with the 532 nm laser.

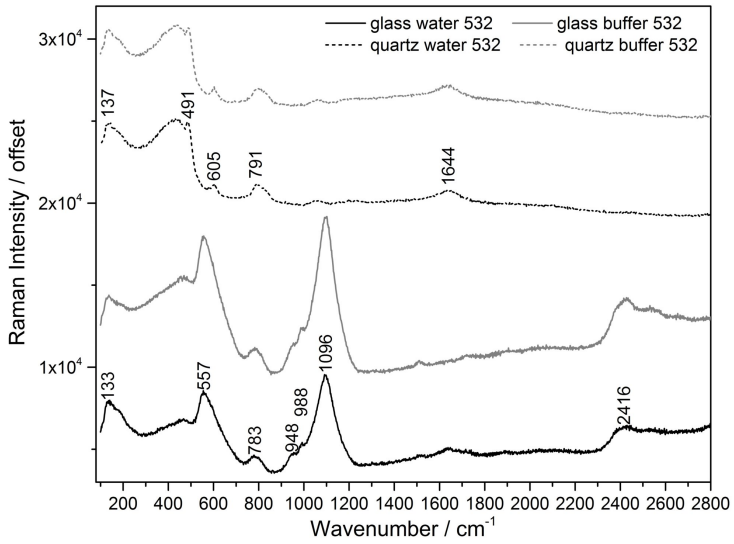


Figure 3.5: Raman signal of a glass (solid line) or quartz (dashed line) setup when acquired with the 532 nm laser with either water (black) or DPBS (grey) inside. The region above 2800 cm^{-1} was omitted as it only showed water contributions.

As the cells require at least a buffered and isosmotic solution to stay viable during the measurements, the spectra of DPBS[1] was measured (see Figure 3.5) and no visible Raman features could be found nor for glass or quartz. However, as cell debris and exosomes can

[1] Dulbecco's Phosphate-Buffered Saline

be present when measuring, the background of the surrounding buffer was measured and subtracted for all cell experiments.

3.2.3 System Confocality

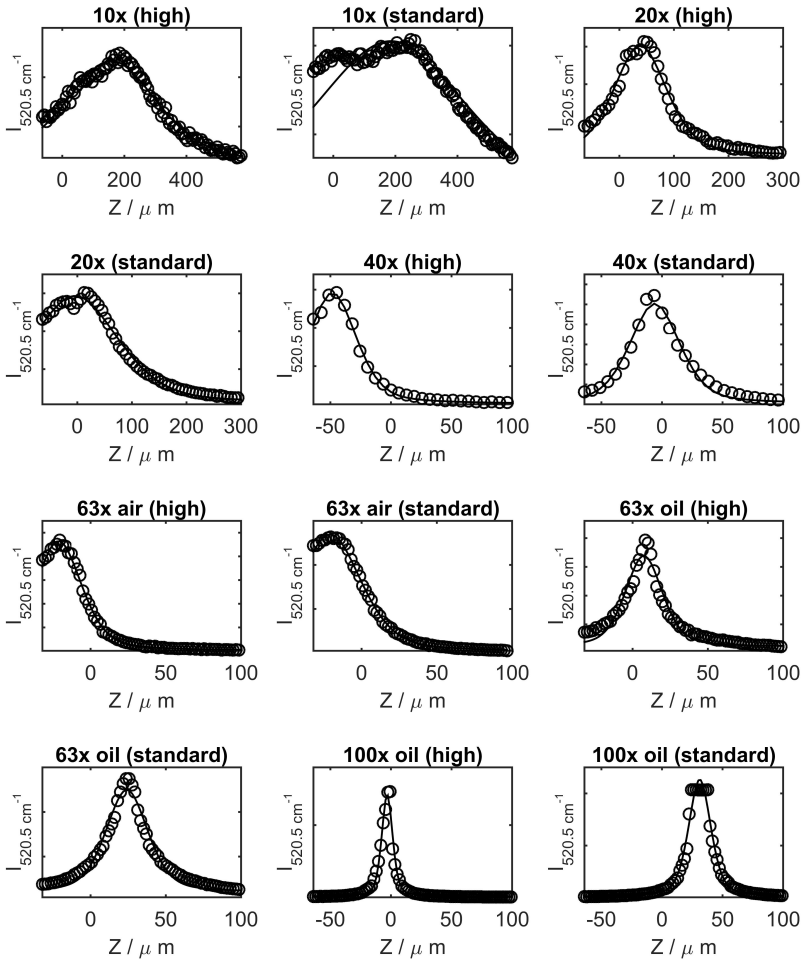


Figure 3.6: 520.5 cm^{-1} peak maximum intensity with distance when exciting with the 532 nm laser and using different objectives and different confocality slit openings (standard: 65 μm, high: 20 μm). The peaks were fitted with a GLS (solid line).

Reducing the system confocality can significantly reduce the contributions of the substrate background to the cell spectra. All the available objectives were tested with a silicon sample with slit sizes of either "high confocality" (20 μm slit opening) or "normal confocality" (65 μm slit opening). Measurements were done by acquiring the silicon peak spectrum in a Z-scan from past the focal plane, taking care to obtain enough points around the maximum intensity. The maximum intensity of the 520.5 $cm^{-}1$ for each Z-range and the fitted GLS[1] curves are shown in Figure 3.6. Confocality was expressed as the FWHM[2], obtained from the width of the fitting of the peaks. Error bars correspond to the 95% confidence interval. The

[1] Gaussian-Lorentzian sum
[2] Full-width half maximum

FWHM and the intensity results are summarized in Figure 3.7. Results showed that reducing the slit size to the *high* confocality mode considerably reduced the acquired intensity of the silicon signal. This is expected, as the acquisition volume is smaller. Reducing the slit size from *standard* to *high* increased the FWHM by even 50% for the 100× oil objective.

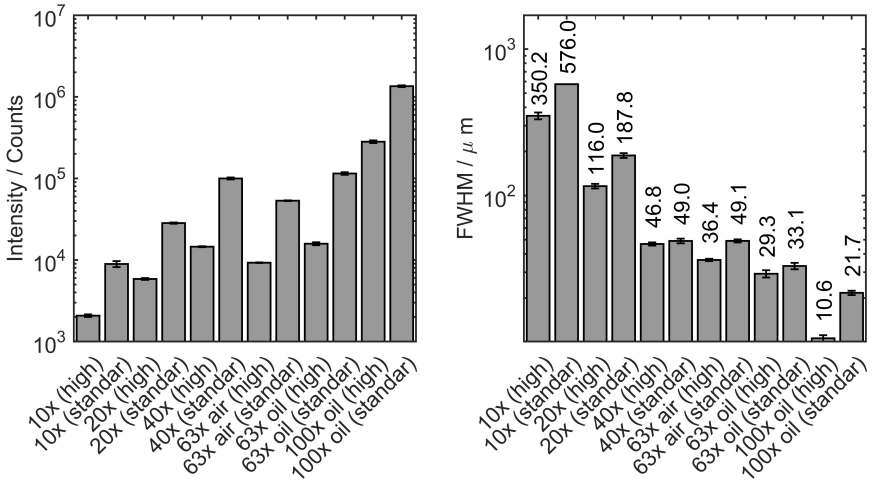


Figure 3.7: Summary of maximum intensity and FWHM results for each objective when using the "standard" and "high" confocality modes.

Best confocality values were obtained when using the higher magnification oil objectives for both confocality settings. The 100× oil immersion objective also gave higher signal intensities than other objectives. However, the 40× objective showed similar confocality values to the 63x air objective with considerably higher signal intensity and was thus chosen as the prefered objective for those samples where the oil objective was not suitable. For single cell spectroscopy measurements, the high confocality setting with the 100× oil immersion objective gave a FWHM of 10.6 μm. Using a Gaussian approximation for the band and the 68-95-99 rule, this means that 99% of the intensity is acquired from a Z-range of $\approx$ 13.5 μm, covering most of the cell volume. Even though the acquired intensity will not be homogeneous from this volume, we considered this method to be optimum for fast acquisition of a cell average signal for fast spectroscopy of single-cells in a confocal manner and suitable for acquisitions in high background chambers such as PDMS[1] microfluidic devices.

3.2.4 Comparison of Excitation Lasers Available

As discussed in subsection 2.1.3 (page 29), Raman scattering intensity is inversely proportional to the wavelength to the power of 4, meaning that using a visible laser instead of a near-IR[2] will give a stronger signal. A visible laser also gives a weaker background when using glass, whereas a near-IR gives a strong background signal with glass. On the other hand, if the sample presents autofluorescence, this can show as a strong baseline in the spectrum of the visible laser, and the visible light is also more phototoxic than the near-IR one.[237] Figure 3.8 shows the raw data of the same SW480 cell measured with the 532 nm laser (top) and the 785 nm laser (bottom). The abscissa axis shows the wavelength for each laser and

[1] Polydimethylsiloxane
[2] Infrared

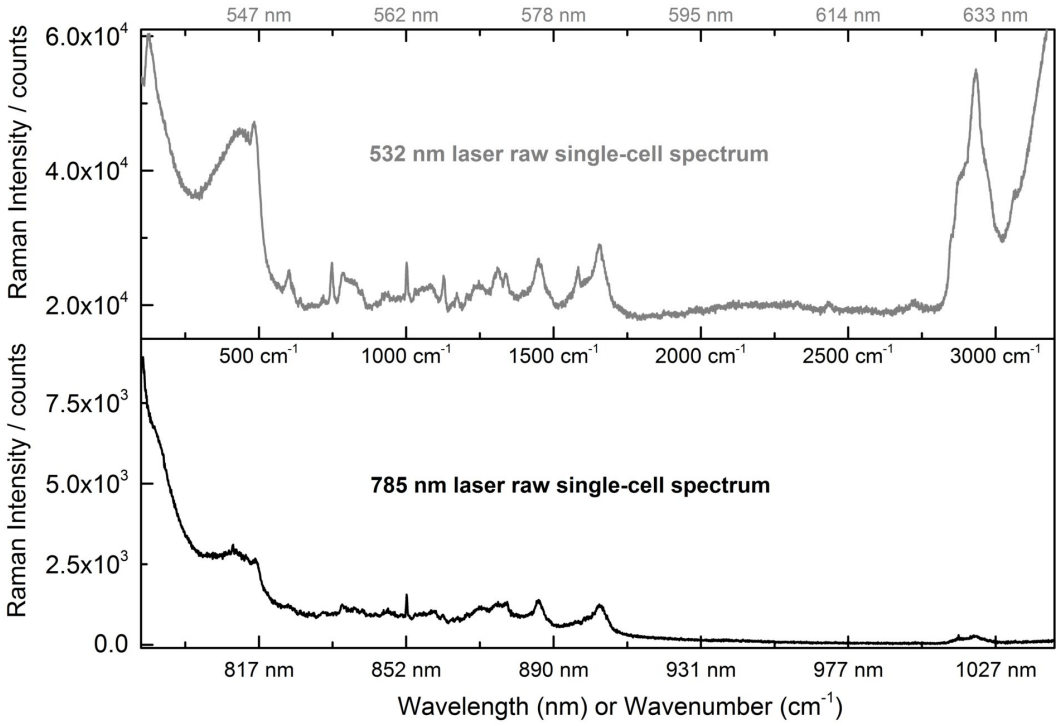


Figure 3.8: Same single-cell raw spectra (SW480 cell) acquired with either the 532 nm laser (grey) or the 785 nm laser (black). Upper abscissa axis shows the wavelength when exciting with the 532 nm laser, bottom axis shows the wavelength when exciting with the 785 nm laser, and the middle axis shows the common wavenumber values for the Raman spectra.

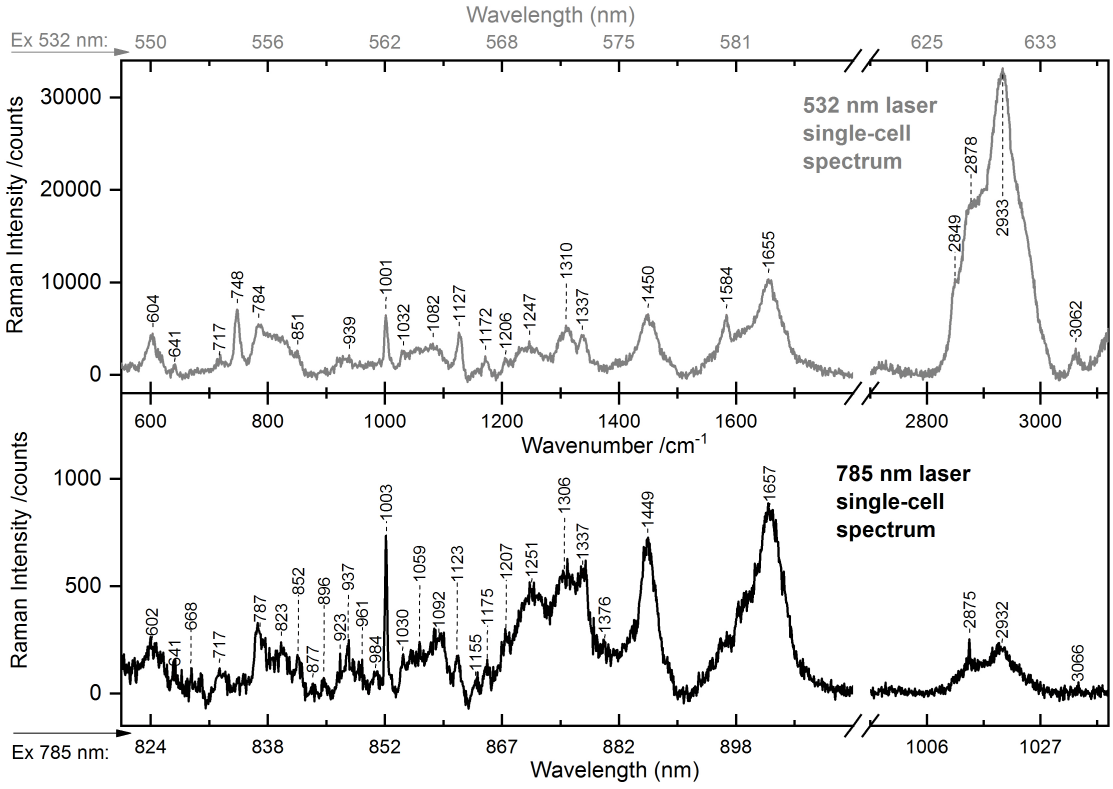


Figure 3.9: Data from figure 3.8 after baseline subtraction, where main peaks have been labelled.

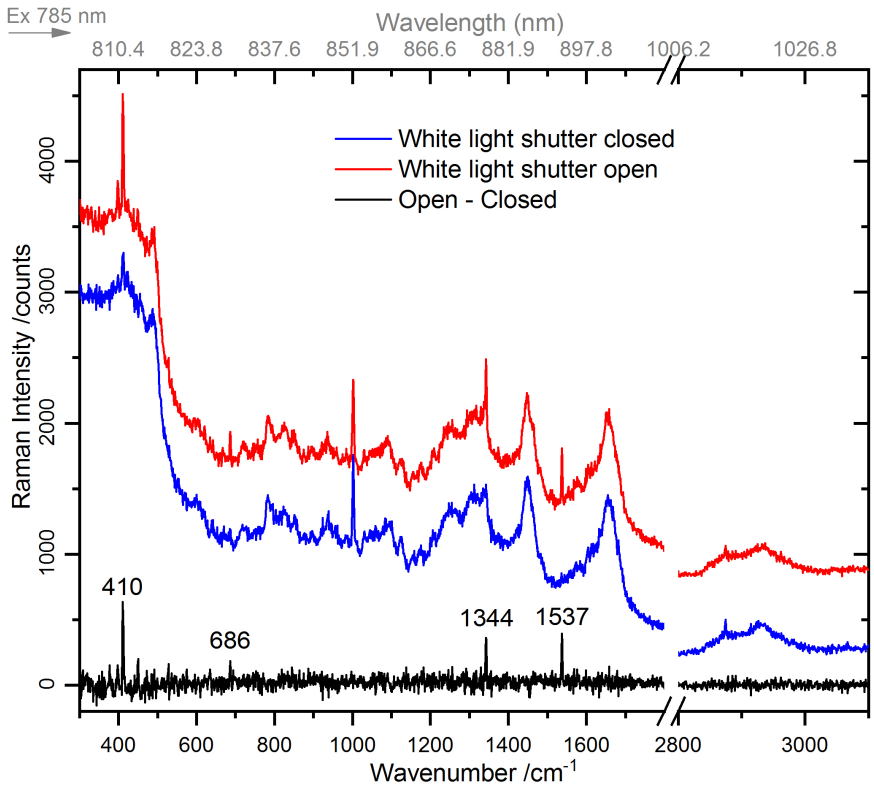


Figure 3.10: Same SW480 cell signal acquired with the manual shutter of the white light lamp open and closed, showing the appearance of a series of sharp peaks due to light contributions to the scattering spectrum. The difference spectrum has been calculated and the wavenumbers of the artefact peaks have been labelled. The top axis shows the equivalent wavelength of the Raman wavenumbers for the 785 nm excitation laser.

the shift or wavenumber value, common for both lasers. The wavelength for each laser is also shown in the top and bottom horizontal axis. When exciting with the 532 nm laser, the overall scattering intensity was stronger, with a total number of counts of $\sim 6 \cdot 10^4$, and the spectrum showed a constant baseline value of 2×10^4 counts likely due to cell autofluorescence. The water band around 4000 cm^{-1} gave a strong baseline at the higher wavenumber ends of the spectrum. 785 nm laser excitation showed weaker signal at higher wavenumbers, lowering the signal-to-noise ratio of the data from the CH stretching region.

After baseline subtraction, the spectra can be compared for both lasers, shown in figure 3.9. Both spectra showed an overall similar shape, where the positions of the main peaks have been labelled. The intensity of the visible laser signal is more than an order of magnitude stronger both in the Amide I region (1657 cm^{-1} peak) and ever more in the CH stretching region ($>$ 2800 cm^{-1} band). Some bands such as the band at 748 or 1584 cm^{-1} are absent on the 785 nm spectrum but are very strong in the 532 nm spectrum. These bands have been attributed to RR[1] of cyt C[2] (see section 2.1.5 on page 32).

Additionally, when using the 785 nm laser other light artefacts may need to be considered. In our setup, even when changing the microscope to the Raman mode, light from the white light lamp can interfere with the 785 nm laser spectrum. This can be seen in figure 3.10

[1] Resonant Raman
[2] Cytochrome C

when obtaining the signal from the same SW480 cell with and without closing the mercury lamp shutter. When the fluorescents light of the room is on, they are also present in the acquired spectrum. When mapping, these peaks will sometimes appear and disappear or change in intensity likely due to internal reflections of the microscope, and thus the shutter of the bright field lamp should be always closed when acquiring Raman spectra with the 785 nm laser. These peaks at 410, 685, 1344 and 1537 cm^{-1} correspond to peaks at 811, 830, 878 and 892 nm.

3.2.5 Comparison of Detectors and Scanning Options

There are two available detectors in the system: Renishaw (1024×256 px, IR-enhanced) and Andor (1600×200 px).

1. **Renishaw detector**: it is IR-enhanced and thus can be used for both the visible and the 785 nm laser. However, it has fewer pixels available (1024 px), giving lower wavenumber resolution.
2. **Andor detector**: is not IR-enhanced, and thus can only be used for the 532 nm laser. It has better wavenumber resolution (1600 px available) and it has the option of increasing the gain.

There are two possible scan types:

1. **Step scan**: a certain window of Raman wavenumbers are acquired at a time by projecting them onto the detector. The width of the window depends on the number of available pixels in the detector and the grating used. This scan is very fast, with no minimum exposure time.
2. **Synchro scan**: here, the grating moves over time and the detector collects just one wavenumber at a time. This scan is very slow, with a minimum exposure time of 10s, but gives very good signal-to-noise ratio.

To this end, two different experiments were conducted. First, the electronic noise of the detector was recorded at different settings by acquiring the signal in the absence of laser intensity (dark noise). This was done by acquiring with the chosen detector with the misaligned manual mirror position. For the Renishaw detector, this was done for the step scan and for the synchro scan. Secondly, the same single cell was measured with each of the detectors, each of the scan types and with a number of selected gains, keeping the total exposure time the same.

Figure 3.11 shows the electronic noise of each of the detectors with different scan settings (step/synchro scan) and different gain settings for the Andor detector. Some of the high-end gains for the Andor Camera showed strong noise components, but all showed an overall flat dark noise spectrum. When using high gain settings, this noise showed as positive only spikes on the otherwise approximately flat spectrum. When integrating the intensity and estimating

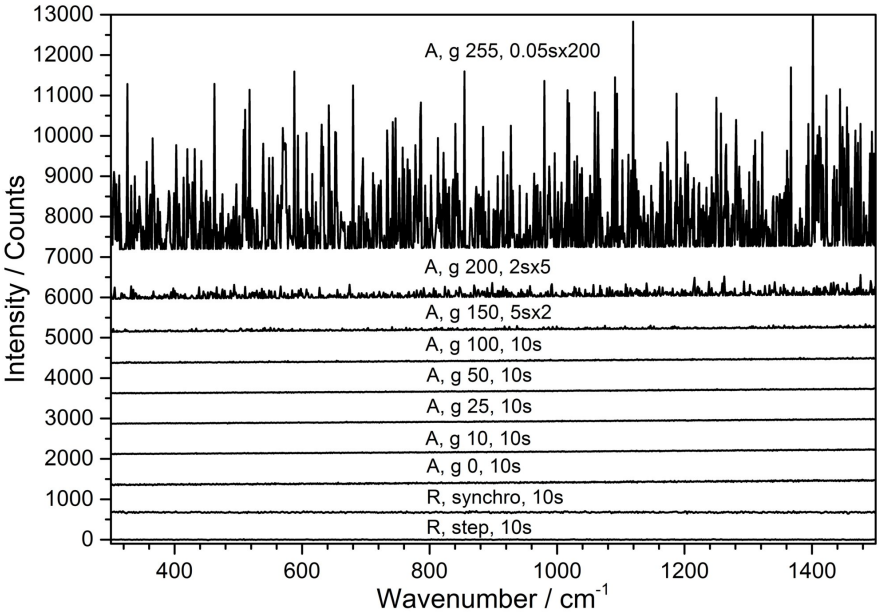


Figure 3.11: Dark noise experiment, where the total noise of the detector when in darkness was acquired. A stands for Andor, R stands for Renishaw and g stands for gain. For the Renishaw detector, both the step and synchro-scan were compared. For the Andor detector, a selection of gains was tested to see how much the noise increased when increasing the detector gain.

the noise as the standard deviation of the signal, the barplot shown in figure 3.12 is obtained. The Renishaw detector had very low electronic noise compared to the Andor detector, even at zero gain. The setting with the lowest dark noise was the Renishaw step acquisition. For the Andor detector, the noise was undetectable with this method below 150 gain, and spiked for values of gain of 200 or higher. The Andor detector also showed a baseline even for the zero gain values whereas the Renishaw detector has an approximately zero baseline.

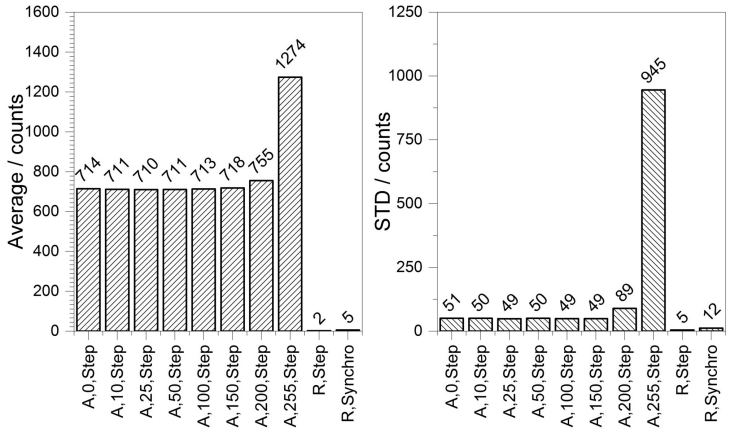


Figure 3.12: Average and standard deviation of the data from figure 3.11, where R stands for Renishaw, A stands for Andor, and the number and text next to the detector corresponds to the gain and scan type used.

However, the noise level is irrelevant if the acquired intensity of the sample is considerably enhanced. To test this, different settings were used to measure on the same live HT29 single cell. figure 3.13 shows the normalised data for each of the settings, where the noise intensity can be measured at wavenumber values $1850 - 2280\ \mathrm{cm}^{-1}$ as the standard deviation of the normalised data for three different areas, used to calculate the standard error of the noise.

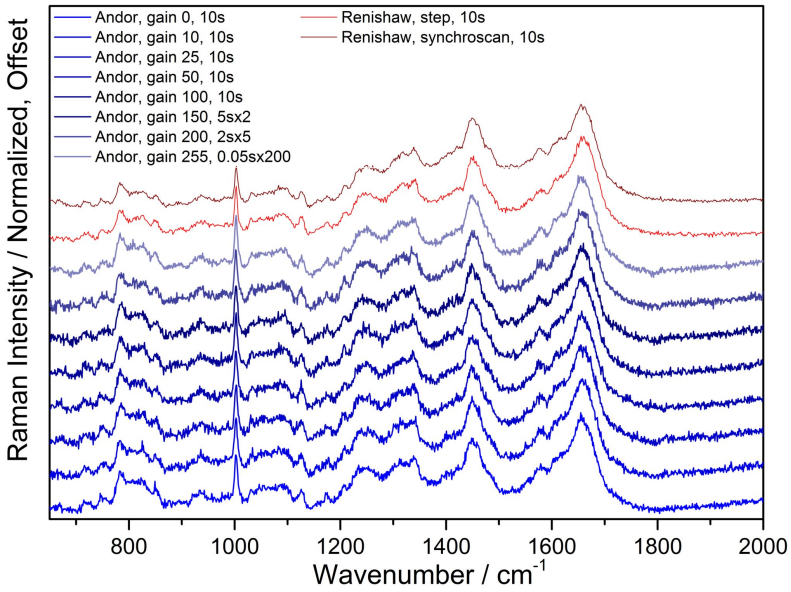


Figure 3.13: Subsequent measurements on the same HT29 live single cell using the 532 nm laser and the detector settings used for the data on figure 3.11. The data was baseline subtracted and normalised so the quality of the data can be appreciated independently of any zero baseline or any increase of intensity that proportionally increases the noise.

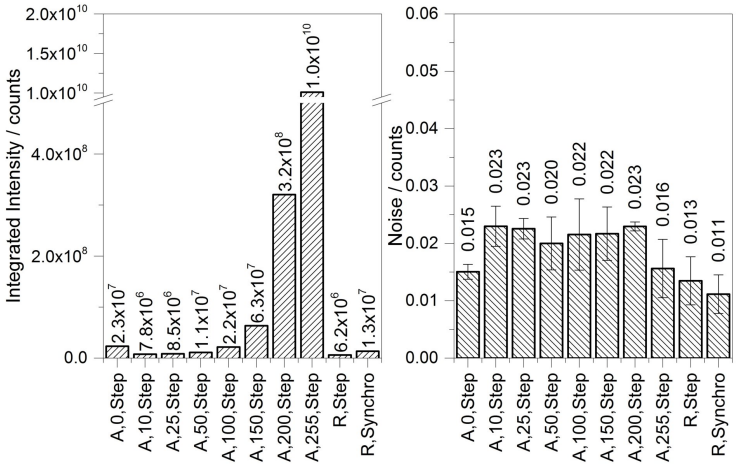


Figure 3.14: Bar plot of the (*left*) integrated intensity of the HT29 single cell baseline subtracted data shown in figure 3.13 before normalisation, showing the enhancement factors due to gain increase and (*right*) the noise of the silent region of the normalised data with standard errors. A stands for Andor, R for Renishaw, and the gain and scan types are stated for each measurement.

The signal intensity was calculated as the area under the curve after baseline subtraction using the function `trapz` (Matlab). The results are shown in figure 3.14, showing that even though the signal integrated intensity increases for higher gains, the noise of the normalised data does not significantly change between the Andor and the Renishaw detector, and the noise of the Andor detector increases when using higher gain settings. This showed that using higher gains even though it increases the overall signal intensity, is actually counterproductive, as it also increases the noise of the spectrum. Based on this, the Andor detector should always be used for the visible laser with no gain, as it is faster and the final signal quality is equivalent, and the Renishaw detector should always be used for the 785 nm laser as it is IR-enhanced.

3.2.6 Beam Expander

The laser spot can be imaged in the accessory camera of the microscope when measuring on a silicon sample. However, the gain of the camera and the intensity of the laser can give different laser spot apparent sizes. figure 3.15 shows the laser spot for 0.05% and 100% intensity, where the apparent laser spot size changes with the intensity. This could be explained by internal reflections of the microscope that are of weaker intensity than the

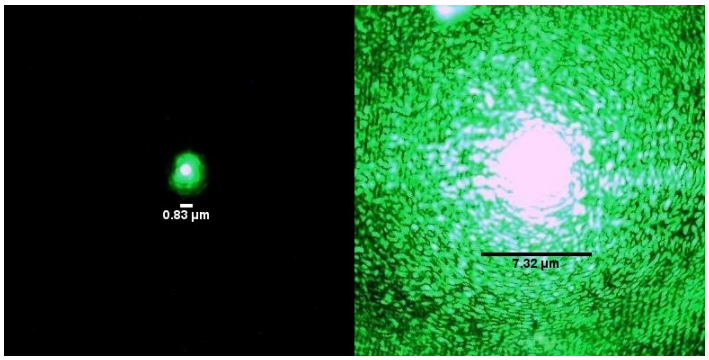


Figure 3.15: Image shows the 532 nm laser spot size onto a silicon sample when the image is acquired with the accessory camera using 0.05% laser intensity (left) or 100% laser intensity (right), with a 0% defocussed beam.

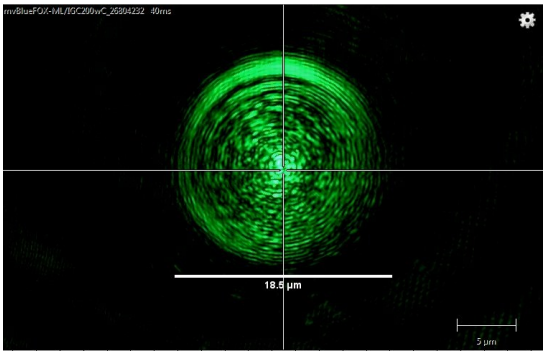


Figure 3.16: Beam expanded laser spot to 20% when measuring on an HT29 live cell with a width of ≈ 18.5 ¯m. The light seen comes from scattering from the cell and the close surface where the cell is attached, so the apparent beam expanded size can vary if a cell is smaller than the beam, as a lot of light is not reflected and the beam appears smaller.

centre of the laser and are filtered out by the ND filters. The camera saturates in the centre of the spot, so we can not obtain information about the laser light energy distribution by using the accessory camera.

The beam expander allows increasing the size of the laser beam. When working in a none confocal mode, this can be achieved by defocusing the objective, but the high confocality of the system means that changing the focus of the objective removes the data acquired from that region. When using the beam expander, this is solved as long as the beam fits through the confocality slit, so a beam bigger than 20 μm would be cropped when using high confocality, and values higher than 65 μm would be cropped when using the standard confocality. When using the beam expander, 0% indicates the laser is tightly focussed, while 100% indicates it is completely defocused by the beam expander. Defocusing reduces the power density at the sample, and so can reduce sample damage in sensitive samples, but also loses intensity and can reduce the signal at the sample.

Here the beam expander is used to obtain the average signal of the cell while still acquiring confocal data with minimum background contribution. When observing the beam shape on a silicon sample, 20% defocussed beam showed a diameter of ≈ 9.7 μm, whereas a 50% defocussed beam showed a diameter of ≈ 10.9 μm. Measuring the expanded beam size in the accessory camera in a silicon sample was challenging, as any focus drift or changes in the camera gain gave changes in the beam apparent size. When measuring on a single cell, both the 50% and 20% defocussed beams showed internal reflections on the whole cell area when looking at the images from the accessory camera. A picture of the 20% expanded beam when measuring on a single cell can be seen in figure 3.16 showing reflections over the whole cell could be observed in the accessory camera. First, the experiments were carried out using a 50% defocussed laser to ensure the whole cell area was covered. Further experiments were done using a 20% defocussed laser to increase signal even though some outer regions of the

cell could be lost by doing this. All experiments were always done by centring the cell in the field of view.

3.2.7 Test of the Phototoxicity of the 532 nm Excitation Laser

For an initial test of photo-toxicity of the visible laser light on a live cell, an HL60 cell was tested. Phototoxicity of the light depends on the wavelength, the intensity, the dose but also the cell line tested. In this test, pictures of the cell were taken between subsequent 1 s acquisitions of both the fingerprint and the CH stretching region of the spectrum, with the laser defocused to 50% as shown in figure 3.17. Given the 532 nm laser power at the sample is 22 mW, each acquisition was a total dose of 44 mJ. The cell was given doses up to 968 mJ with a total experiment duration of 8 min 27 s. No morphology changes were observed on the cell during this time. PCA showed no component with a significant trend with subsequent acquisitions. It was concluded that these doses were safe for the cells, and thus the optimised acquisition was set to 5 s per region, with a total dose of 220 mJ.

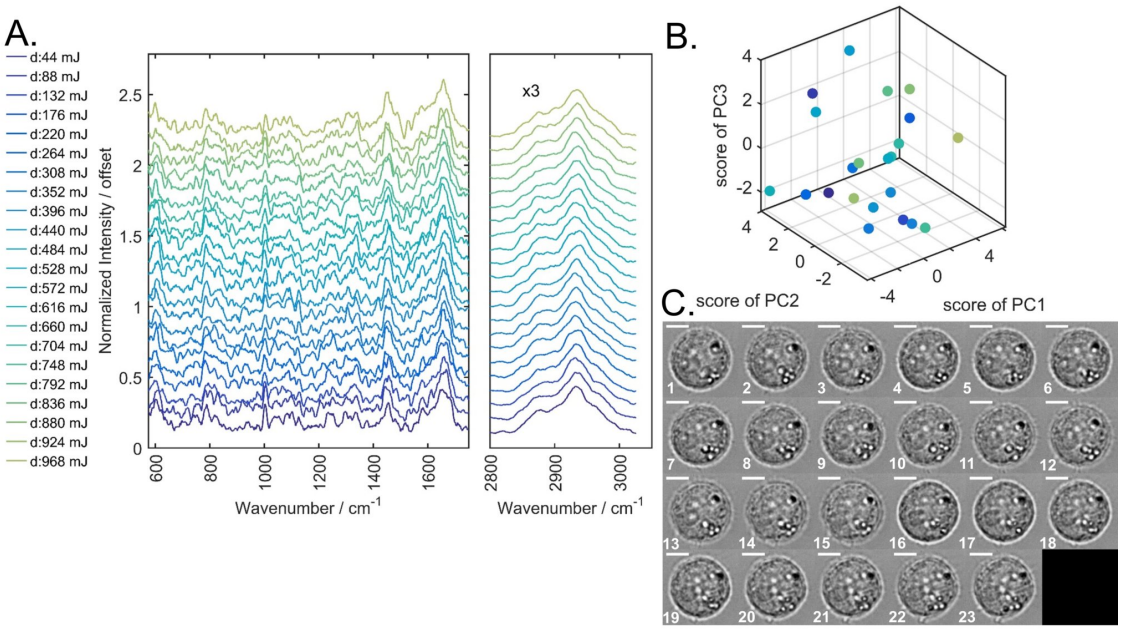


Figure 3.17: Visible laser dose test on HL60 cell. **A.** Time course of spectra on the cell showed no clear trend with the dose within the noise limitation. **B.** PCA on the spectra from A shows that the axis of maximum variability of the system did not correlate with the dose on the cell, where color legend from A is used in B. **C.** Images taken between spectra showed no major morphology changes.

3.2.8 Focus Drift

It was observed that, when mapping, the focus of the sample would vary over time, with a quick drift that made mapping difficult, especially when using higher magnification objectives with higher confocality. To characterise the drift, a constant acquisition was set after focusing the sample for around 1.5 h.

The top panel of figure 3.18 shows the results of this experiment, showing an oscillation with a period of around 11 min respect to an equilibrium position. The Renishaw stage was then

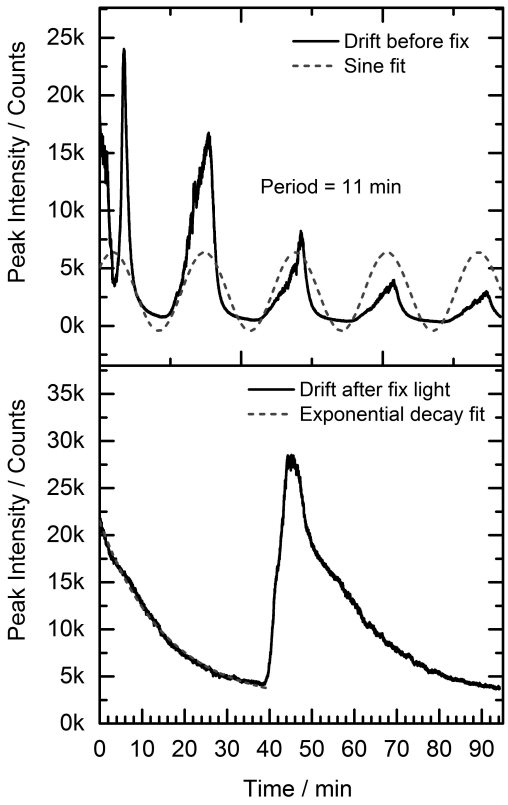


Figure 3.18: The focus drift was characterised by measuring in the same focal plan over time before and after fixing the stage, showing the changes on the focus drift in either case.

fixed by removing a plastic piece that was noted to be causing some stress on the stage, and the experiment was repeated, with results shown in the bottom panel of figure 3.18. Additionally, the aircon setup was revised to ensure that temperature changes were not causing the drift. The microscope still showed a drift with an exponential decay for around 40 min, followed by a strong secondary peak - likely an artefact from when the oil detached from the sample.

The size of the drift of the microscope after fixing the stage was estimated using the Depth/Intensity information from figure 3.6 (page 55) and the Time/Intensity data from figure 3.18 (bottom panel). The final relation obtained is plotted in figure 3.19, where the sample had a drift of around 4 μm per hour (when extrapolating from the 40 minutes drift data). This drift should allow mapping, given that the 100× oil with high confocality has a full-width half maximum of around 10.6 μm. However, long acquisitions in high confocality settings could mean losing the signal of the sample at some point during acquisition. This could be fixed in future acquisitions by implementing autofocus feedback.

3.3 Raman optimised Parameters

3.3.1 Raman Single-Cell Spectroscopy

For single cell spectroscopy, the aim was to obtain the average of a single cell in the minimum time possible with enough signal quality to be able to perform classification of different cell

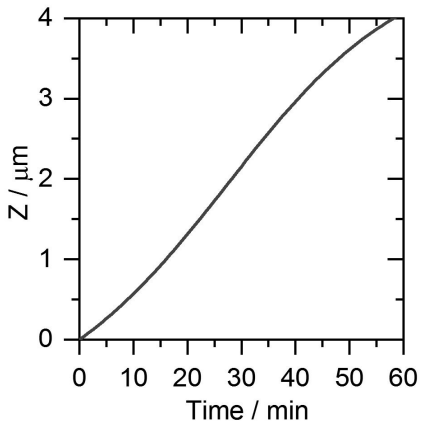


Figure 3.19: Estimation of the focus drift in the axial axis (Z) over time obtained from the data from figure 3.18 and the data from figure 3.6.

lines. For this, quartz slides (UQG Optics, 75x25x1 mm) and coverslips (25.4×25.4×0.15-0.25 mm Alfa Aesar) were sonicated with acetone (VWR Chemicals), 2-5% Decon 90 (VWR Chemicals) and rinsed with MilliQ. Hydrogen peroxide 30% (Thermo Fischer) and sulfuric acid >95% (Thermo Fischer) were mixed in a 3:7 proportion (Piranha solution) and used to clean the slides for 20 minutes. Slides and coverslips were stored in MilliQ and dried under a stream of nitrogen immediately before the experiment. Spacers were prepared using a 50 μm PET[1] film (Goodfellow, UK). A nitrocellulose-based solution was used to bond the coverslip to the slide and was dried at 80°C for 30 min.

The cell solution was pipetted into this chamber immediately before measuring. Alternatively, the samples were sealed with wax after pipetting the cell solution to avoid evaporation and measured straight away. All experiments were done at room temperature and samples were measured for 1-1.5 h. When pipetted into the setup, cells sedimented onto the coverslip and showed no visible Brownian motion, remaining in a spherical shape. The objective used for the cell measurements was initially the 63× oil immersion objective, and later the 100× oil immersion objective (see table 3.1 on page 51). Prior to every experiment a spectrum of a silicon sample was collected and the microscope was calibrated to the peak position (520.5 cm^{-1}).

The cell spectra were obtained using a slit size of either 20 μm (high confocality mode) of 65 μm (standard confocality mode). The laser spot was defocused by either 50% or 20% using a beam expander, generating a laser spot of approximately 20-18 μm diameter respectively. An example of acquisition settings would be the following: each cell spectrum was obtained using a step configuration with 1 s exposure time and 5 accumulations in two different windows (300 - 1800 cm^{-1} and 1800 - 3200 cm^{-1}) which gave a total exposure time of 10 s per cell. Background spectra from cell-free regions of the sample were measured at the same Z-position as the cells for each experiment for subsequent background subtraction. 50% defocus was used for all experiments in chapter 4 (page 91). For experiments in chapter 5 (page 111), 20% defocus was used.

Alternatively, in some experiments, the fingerprint region was prioritized to reduce the acquisition time between measurements. The confocality can also be changed to standard when the background contribution is easy to subtract like when using quartz. For this reason, in the time study of the UV activation of DC473 in chapter 5, the confocality was set to

[1] Polyethylene Terephtalate

standard and only the fingerprint region and part of the silent region of the spectrum was acquired.

3.3.2 Raman Mapping

For Raman mapping, the final goal is to obtain an image where the pixels have enough information to be classified in the different parts of the cell, whilst still taking as little time as possible. When mapping on fixed cells, the time constraint is reduced. However, the objective drift means that long maps will likely drift out of focus. When mapping live cells, both the mapping time and the light dose need to be carefully considered to avoid phototoxicity or changes in the cell during the mapping.

For DC473, it was found that mapping on fixed cells with the 785 nm laser gave the best results using the Renishaw detector. Prior to every experiment, a spectrum of a silicon sample was collected and the microscope was calibrated to the peak position (520.5 cm^{-1}). The cell spectra were obtained using the 100× oil immersion objective with a slit size of 20 μm. The step size was 0.7–1 μm and exposure time was 1.8–2 s/px.

Maps were cosmic ray filtered in WiRe software and exported as txt. for subsequent pre-processing and analysis in Matlab in-house scripts. Pre-processing was done in a similar fashion to the single cell data, creating map datasets that contained the positions and times for each pixel spectrum acquisition (`DatasetCreator_Maps_Example`). Maps were then analysed in a different Matlab script (`DatasetAnalysis_Maps_Example`), where PCA was performed in a similar manner to single cell data pca. Maps were created in Matlab by rearranging the spectra onto the map shape and using the Matlab function `imshow` with the standard grey colormap, where the contrast of the map was set so 1% of the pixels were saturated in each side of the colorbar, using the function `percentile` over the observations for each integrated intensity or loading scores. The scripts are available to download as supplementary material. Maps were exported from Matlab and loaded in ImageJ, where images were smoothed with a Gaussian filter of size half the pixel size, and a colormap of *fire* was used for all the Raman maps.

3.3.3 Raman Spectra from Standards

To obtain the Raman spectrum of 5-ALA[1], crystals from 5-ALA HCl (Sigma Aldrich) were placed onto a glass coverslip chamber (8 well chamber, LabTek) and measured using the 532 nm laser and the 40x objective, using intensity of 100% for 2 s and 15 accumulations and step scan. The spectrum was baseline subtracted using the function developed by Koch *et al.* [238] a smooth width of 30, a bandwidth of 350 and 20 iterations. For the Raman spectrum of PPIX[2] and cyt C[3], cyt C was bought lyophilized from Generon, Bio Basic (100 mg) and prepared as a stock solution in DPBS at a concentration of 200 μM. It was stored at -20°C and kept in ice and protected from light until measured. Disodium salt PPIX was obtained

[1] 5-aminolevulinic acid
[2] Protoporphyrin IX
[3] Cytochrome C

from Sigma Aldrich and stored at 4°C. It showed poor solubility in DMSO[1], chloroform or dichloromethane. A solution of 0.16 mg/mL was prepared in milliQ. Both compounds were measured with the 532 nm laser to achieve RR[2] in order to minimise the limit of detection and the 40× objective, in a glass coverslipped chamber (LabTek, 8 wells). To achieve high signal to noise spectra, the exposure time was 1000 s using 22 mW laser power for the cyt C. For the PPIX, the exposure time was 200 s using 22 mW. The DMSO used for the standard measurement was > 99% (Fischer Scientific), measuring with the 785 nm laser in a quartz slide/coverslip setup with 50 μm spacers using the 40× objective, standard confocality, 15 accumulations of 1 s exposure time and 75 mW laser power. For measuring DC473, DC271 and EC23, solutions 1 mM in DMSO were provided. Solutions were dried onto a quartz slide and Raman measurements were taken on visible crystals on the slide using the 40× objective with both the 532 and the 785 nm excitation lasers. DC473 crystals were measured with 25 accumulations of 1 s exposure time and standard confocality. DC271 and EC23 crystals were measured using 0.2 s exposure time and 50 accumulations.

3.3.4 Raman Spectra Pre-Processing

A raw single-cell Raman spectrum can have variable background contributions depending on the exact location of the cell within the measuring chamber, baseline due to variable autofluorescence and due to scattering, variable intensity depending on the cell size with respect to the acquisition volume and contributions of cosmic rays. For reliable analysis of data from multiple experiments, very careful pre-processing should be done of the spectra, using robust methods equally over all experiments. Here, all the pre-processing steps will be illustrated following an example spectrum of an HL60 cell excited with the 532 nm laser (1 s, 22 mW, 5 accumulations, step acquisition) with a beam expander focus of 50%. All these steps were done in a single Matlab script for dataset creation, that saved the pre-processed data for later quick loading and further analysis in a different Matlab script for analysis. Example scripts with all the steps for both the dataset creation and the analysis, and all the required functions for them are attached to the e-thesis as supplementary materials (`DatasetCreator_Example` and `DatasetAnalysis_Example`).

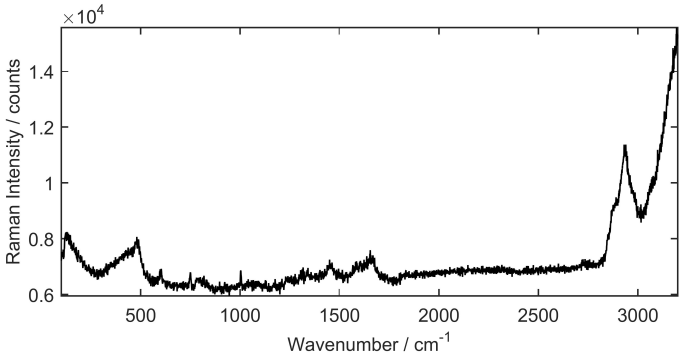


Figure 3.20: Example of a raw HL60 single-cell Raman spectrum excited with the 532 nm laser.

Cosmic rays were rarely present in single cell spectra but were more common in Raman maps. They were removed from the spectrum using the WiRe software cosmic ray removal option before exporting the spectra to txt for further analysis in Matlab. After importing the data

[1] Dimethyl sulfoxide
[2] Resonant Raman

in Matlab, the first step was to ensure that all spectra shared the same reference calibration and the same wavenumber vector. For each experiment, the grating motor was aligned to the silicon peak at 520.5 cm^{-1}, considering good any value between 520.3-520.7 cm^{-1}. The spectrum was interpolated to a reference wavenumber array with step 0.5 cm^{-1}, and the shift between the calibration silicon peak position and the reference position was used for perfect alignment. This is shown in the following Matlab code, where *k* corresponds to the spectrum index and *j* corresponds to the experiment index, for which the *CalibrationPeak* was recorded from the silicon sample. The final spectrum was interpolated to the chosen vector `Xintvalues`, ensuring that data from different experiments shared the same wavenumber values. Choosing this vector was also an opportunity to truncate the side regions of the spectrum that were considered to be most affected by the substrate background.

```
RamanSpectraInt{k}=spline(RamanSpectra{k}(:,1)+...
    (520.5-CalibrationPeak{j}),RamanSpectra{k}(:,2),Xintvalues);
```

An example of this is shown in figure 3.21, where the raw spectrum is shown as a reference, and the interpolated spectrum is shown with an offset for clarity. The shape of the spectrum is identical to the raw data, and, in this example, the calibration peak was at 520.4820 cm^{-1}, so the shift is of only 0.018 cm^{-1}.

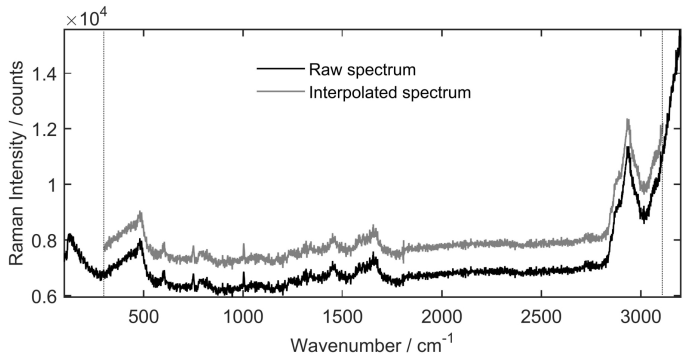


Figure 3.21: Example of a raw HL60 single cell spectrum interpolation and alignment. The interpolated spectrum has been offset for clarity.

For the next step, 5-6 backgrounds were obtained from each experiment from the same focal plane as the cells but a region with no visible cells. These backgrounds were interpolated and averaged, and the average was smoothed before subtraction using a Savitzki-Golay filter (function `sgolayfilt`), with polynomial order of 2 and frame of 12.5 cm^{-1}. This ensured that the minimum amount of noise was introduced when subtracting the background. For subtraction, the maximum between 540-720 was considered to obtain a multiplier for the background contribution for each cell. A constant baseline calculated as the minimum value of the smoothed spectrum was previously subtracted from the cell spectrum and from the background spectrum. This subtraction was done using the following example code. A better subtraction can be done by fitting the background to the cell spectra, implemented in Chapter 6 (page 6.3) that was developed for cases where the background is high compared to the cell signal, such as the PDMS[1] contribution when doing Raman *on chip* single-cell measurements. Background subtraction on the spectra from figure 3.21 can be found in figure 3.22.

[1] Polydimethylsiloxane

```
BackgroundAverage=mean(RamanBSpectraInt);
BackgroundAverageZero=sgolayfilt(BackgroundAverage-...
min(sgolayfilt(BackgroundAverage,2,25)),2,25);
IndexQuartz(1)=find(round(Xintvalues)==540,1)
IndexQuartz(2)=find(round(Xintvalues)==720,1)
IntPeakRef=max(BackgroundAverageZero(IndexQuartz(1):IndexQuartz(2)));
% Background subtracting the spectrum:
RamanSpectraZero{k}=RamanSpectraInt{k}-...
    min(sgolayfilt(RamanSpectraInt{k},2,25));
IntPeak=max(...
    sgolayfilt(RamanSpectraZero{k}(IndexQuartz(1):IndexQuartz(2)),3,15));
RatioSubs=IntPeak/IntPeakRef;
RamanSpectraQuartz{k}=RamanSpectraZero{k}-BackgroundAverageZero*RatioSubs;
```

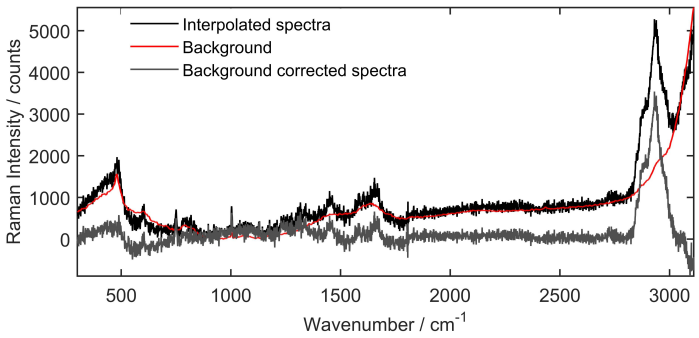


Figure 3.22: Example of background subtraction to a previously interpolated and aligned HL60 single cell spectrum. The final spectrum is shown offseted for clarity.

To minimise noise effects, Raman data is often smoothed. Here we used a Savitzki-Golay filter with polynomial order of 2 and frame of 8.5 cm^{-1}, as shown in the following code. An example of the effects of this filter on the data is shown in figure 3.23.

```
RamanSpectraSmooth{k}=sgolayfilt(RamanSpectraQuartz{k},2,17);
```

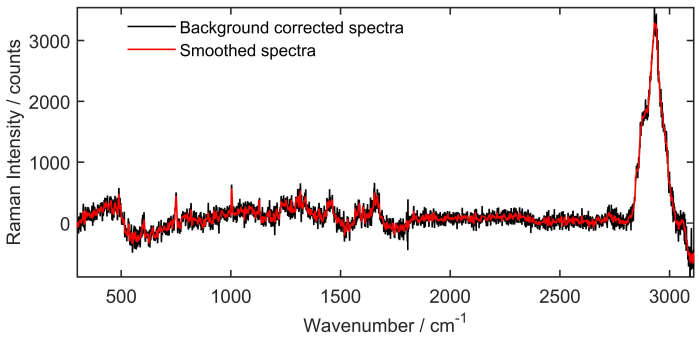


Figure 3.23: Example of smoothing on the HL60 single cell spectrum from figure 3.22.

Additionally, the contributions lower than 730 cm^{-1} were truncated from the spectrum as they were seen to contribute with background variability. The following step was baseline subtraction, using the algorithm developed by Koch *et al.* (`f_baseline_corr`, available in supplementary materials).[238] It was found that two subtractions – a more aggressive subtraction with lower frame size followed by a gentle subtraction with a broader frame size to remove the *zero*-baseline – gave the best results. An example of this subtraction is shown in figure 3.24, using the parameters from the following code:

```
[RamanSpectraB{i}, varargout{i}]=f_baseline_corr(XvaluesTrunc,...
    RamanSpectraTrunc{i},40,400,10);
[RamanSpectraB2{i}, varargout2{i}]=f_baseline_corr(XvaluesTrunc,...
    RamanSpectraB{i},60,800,3);
RamanSpectraBaseline{i}=RamanSpectraB2{i}.';
```

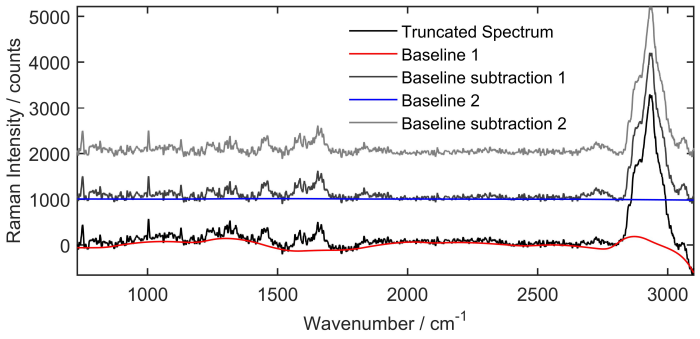


Figure 3.24: Example of baseline subtraction on the HL60 single cell spectrum from figure 3.23 using the algorithm developped by Koch *et al.*[238]

Additionally, it was found that using the silent region of the spectrum for classification could select variations related with the substrate baseline contribution, and thus the region between 1750-2800 cm^{-1} was removed from the spectrum (*zapped*) as shown in figure 3.25.

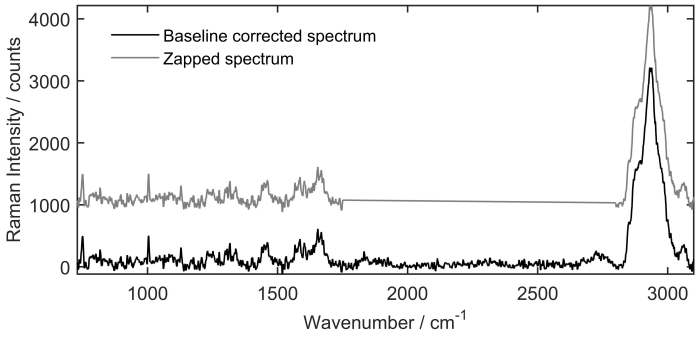


Figure 3.25: Example removal of the silent region on the HL60 single cell spectrum from figure 3.24, removing the region between 1750 and 2800 cm^{-1}.

To finalize, the spectrum was normalised so the Amide I had an intensity of 1. It was found that normalising to the maximum of the Amide I band (between 1600-1700 cm^{-1}) caused variability due to the noise of this band, and calculating the maximum using a spectrum with additional smoothing gave more robust results. The final normalised data is shown in figure 3.26, and was saved in a Matrix as a `.mat` file, imported for further analysis in an *Analysis* script.

```
x1=find(round(XvaluesTrunc2)==1600);
x2=find(round(XvaluesTrunc2)==1700);
RamanSpectraNorm{k}=RamanSpectraZap{k}/...
          max(sgolayfilt(RamanSpectraZap{k}(x1(1):x2(1)),2,31));
```

3.3.5 General Pre-Processing Final Settings

The spectra obtained were cosmic ray filtered (WiRE® software) and exported as text files for further analysis using Matlab's Statistics and Machine Learning Toolbox (MathWorks). The silicon peak of a calibration sample was used to calibrate the wavenumber axis of each

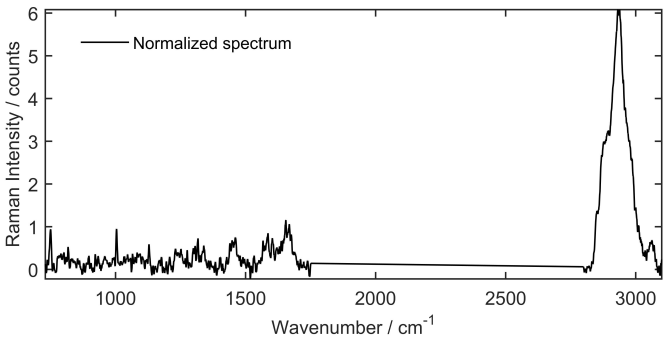


Figure 3.26: Example of a single cell spectrum normalised to the Amide I band.

spectrum and the spectra were translated vertically, such that the minimum intensity was zero. For each cell spectrum, the average background spectrum was multiplied by an adjustment factor before being subtracted from the cell spectrum to ensure the quartz band at around 480 cm^{-1} was fully corrected. In general, the spectrum was smoothed using a Savitzky-Golay filter and truncated to only consider the regions between 730 cm^{-1} and 3100 cm^{-1}. The spectrum was baseline corrected using the algorithm developed by Koch et al (2017). [238] The silent regions of the spectra between 1750 cm^{-1} and 2800 cm^{-1} were not considered for cell analysis except when stated otherwise (DC473 experiments). In order to normalise to the protein content, for comparison with biochemical literature data, each spectrum was normalised such that the Amide I peak was unity.

3.4 Raman Analysis

Statistical errors Unless stated otherwise, all values are expressed ± the standard error calculated as $\sigma/\sqrt{N}$, where σ is the standard deviation and N the sample size. Performance of the multivariate models was calculated as the accuracy of the model using 10-fold cross-validation with 5 repetitions.

Correlation Matrix The correlation matrix of all the pre-processed data was calculated to help with the peak assignment. The function used was `corrcoef`, that returns both the correlation matrix and the p-values matrix. To simplify the correlation image, points with p-values > 0.0001 were set to zero, and only the peaks that showed an absolute value of correlation greater than 0.3 were considered in the analysis.

PCA The edited data was truncated to 730-1750 cm^{-1} and 2800-3000 cm^{-1} and standardised using Standard Normal Variate, where each spectrum is normalised by subtracting the overall average and making its standard deviation of 1. This method of normalisation is useful when much of the signal in a sample is the same. After normalisation, PCA[1] was done using the function `pca`.

[1] Principal Component Analysis

LDA was performed with a limited number of PCs, following the 95% rule of truncating so at least 95% of the viability was explained. LDA[1] was done using the function `fitcdiscr` using a *'linear'* discriminant type and was performed onto the training data. The classifier obtained was then used to calculate the performance in the testing group. Training and testing groups were obtained using a random 10-folding of all the available data. More about k-fold validation is given in Chapter 2 in section 2.2.2.1 (page 38). It must be noted here, that LDA performed worse when the number of observations was close to the numbers of PCA loadings for training, and thus this must be taken into consideration when choosing the model – this can be simply fixed by adding more PCA loadings in these particular cases.

DT The function used was `fitctree` using the *exact* algorithm, that fits a binary DT[2] to the data. The classifier obtained was then used to calculate the performance in the testing group. Training and testing groups were obtained using a random 10-folding of all the available data.

C5.0 Fitting to the C5.0 trees was done by our collaborator Michael Lones from Heriot-Watt University. R's C5.0 package was used to train DT ensembles based on R. Quinlan algorithm, and the *caret* package was used to optimise training parameters. The classifier obtained was then used to calculate the performance in the testing group. Training and testing groups were obtained using a random 10-folding of all the available data.

SVM SVM[3] fitting was done by collaborator Michael Lones from Heriot-Watt University. R's `kernlab` package was used to train SVM models, and the *caret* package was used to select an optimal kernel function (from amongst linear, polynomial and Gaussian kernels). As all the tested kernels showed a similar performance, the linear kernel was selected.

PLSR function `plsregress` was used for the analysis. Scores in each of the components were compared in pairs using an unpaired two sample one-tailed t-tests, and the number of components was determined so cell lines showed a significant ($p > 0.01$) increase with the adenocarcinoma stage. Final considered values all showed at least $p < .001$. Duke's stages [B primary, C primary, C metastasis] were fitted as [1,2,3].

3.5 Gaussian-Lorentzian Sum Fitting of Individual Single-Cell Spectra

Raman bandwidths often follow Voigt profiles, that can be easily approximated by GLS[4] fittings (see section 2.1.4 on page 30). For fitting of hundreds of peaks on single cell spectra, the following code was implemented in Matlab, based on the GLS function from

[1] Linear Discriminant Analysis
[2] Discrimination Trees
[3] Support Vector Machines
[4] Gaussian-Lorentzian sum

Jain *et al.*, defined on equation (2.19) (page 31), rewritten here with the same nomeclature followed in the Matlab script, where a is the amplitude, σ (`sigma`) is the width, ν (`nu`) is the mixing parameter, x_C (`xc`) is the band position, and c_0 (`c0`) was introduced as an initial intensity baseline to be able to compensate for baseline subtraction artefacts in the fitting region.[213]

$$GLS(x, \sigma, x_C, a, \nu, c_0) = c_0 + (1-\nu) \cdot a \cdot e^{-4ln(2)\frac{(x-x_C)^2}{\sigma^2}} + \nu \cdot \frac{a}{1 + 4 \cdot \frac{(x-x_C)^2}{\sigma^2}} \tag{3.1}$$

The following Matlab code is an example used to fit the DC473 alkyne band at 2197 cm^{-1}, when fitting spectrum `i`:

```
GLS = @(a,xc,sigma,nu,c0,x) a.*...
    (...
    (((nu/sigma)*sqrt(log(2)/pi())).*...
    exp(-4*log(2).*((x-xc)./sigma).^2))+...
    (1-nu)./(pi().*sigma.*(1+4.*((x-xc)./sigma).^2))./...
    ((nu/sigma).*sqrt(log(2)/pi())+(1-nu)/(sigma*pi()))...
    +c0;

% Starting parameters
starta=0.20;     startnu=0.5;     startc0=0.02;

% Limits
fittinglimits=[2140 2275];
fittingIndex=[find(XvaluesTrunc2==fittinglimits(1))...
    find(XvaluesTrunc2==fittinglimits(2))];

fitoptionGLS=fitoptions(GLS);

peakposition(1)=2194;     peakvariation=10;
peakwidth(1)=20;       widthvariation=10;

% Variation of the parameters
fitoptionGLS.Lower=...
    [0, peakposition(1)-peakvariation, peakwidth(1)-widthvariation,...
    nuvalue, -inf];
fitoptionGLS.Upper=...
    [inf, peakposition(1)+peakvariation, peakwidth(1)+widthvariation,...
    nuvalue, inf];
fitoptionGLS.StartPoint=...
    [starta, peakposition(1), peakwidth(1),...
    startnu, startc0];

% Fitting
fitparm{i}=fit(Wavenumber(fittingIndex(1):fittingIndex(2)).',...
    DataMatrix(fittingIndex(1):fittingIndex(2),i),...
    GLS,fitoptionGLS);
```

Multipeak fitting can be done by simply defining a fitting function composed from multiple

GLS functions, as it was done to fit the phenyl DC473 band that falls on the side of the Amide I, where 5 GLS bands were used:

```
GLSmodel = @(a,xc,sigma,a2,xc2,sigma2,a3,xc3,sigma3,...
        a4,xc4,sigma4,a5,xc5,sigma5,nu,c0,x) ...
        GLS(a,xc,sigma,nu,c0,x)+GLS(a2,xc2,sigma2,nu,c0,x)+...
        GLS(a3,xc3,sigma3,nu,c0,x)+GLS(a4,xc4,sigma4,nu,c0,x)+...
        GLS(a5,xc5,sigma5,nu,c0,x);
```

This approach was used for analysing the single-cell data from Chapter 5. The fittings shown in Chapter 4 were done on the average spectrum and not on the single cell spectrum, and were done using Gaussians in Origin software.

3.6 Cell Culturing

The cell lines cultured for this project were SW480 (ECACC 87092801 or ATCC CCL 228), SW620 (ECACC 87051203 or ATCC CCL 227), HT29 (ECACC 91072201 or ATCC HTB-38), HTC116 (ECACC 91091005 or ATCC CCL 247) and HL60 (ECACC 98070106). All cell line manipulation was done under sterile conditions in a class II biohood unless stated otherwise.

3.6.1 Adherent cell lines

Cells were gently provided by PhD student Catriona Marshall (St. James University Hospital) after STR profiling and mycoplasma testing. SW480, SW620, HT29 and HCT116 cell lines were grown in DMEM[1] (Gibco) 10% FBS[2] (Sigma) 2 mM Glutamax (Thermo Fischer Scientific) and penicillin 100 units/mL streptomycin 100 μg/mL (Sigma), in a 37°C, 5% CO_2 and high humidity environment. Cells were split when reaching 80% confluency (approx. every 5 days). Cell splitting was done by first gently wash with sterile DPBS[3], covering with TrypLE (ThermoFischer) and 5 min incubation and 1/6 to 1/10 seeding in a new flask (T25 or T75) with 5-10 mL of new media. The cell lines were frozen for long term storage by concentrating 2-20 million cells in a cryovial (Fisherbrand) with complete media and 10% sterile DMSO[4] (Thermo Fischer). The cryovials were slowly frozen in a Mr Frosty container filled with IPA[5] at −80°C and then stored in liquid Nitrogen containers. Cells were never used past passage number 50. Cells were seeded 2 days in advance of an experiment at a concentration of 3.5×10^5 cells/well in a 6 well plate, and were detached by firstly washing with DPBS and incubating with TrypLE for 5 min. Cells were then washed by centrifugation 3 times in DPBS, and resuspended in Cell Dissociation Buffer.

[1] Dulbecco's Modified Eagle Medium
[2] Foetal Bovine Serum
[3] Dulbecco's Phosphate-Buffered Saline
[4] Dimethyl sulfoxide
[5] Isopropanol

3.6.2 Non-adherent cell lines

The HL60 cell line is non-adherent and stable in solution and does not require anti-clumping agents. HL60 can also differentiate when induced with compounds such as butyrate, hypoxanthine, phorbol myristic acid (PMA, TPA), DMSO (1% to 1.5%), actinomycin D, and retinoic acid. The HL60 cell line was purchased frozen and resuscitated by fast thawing at 37°C, adding it drop by drop to 4mL of preheated RPMI 1460 20% FBS 2 mM Glutamine. The cells were immediately centrifuged at 100gs for 4 min to avoid the toxic/differentiating effects of DMSO. Fresh media was added to the pellet to a final cell concentration of 3.5×10^5 cells/mL. Viability and concentration were monitored for the next 7 days by cell counting in an improved Neubauer chamber (using a 1/2 dilution with trypan blue, Sigma), where final viability was above 90%. Viability and concentration were monitored during the culture, keeping the concentration between 3×10^5 and 10^6 cells/mL and viability always above 95%. Cell splitting was done every 2-3 days to a final concentration of $1\text{-}3\times10^5$ cells/mL (1/5-1/10 split). For measuring, cells were retrieved and washed 3 times by centrifugation, resuspending in DPBS. Cells were frozen (10-20 million cells per cryovial) in 90% FBS 10% DMSO using a Mr Frosty container filled with IPA at −80°C and then stored in liquid Nitrogen containers.

3.6.3 Model System Cells Staining and Imaging

Cells were grown in NuncTMLabTekTM 8 well chambers (155409) with a well area of 0.7 cm^2 and coverslip bottom. Wells were Laminin coated, with an initial pre-coat with pre-filtered Poly-L-Lys (Sigma) 100 $\mu g/mL$ (stock concentration 0.01% = 100 μg/mL). Poly-Lys was first incubated for 30 min (37°C), washed with sterile MilliQ and allowed to dry. Laminin (Engelbreth Holm-Swarm Murine 1-2 mg/mL, 100x) was thawed in ice to avoid gel formation and a 1/1000 dilution was prepared in DPBS. The solution was added to the wells (1.4-2.9 $\mu g/cm^2$), that were covered with parafilm to avoid evaporation and left overnight at 37°C, and removed prior to cell seeding. Cells were seeded at concentrations of around 16000-19000 cells/well, with viabilities >97.6% assessed with Trypan blue assay. Cells were grown for 2 days, washed twice with DPBS and fixed with PFA[1] 4% for 10 min. PFA was then washed with DPBS and Hoechst 33342 (New England BioLabs) at a final concentration of 100 $\mu g/mL$ and Nile Red (Sigma-Aldrich, >98.0%) at a final concentration of 100 $\mu g/mL$ were added. Images were acquired using sequential scanning, with 488 nm excitation for Nile Red and two emission windows: 500-600 nm (green) and 600-700 nm (red) for triglycerides and phospholipids respectively. Hoechst 33342 was excited at 405 nm and emission was collected at 410-500 nm. Hoechst 33342 staining was also used for regular mycoplasma testing, as the mycoplasma are undetectable in bright field images but show as blue spots on the cell surface when Hoechst stained. No cultures showed any presence of mycoplasma under this staining.

[1] Paraformaldehyde

3.6.4 Incubation with Photosensitisers

5-ALA to PPIX experiments Cells were seeded 2 days in advance at a concentration of 35k cells per well in 6 well plates. 5-ALA[1] solution of either 100 mM or 200 mM were prepared as stock the morning of the experiment in DPBS from 5-aminolevulinic acid hydrochloride (Sigma Aldrich, >96%, powder) that was immediately filtered with a 200 nm filter and added to the cells in fresh complete media. 5-ALA has a half-life at pH 7 of 3h [239] so preparation of the stock immediately before each experiment is essential. Control cells were then DPBS washed and detached for Raman measurements using TrypLE while 5-ALA cells were incubated for 4 h. After 4 h, 5-ALA cells were DPBS washed and detached for Raman measurements. All cells were centrifuged twice and washed with DPBS to remove any excreted PPIX[2] from the surrounding media.

DC473 experiments Similar to 5-ALA experiments, 35k cells per well were seeded in a 6 well plate 2 days in advance. 4 μM DC473 was added to the cells and an equivalent volume of DMSO was added to the control (0.4% DMSO). Control cells were measured first while incubation of the treated cells with DC473 was taking place for 4 h, and then treated cells were measured. Cells were washed with DPBS and detached using TrypLE, and then were centrifuged twice and washed with DPBS, ensuring removal of all remaining DC473 not uptaken by the cells.

Activation of DC473 Cells were incubated with 4 μM DC473 as indicated previously and retrieved detached to measure on a quartz slide-coverslip setup. Every single cell was measured every 15 s for 15 acquisitions using the 532 nm laser line with an intensity of 22 mW and laser defocused at 20%, and acquiring only the fingerprint region for 5 s. For those samples where UV light was used, the cell was exposed to the light (360/40 nm centre/bandwidth) using 1 degree aperture with the 100× oil objective for 3 s (711 mJ/cm^2).

3.7 Rapid Prototyping Facilities

A clean room-*like* environment was available for rapid prototyping and soft lithography fabrication. The room has a double door and requires the use of hair nets, lab coats and plastic covers for shoes to access, with adhesive films for shoes' particles in the double entrance and frequent vacuum cleaning of the lab. All lights are filtered with the blue components removed to avoid cross-linking of the photoresists when manipulating them in the room. The main instruments available in the room are shown in figure 3.27. It includes an EMS 6000 Photo Resist Spinner (Electronic Micro Systems Ltd) with speeds up to 10,000 rpm with a remote control unit and capable of fitting samples from $\approx 1\ cm$ size to up to 10 cm. It works with a vacuum connection that maintains the sample in place during the spinning. Two hot plates EMS model 1000-2 ($150 \times 150\times < 25mm$) are available, with a $\pm 1\%$ temperature accuracy across the working surface.

[1] 5-aminolevulinic acid
[2] Protoporphyrin IX

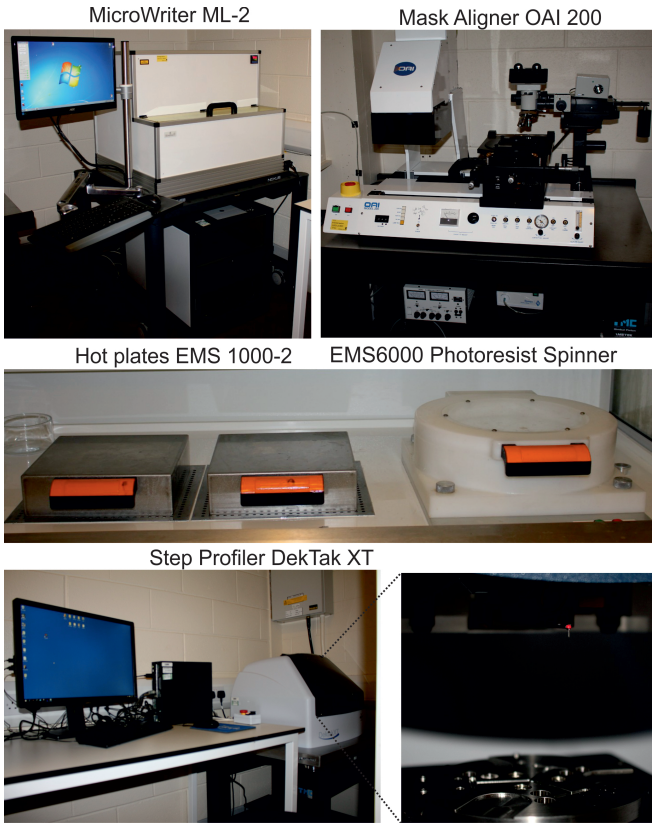


Figure 3.27: Main facilities in the Rapid Prototyping room: direct writing system, mask aligner, hot plates and spin coater and step profiler.

For rapid prototyping, two options are available: a direct laser writing system, where a pattern is programmed from a file and scanned along the surface, or mask alignment systems, where a mask is required and the system projects the mask onto the sample with a very accurate dose. For this project, the main system used was the direct writing system, a Micro Writer ML2 (Durham Magneto Optics) with both 405 nm lasers and a 375 nm laser available. The 405 nm lasers are available on different sizes, from 5-0.6 μm size, with a turbo option where multiple lasers are used at the same time to expose a larger area. The 375 nm laser is only available with a 2 μm size. However, different quality settings are available where the same size laser can be focussed or defocussed for coarser or finer results depending on the pattern. The writing system renders the designs into a grid of pixels to expose, and there are 4 quality settings available depending on the sizes of the pixels compared to the laser size: *draft* (pixel size larger than the laser size), *normal* (pixel size equals the laser size), and *high* and *super* (pixel size smaller than the laser size). Pixels smaller than the laser size give smoother angles or curved edges but take longer to expose.

3.8 Microfluidic Designs and Fabrication

3.8.1 Microfluidic Devices Fabrication

Devices in this project were fabricated in PDMS[1]. In order to generate the final PDMS device, a "master" is first prepared. Multiple photoresists are available in the market, that can be either positive or negative, as explained in figure 3.29A. The negative photoresist becomes

[1] Polydimethylsiloxane

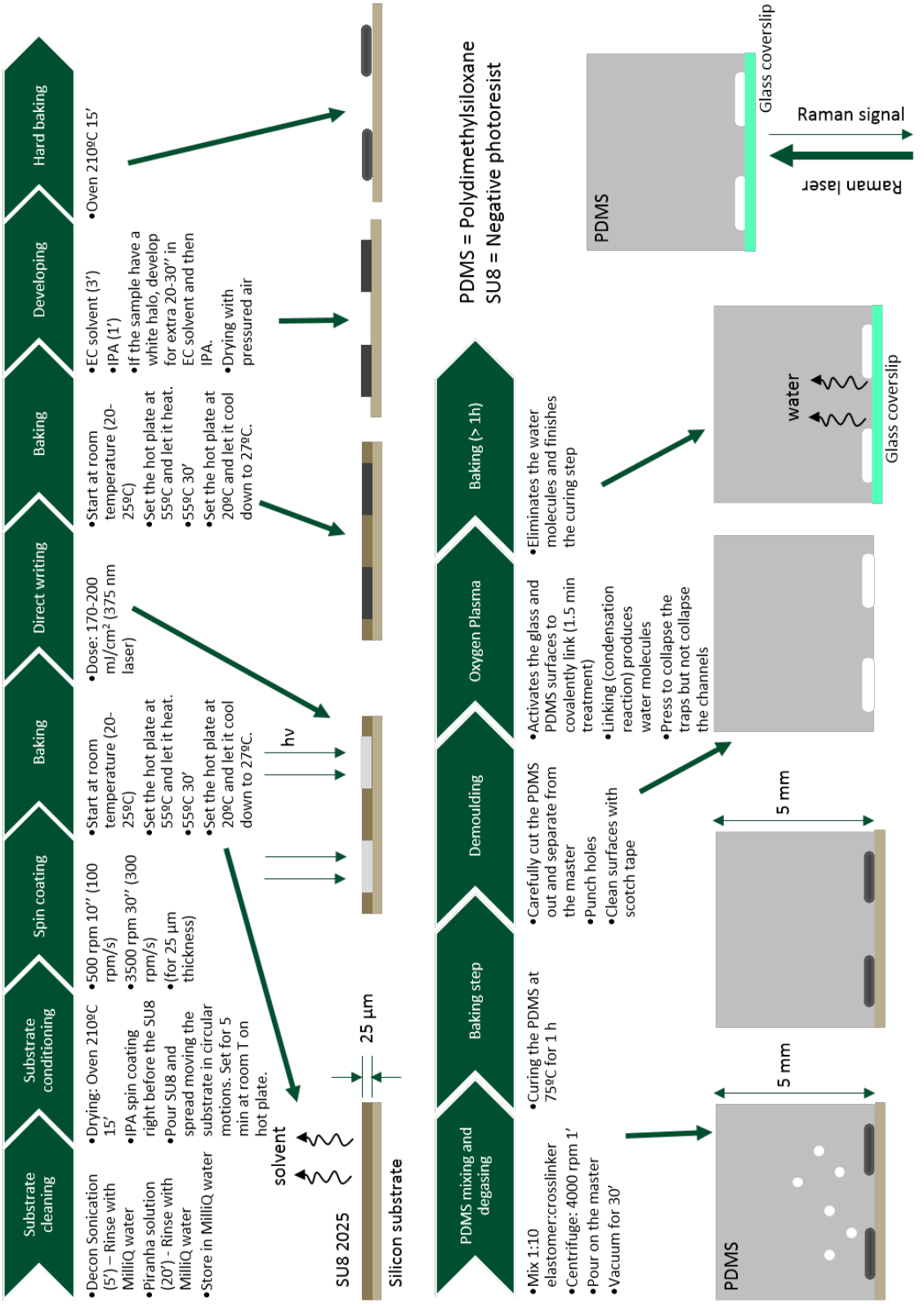


Figure 3.28: Summary of the protocol for the preparation of the SU8 2025 master and the PDMS-glass microfluidic devices.

polymerized by laser exposure and baking, whereas the positive photoresist becomes soluble by laser exposure. As the desired chamber depth was between 20-25 μm, it was chosen to use the negative photoresist SU8-2025 (Microchem).

Substrates can be in general either glass or silicon. As silicon showed a better adherence, the photoresist was spin coated on a P doped Silicon wafer (PI-KEM Ltd, reclaimed, 3 inches). Figure 3.28 shows the detailed protocol followed for the fabrication of microfluidic devices. The silicon wafer was sonicated in Decon for 5 minutes and Piranha cleaned (2:1 solution of 95% H_2SO_4 and 30% H_2O_2) at $> 80°C$ for 20 minutes, stored in MilliQ water until usage. Before usage, the substrate was oven dried at 210°C for 20 minutes.

Quick spin coating of IPA[1] onto the silicon before pouring the SU8 also caused increased adherence. When pouring the SU8 prior to spin coating, it was seen that the area of first contact of the SU8 to the wafer always showed a step compared to the surrounding area of the wafer. It was found that this was avoided if the SU8 was pre-spread onto the wafer before spin coating. For this, pre-warmed SU8 to room temperature was poured and manually spread over the wafer by tilting it and letting it flow towards the edges, letting it sit in a room temperature plate for 4-5 min before spin coating. The spinning conditions specified in the SU8 2025 datasheet were used to obtain the desired thickness: first 500 rpm for 10s using an acceleration of 100 rpm/s for spreading, followed by 3500 rpm for 30 s using an acceleration of 300 rpm/s for a 25μm thickness. However, over time, the SU8 thickens, and the spin coating setting had to be adjusted to obtain the same SU8 thickness.

Both before direct writing and after, a *soft-baking* step and a *post-exposure baking* step were carried out. According to the SU82025 datasheet, this should be done at 65°C (3-6 min) and at 95°C (3-6 min). However, quick changes in temperature were found to cause cracks and loss of adherence in the sample, critical for small structures and thin channels. Figures 3.29B, C and D show the main issues found while optimising the protocol that were cracks, ripples and lack of adherence. All of them were linked to either the use of low doses or drastic changes of temperature. Substituting the baking steps from the datasheet by longer baking steps at lower temperatures (55°C for 30 min) minimised these effects (Suggestion by Dr. Ben Johnson). The samples were heated and cooled gradually by placing the sample with the hot plate off and letting it reach 55°C (around 30 min). After a temperature of 55°C was reached, the sample was left for 30 min and then the hot plate was set to room temperature and cooled down slowly (around 1.5 h). Optimal baking left the SU8 layer dry of solvent, and thicker SU8 samples could require a 1 h soft baking step instead of 30 min. This temperature ramping was used for both the *soft baking* and the *post-exposure baking*.

Two methods are available for exposing the SU8: either using a mask, that allows exposure of the whole wafer at the same time or using a direct writing system, that moves lasers in a raster fashion to expose the sample. Masks are expensive but can be used many times with the same design. It was decided to prepare the wafers using the direct writing system while optimising the microfluidic designs and to create masks when the designs were definitive.

The dose was optimised for each design. Lower doses were found to obtain a better resolution and development of the traps, but lower adhesion. It was decided to use a dose of

[1] Isopropanol

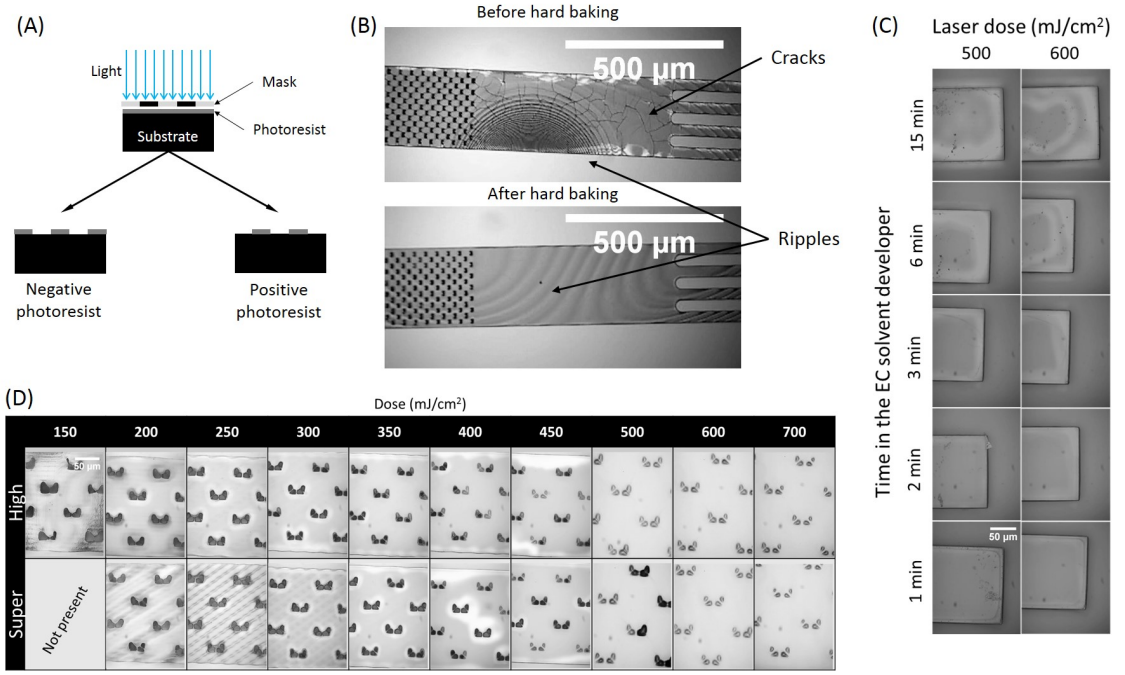


Figure 3.29: (A) Main difference between a positive and a negative photoresist when using a Mask Aligner. **(B)** Examples of the main issues found when optimising the rapid prototyping protocol. Hard baking was found to reduce the cracks whereas ripples were attributed to drastic changes of temperature, and were reduced using longer baking steps at a lower temperature. **(C)** optimisation results for EC solvent development showed that more than 3 minutes in the solvent reduced the adhesion of the photoresist. **(D)** Example of the dose and *quality* optimisation for one of the early cell trap designs, where the traps are holes in the SU8 that appeared in black. Lower exposed SU8 tended to show poorer adhesion, whereas overexposed SU8 did not resolve the trap. *Quality* values compared are *high* (lateral resolution between 2-1.8 μm) and *super* (lateral resolution of 1.7-1.6 μm).

400 mJ/cm^2 for wider channels of the chip, and a lower dose of 170 – 180 mJ/cm^2 for the trapping area, that was kept in place by the over-exposed remaining structure. See Figure 3.29D for an example of a dose test for traps. The *quality* used was *super* for the traps (lateral resolution of 1.7-1.6 μm) and *fastest* for other structures (lateral resolution >2 μm). Developing was done using EC solvent (Microposit) for 3 minutes and 2-Propanol Analytic for 1 min, drying with pressurised air. If the IPA revealed a white crust onto the devices, this was a sign of longer developing required, and 3 min were added to the developing step until complete development was achieved. Using a spray was found to help ease the developer into gaps of small structures such as microfluidic traps, removing the unexposed SU8. Hard baking was found to reduce the cracks, but can also round the SU8 corners, and so was only done when cracks were present (15 minutes at 210°C).

PDMS (Sylgard 184 Silicon Elastomer, Sigma) was prepared by mixing 1:10 polymer:crosslinker. Degassing was done by centrifugation (4000 rpm 1 minute) and applying a soft vacuum on the sample (30 min - 1h). The PDMS layer was approximately 3-5 mm thick. A baking step was done at 75°C for 1 h. The PDMS cast was removed and holes were punched. PDMS was covalently joined to coverslips (cleaned by Decon sonication for 5 minutes) by Oxygen Plasma treatment for 1.5 minutes at 0.3-0.5 mbar in a Diener Oxygen Plasma system. Gentle pressure was applied, especially in the trapping area, to ensure the

contact of the trap structures with the glass. An additional baking step (75°C for 1 h) was used to eliminate the water generated in the covalent bonding.

For the fabrication of channels of different heights, SU8 2075 was also used changing the spin coating settings according to the product datasheet. For thicker devices, both the soft bake and post-exposure bake were done for 1 h to ensure complete removal of the solvent. Developing was also done for longer until the IPA washed did not show a white crust on the surface of the photoresist.

3.8.2 Microfluidic Experimental Procedures

Before use, the devices were washed with ≈ 1 mL Ethanol 70% as a wetting agent and as a sterilization mechanism, followed by ≈ 1 mL of MilliQ. The designs incorporated built-in filters to avoid blockages. Fine bore polythene tubing (Smiths MedicalTM, Inner diameter of 0.38 mm) was used to connect the inlet and outlet holes with the syringes (1 mL, SGE gas-tight syringe, Sigma, Section diameter of 4.6 mm), that were previously washed by alternating Ethanol 70% and MilliQ (total of 3 mL each). The pumps used were PHD Ultra Syringe Pumps (Infuse/Withdraw Programmable, Harvard Apparatus).

3.8.3 PDMS Metal Coating Experiments

The PDMS was prepared as indicated in section 3.8.1, but before binding with the glass coverslip the metal (either Aluminium or Chromium and Gold) was evaporated onto the PDMS using an Edwards AUTO 306 Turbo pumped evaporator. A high vacuum of approximately 10^{-6} bar was reached before evaporating the metal. In the first case, 50 nm of Aluminium was evaporated on the PDMS, and in the second case 5 nm of Chromium was evaporated first to increase the Gold adherence to the PDMS, and then 50 nm of Gold was evaporated on top. The regions around the chip were either rubbed cleaned or cleaned with Scotch tape to remove the metal coating in those areas and promote their binding to the glass coverslip, using oxygen plasma.

3.8.4 Trap Dimensions Optimisation

By using the equations (2.40) (on page 47) introduced in chapter 2, the resistance of a microfluidic channel can be calculated. When a microfluidic device has multiple possible paths for the flow, the ratio of the resistances of each of the paths allows calculating the ratio of expected flows between the paths. This is particularly useful when designing microfluidic traps.

When considering a trap array, such as the one shown in figure 3.30, showing Di Carlo-like traps design [240], objects will only go through the trap instead of around it if the flow through the trap is high enough. There is thus a minimum value of R_1/R_2 needed for object trapping. Previous work in optimisation considered the trapping condition as $R_1/R_2 < 1$, however, particles in solution have inertia and will not simply follow flow lines, making trapping slightly easier.

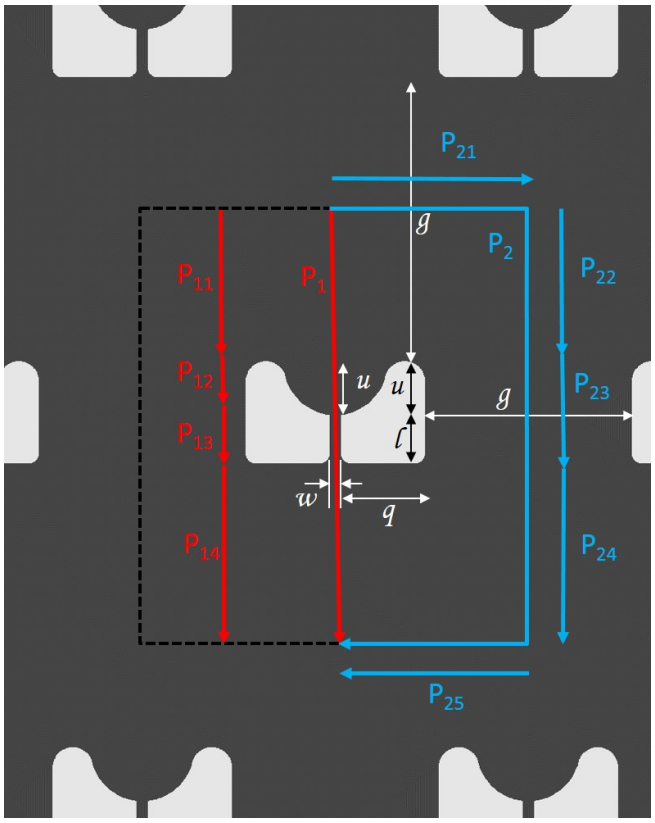


Figure 3.30: Di Carlo-like [240] trap schematics used to show the different parameters of trap design optimisation. In red, the path through the trap P_1 divided in 4 different subpaths. In blue, the path around the trap P_2, divided in 5 different subpaths.

Each of the paths will have a different flow Q_1 or Q_2 and pressure P_1 and P_2. We can thus consider the resistances following the pressure nomenclature from figure 3.30:

$$\begin{aligned} R_1 &= R_{11} + R_{12} + R_{13} + R_{14} = 2R_{11} + R_{12} + R_{13} \\ R_2 &= R_{21} + R_{22} + R_{23} + R_{24} + R_{25} = 2R_{21} + 2R_{22} + R_{23} \end{aligned} \tag{3.2}$$

However, for R_{11} and R_{22}, the width of the channel is the width of the whole chamber, making their contribution negligible. This simplifies each resistance as follows:

$$R_1 = R_{12} + R_{13} \qquad R_2 = 2R_{21} + R_{23} \tag{3.3}$$

For the particular case of R_{12}, the circular shape of the trap complicates the optimisation as the resistance would follow the equation:

$$R_{12} = \frac{\mu}{8} \int_0^{w'(x)=w} \frac{f\left(min[\frac{w'(x)}{h}, \frac{h}{w'(x)}]\right)(w'(x)+h)^2)}{(w'(x)\cdot h)^3} dx \tag{3.4}$$

where $w'(x)$ is the width dependence with distance along the circumference, i.e. the length of the cord in a circle with radius r from the centre to w along the trap channel coordinate x: $2r\sqrt{1-(x/r)^2}$. From that, the expression of R_{12} can be calculated as shown in equation 3.8.4.

$$R_{12} = \frac{\mu}{8} \int_0^{\sqrt{1-\left(\frac{w}{2r}\right)^2}} \frac{f\left(\frac{2r\sqrt{1-\left(\frac{x}{r}\right)^2}}{h}\right)\left(\frac{2r\sqrt{1-\left(\frac{x}{r}\right)^2}}{h}+h\right)^2}{\left(\frac{2rh\sqrt{1-\left(\frac{x}{r}\right)^2}}{h}\right)^3} dx \tag{3.5}$$

The rest of the resistances will have a general expression based on equation (2.40). An optimal value for the trap parameters, ensuring the right resistance ratio and minimising the repetition unit area to achieve the maximum number of traps per unit area can be found, with repetition unit area $A = (l + u + g)(2q + w + g)$. A number of constraints to achieve cell trapping need to be applied:

- Correct ratio of $R_1/R_2 < B$.
- Small enough channel to avoid the cell deforming through and leaving the trap: $w < \alpha u$, where α is a percentage of the total cell size.
- Distances between traps need to be big enough to avoid clogging: $g > \beta u$
- Trap area needs to be large enough for easy fabrication: $ql + qu - (\pi u^2)/4 > \gamma$

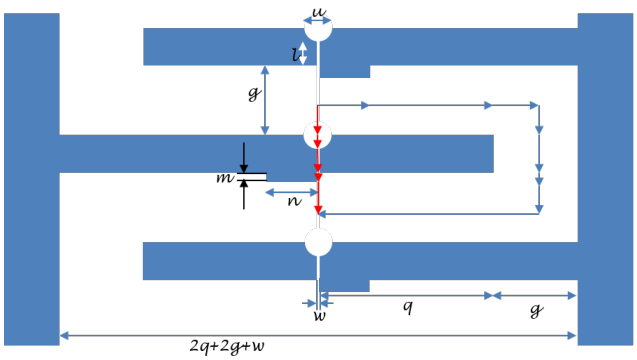
The optimisation algorithm is written in equation (3.6) in terms of the parameters, where the values for B, α, β and γ are estimated based on the system and experimental evidence, the values for u and h are chosen and the algorithm outputs the values for l, w, q and g:

$$min(A) \quad \text{assuming} \quad \left\{\frac{R_1}{R_2} < B;\ w < \alpha u;\ g > \beta u;\ ql + qu - \frac{\pi u^2}{4} > \gamma\right\} \tag{3.6}$$

3.8.5 Takeuchi-like Trap Dimensions Optimisation

An alternative trap design is the Takeuchi-like trap [241]. These traps have lower packing density, but can have really high trapping efficiency and make keeping track of the same cell easier over time, as the cells are aligned in a row. Here, the same optimisation approach was applied to this design, as shown in figure 3.31. Figure 3.32 shows a schematic of the trapping and untrapping capabilities of these traps, allowing retrieving the cells in order – so in the future, these cells could be diverted to a different single cell analysis technique for further interrogation. To be able to untrap the cells by reversing the flow without getting cells trapped on the side opposite to the trap, an $m \times n$ structure was designed next to the trapping channel. The same logic followed above can be applied to this design, where the resistance of each path is calculated based on the width and the length, except the path through the trap, that has the same expression from equation (3.8.4) (page 83).

It was found that these traps had very good efficiency, although the trap density is significantly lower than the Di Carlo design traps. An example of a design that was fabricated is shown in figure 3.33, where figure 3.34 shows SW480 cells trapped in this device. The traps showed between 70-100% trapping efficiency – defined as the percentage of full traps – in two



Path	length	width
P1	$\frac{g}{2}+\left(\frac{g}{2}-n\right)=g-n$	$2q+2g+w$
	l	w
	m	$q+g+\frac{w}{2}$
	integrate	*integrate*
P2	n	$g-m$
	$q+(q-n)+u+l$ $=2q-n+u+l$	g
	$\frac{g}{2}+\frac{g}{2}=g$	$2g+l+u$
	$\frac{g}{2}+\left(\frac{g}{2}-m\right)=g-m$	$2q+2g+w$
	m	$g+q+\frac{w}{2}$

Figure 3.31: Takeuchi-like trap [241] schematics used to show the different parameters of trap design optimisation, similarly to figure 3.30 (page 82). In red, the path through the trap P_1 divided in 4 different subpaths. In blue, the path around the trap P_2, divided in 8 different subpaths. The subpaths widths and lengths are indicated in the table on the right.

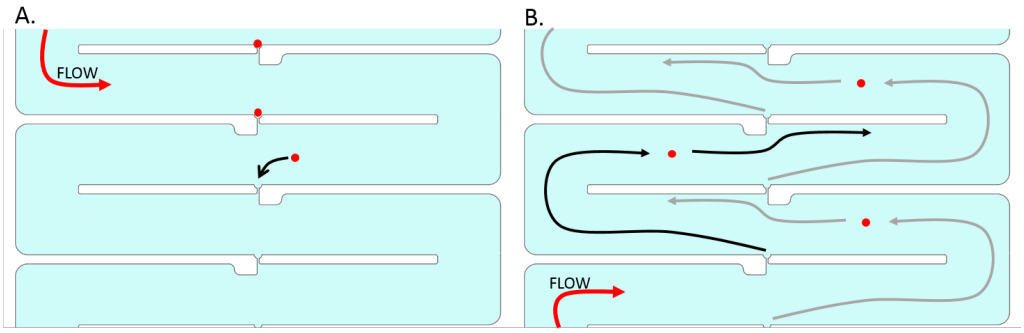


Figure 3.32: Takeuchi-like trap [241] schematics of the process of **A.** trapping, where the cell gets suctioned to the trapping region due to a mix of inertia after the turn and enough flow flowing through the trap, and **B.** untrapping when reversing the flow, where building $m \times n$ pillars in the bottom region of the trap lowers the probability of the cells to get trapped in the bottom part of the trap when collecting the cells.

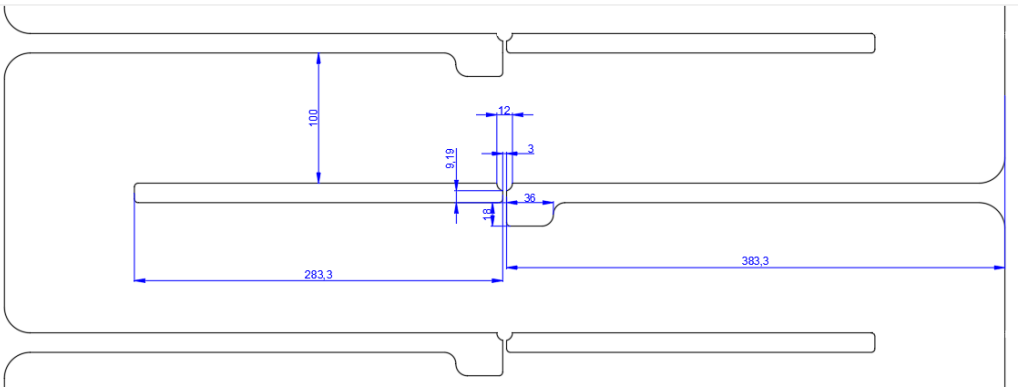


Figure 3.33: Takeuchi-like trap [241] schematics schematics with indicated dimensions, where the resistance ratio of these traps was in theory of around 100. However, fabrication of these traps resulted in a wider channel, with a final resistance ratio of $\approx$ 30-40.

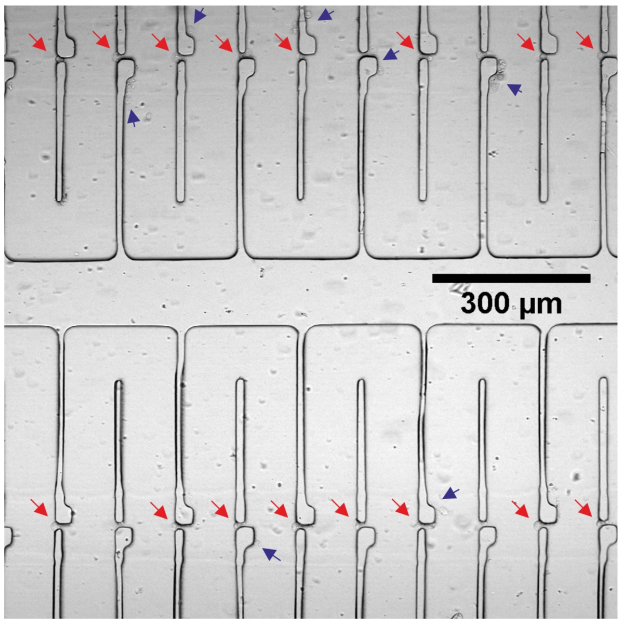


Figure 3.34: Takeuchi-like traps (final resistance ratio of ≈ 30-40) with SW480 trapped cells. Traps showed really good trapping efficiency, with empty traps in most cases due to too big channels caused by faulty fabrication. Red arrows indicate trapped cells and blue arrows indicate debris or cells stuck to the PDMS walls. Efficiency based on this image was of 80% full traps.

independent experiments, where the main issue found was untrapping of cells due to pressure peaks or drops caused by manually changing between different buffers. The traps in figure 3.34 showed ≈ 80% efficiency.

3.8.6 Cell Traps Design and COMSOL Modelling

When designing a cell trap, the naive approach can lead to a massive resistance through the trap compared to the resistance around it, with only some cells getting trapped "by chance" when using high concentrations – likely due to collisions with other cells as they are going through the trapping region.

Using COMSOL fluids simulations, the velocity profile around the traps can be obtained when varying different trap parameters. For the simulations, laminar flow with water and incompressible barriers for the PDMS traps were selected, with periodic boundaries for the sides of the chamber and constant inflow from the inlet (left), and atmospheric pressure for the outlet (right). These simulations are very accurate, but they are usually time-consuming, with an extra coarse mesh setting still taking about 1.5 min to run for a 2D model and hours of computational time when simulating a 3D model with a fine mesh. An example of one of these simulations is shown in figure 3.35, where the array horizontal and vertical spacing was kept constant at 80 μm. In this example, the parameters being explored were the variation of the length and width of the channel and how this affected the velocity field and the resistance inside the channel. However, the velocity in the trapping area is almost stagnant. When calculating the resistance ratio of these traps, for the $l = 5$ and $w = 8$ trap which gives the best results from the array, the ratio is >150, making trapping very unlikely.

Similar simulations were made to explore the variation of the array spacing and the variation of the trap shape. Figure 3.36 shows the velocity profiles for arrays of traps with different shapes. This was the first design to take into consideration the resistance through and around the trap, with a ratio of around 27, showing higher velocities inside the channel. We can see how the trap shape affects the velocity profile of the trap, with the square traps giving

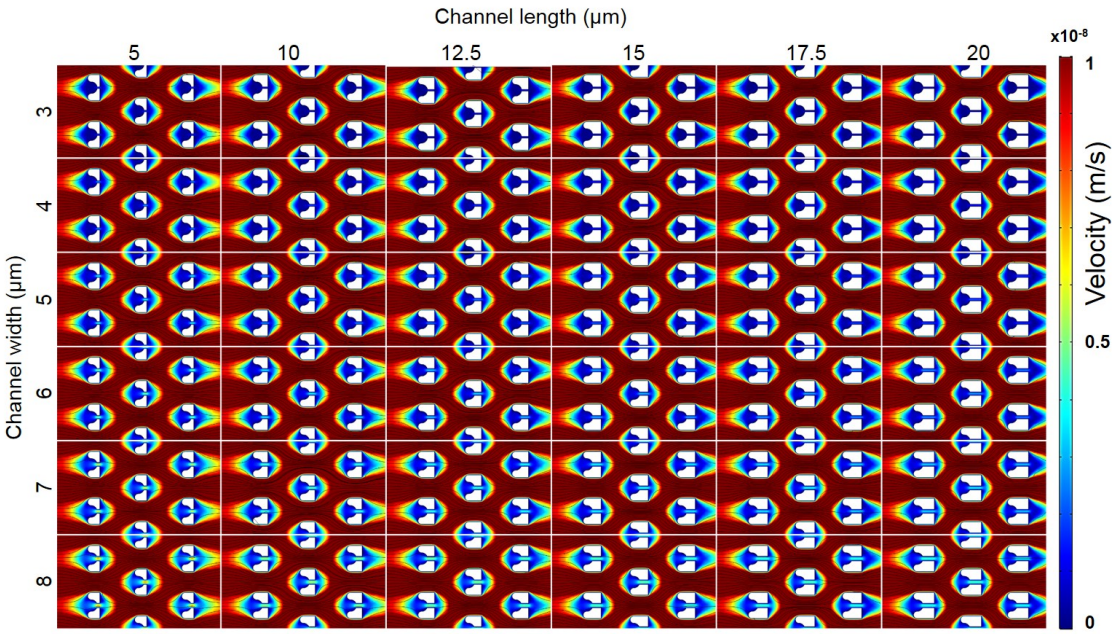


Figure 3.35: Example of the velocity profiles that can be obtained from simulating in COMSOL Di Carlo-like traps with different channel widths w and length l when keeping the same horizontal and vertical spacing between traps.

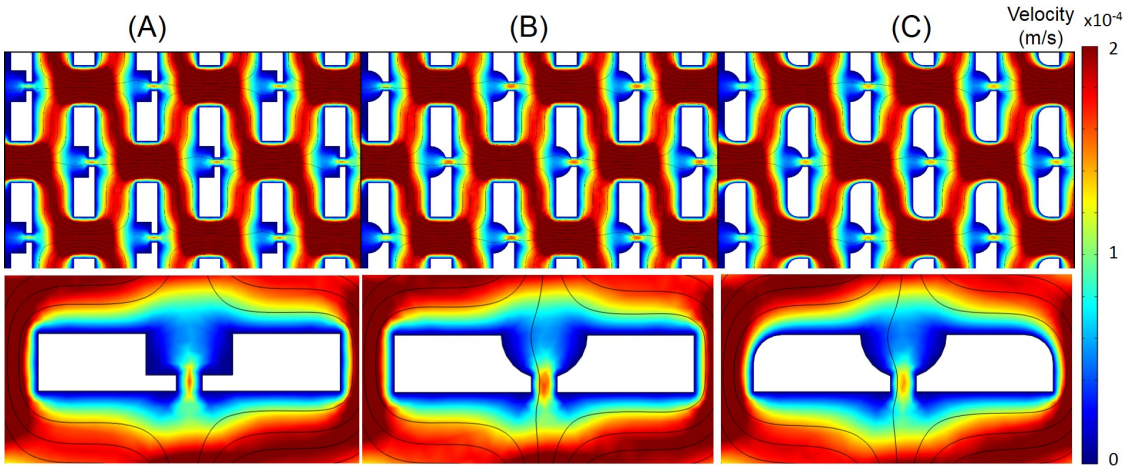


Figure 3.36: COMSOL simulation of Di Carlo-like arrays of traps with different shapes, with A. having square trapping points, B. with round trapping point with square corners elsewhere and C. round top corners.

higher velocity options than the rounded ones. However, there is always a bit of roundness when producing the microfluidic chips, especially when dealing with high aspect ratios such as the one in the channel (height is 25 μm with a width of only 4.5 μm).

Using the approach described on subsection 3.8.4 cell traps resistance can be calculated and optimised parameters can be obtained for the desired resistance ratio. In general, a massless point-like object would follow the flow lines and so will get trapped when the resistance through the trap is lower than the resistance around it – ie. $R_1/R_2 = B < 1$. However, experimentally cells got trapped with resistance ratios up to 40 if the concentration of the solution was high enough, although a lot of cells were missed. A cell will have a certain inertia, and so it will have a resistance to change direction in the S shaped path around the

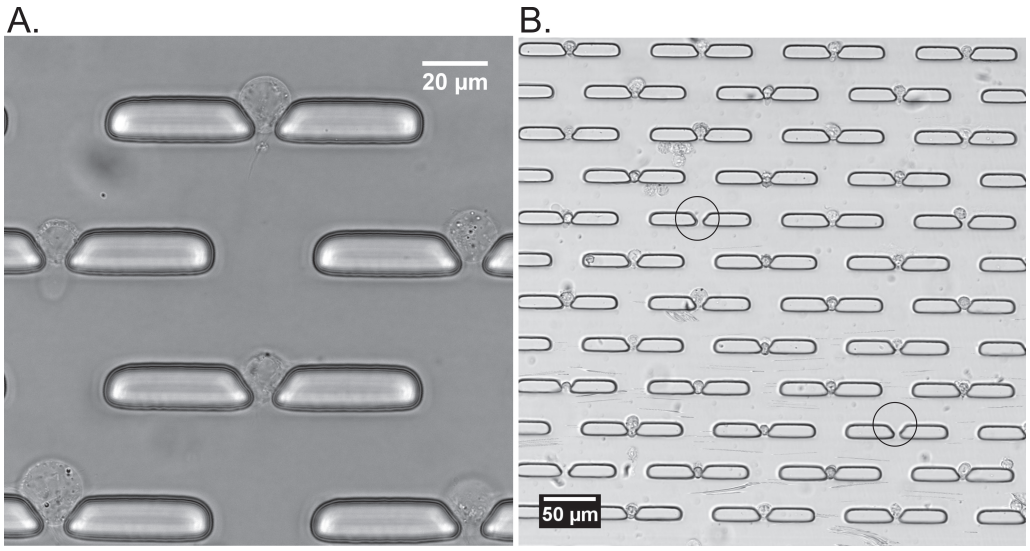


Figure 3.37: **A.** Example of HCT116 cells trapped onto one of the optimised Di Carlo-like trap arrays. **B.** Example of SW620 cells trapped in the same device, with the empty traps indicated with a blue circle.

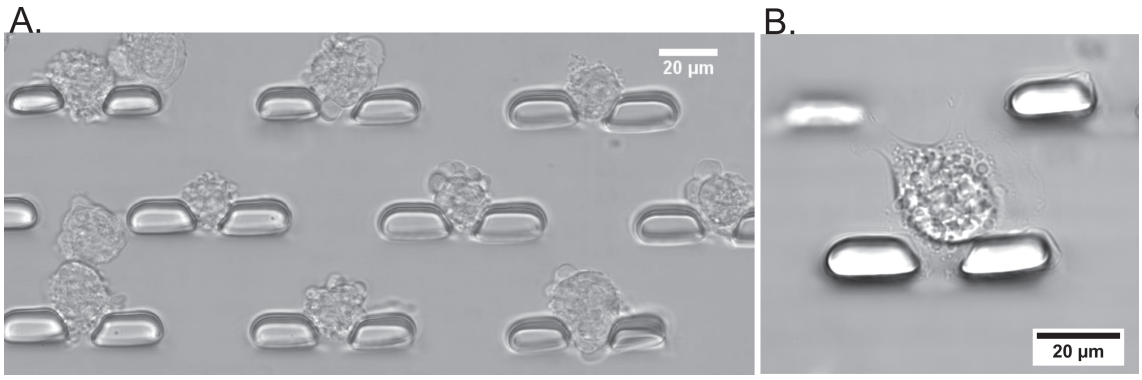


Figure 3.38: **A.** HUVECs cells trapped *on a chip* on Di Carlo-like optimised cell traps. **B.** HUVECs were allowed to adhere onto the chip and were subsequently cultured for multiple days.

trap, giving a bit of flexibility in the resistance ratio that achieved trapping, being able to increase the number of traps per unit area.

An example of cells trapped in one of these arrays is shown in figure 3.37A and B. For those traps, the ratio was around 35-40 for the final fabricated structures, but the chip achieved good trapping (between 85-95%) as long as the cell solution had a concentration of cells between 1-2 million cells/mL. Some cells were seen to deform through the trapping channel but were substituted by other cells, indicating that the channel was probably too wide to successfully retain all the cells.

A sensible group of parameter choices would be the following: being the trap radius u, the gap was chosen to be smaller than 30% the cell diameter $w < 0.3 \cdot (2u)$, with a channel length l at least as long as the width w for reproducible fabrication. The gap between the traps was chosen to be bigger than 2.5 cells to avoid clogging: $g > 5u$.

These traps were used for the experiments from chapter 6, and in other collaborations during the PhD. An example of one of these collaborations is shown in figure 3.38, where HUVECs[1]

[1] Human umbilical vein endothelial cells

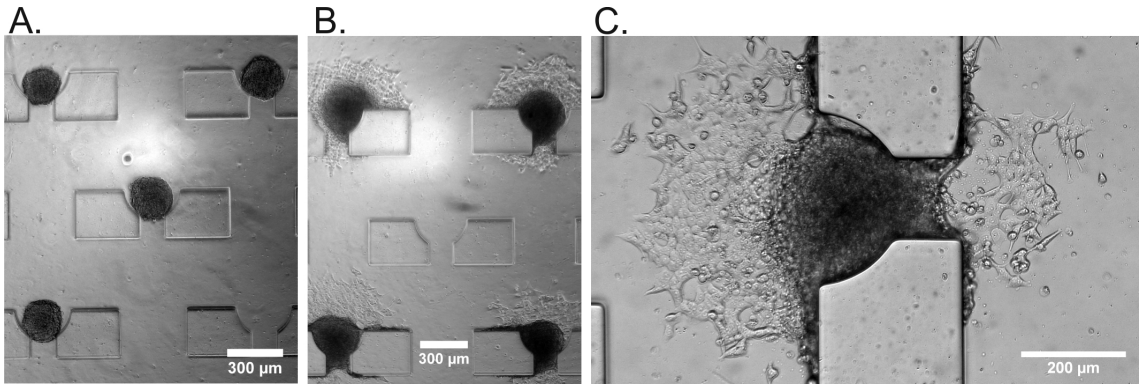


Figure 3.39: A. Spheroids (ca. 200 μm) trapped *on a chip* using Di Carlo-like optimised traps. **B.** Spheroids were allowed to adhere and expand around the trapping site. **C.** Zoom in on the adhered spheroids. Experiment and fabrication done by Matthew Bourn.

were trapped *on chip* and cultured, and subsequently transfected *on chip* by collaborators in Biology (University of Leeds), Dr. Vas Ponnambalam and in Switzerland, Dr. Edgar Delgado Eckert. A different design was used for trapping OLI-MB[1]s and washing them from the LOND[2]s solution with DPBS[3] in Churchman *et al.* [242]. The optimisation method also worked even when increasing considerably the height of the device and the size of the trapped object. An example of this is figure 3.39 of optimised traps for trapping spheroids, by colleague Matthew Bourn, where the trap dimensions were optimised before fabrication.

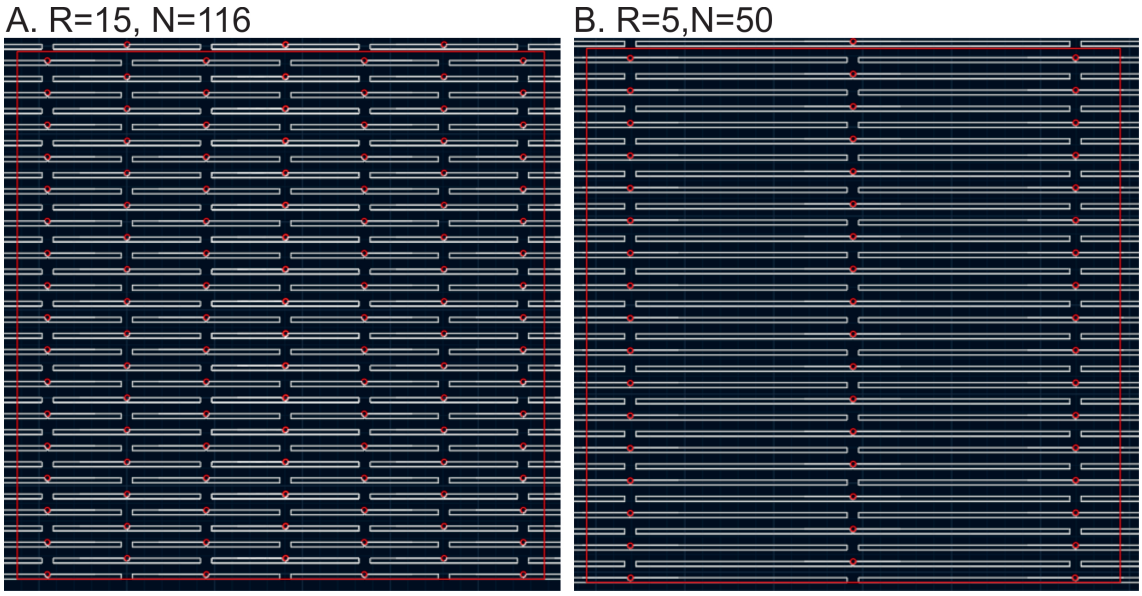


Figure 3.40: A. Di Carlo-like traps with a resistance ratio R of 15, and 116 trapped cells enclosed in the red square. **B.** Traps with a resistance ratio R of 5 and 50 trapped cells enclosed in the red square. The red square enclosing the traps has an area of 1500 $\times$ 1500 μm^2.

However, these traps are still being optimised. A point needs to be made that it can be more beneficial to increase the total number of cells trapped than obtaining perfect occupancy but lower trap density. For instance, if decreasing the resistance ratio to 5 will require a trap with bigger area, it might be better to keep a resistance ratio of 15 with half the occupancy, obtaining more trapped cells. An illustration of how drastically the trap repetition area can

[1] Oil layer inside microbubbles
[2] Lipid oil nanodroplets
[3] Dulbecco's Phosphate-Buffered Saline

increase when reducing the resistance is shown in the schematic of figure 3.40. Traps for cells of diameter 16 μm with a spacing of 2 cells between traps were optimised requiring a resistance ratio of either 15 or 5. In both cases, a square of area 2.25 mm^2 was drawn around the trapping area. For the traps with a resistance ratio of 5, only 50 traps were enclosed in the area, with more than double the traps enclosed in the traps with a resistance ratio of 15 (116 traps). Thus if the experiment requires maximizing the number of cells trapped and experimentally the design from A obtains more than half the efficiency of the design from B, then the design from A is optimum.

Part III

Results

Chapter 4

Raman Spectroscopy on Live Single cells

Previous chapters reviewed the main background and theory and described the Raman system capabilities and the experimental methods used in the current thesis. This chapter presents the first results of a RS[1] study of live cells of multiple colorectal adenocarcinoma cell lines SW480/SW620/HT29 representing different stages of CRC[2] progression and compared these to a carcinoma cell lines HCT116 and a non-colorectal cell line HL60 (leukaemia). Data from SW480/SW620 was first compared using different multivariate methods – PCA[3]/LDA[4], DT[5] and SVM[6] – to find an optimal multivariate method to differentiate between these primary and secondary cancer cells. Then, additional cell lines were added to the analysis to identify possible metastasis biomarkers compared to a greater pool of cells ranging from different disease states, different disease types within the same organ and different cell origins altogether. The results showed the successful classification of these cells with high accuracy and identified potential biomarkers that will require further investigation in clinical samples.

4.1 Adenocarcinoma Model System

4.1.1 Cell lines Imaging

To begin characterising the model system, figure 4.1 shows the phase contrast images of the different cell lines analysed in this chapter, ranging from the non-adherent leukaemia HL60 cells to the epithelial HCT116, SW480 and HT29 cells and the fibroblast-like SW620 cells. Cell size and morphology varies between the cell lines.

Additionally, fluorescence staining was done to characterise the nuclear size and the organelle density, independently staining for DNA and lipid-rich organelles. Figure 4.2 shows nuclear staining (blue) and lipid staining (green and red) for the three model colorectal adenocarcinoma cell lines. Cells were prepared as described in section 3.6.3 (page 75). Nile red emission shifts to longer wavelengths for more polar solvents [243] so two different emission windows

[1] Raman spectroscopy
[2] Colorectal Cancer
[3] Principal Component Analysis
[4] Linear Discriminant Analysis
[5] Discrimination Trees
[6] Support Vector Machines

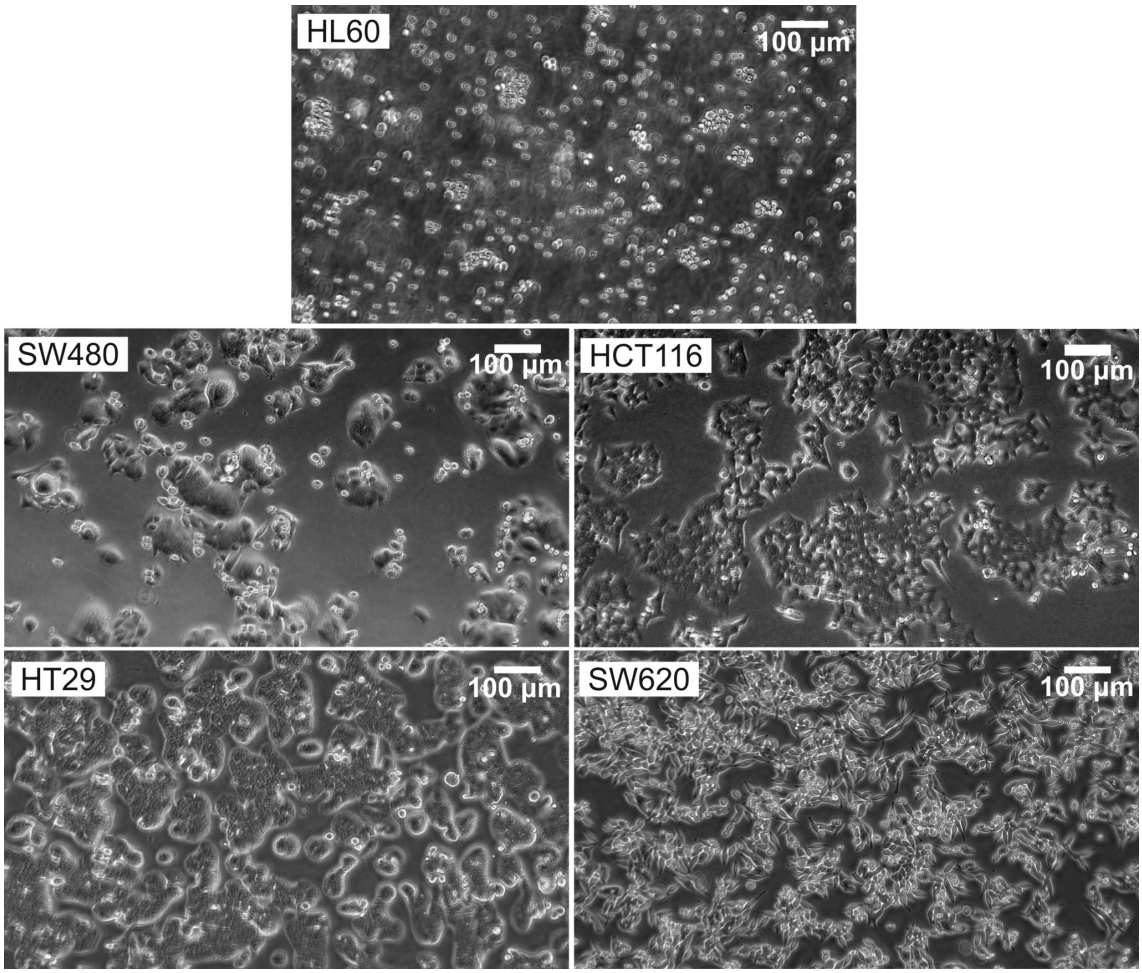


Figure 4.1: Phase contrast images of the cell lines used when grown in a polystyrene T flask with no additional coating. Different cell lines showed very characteristic morphologies.

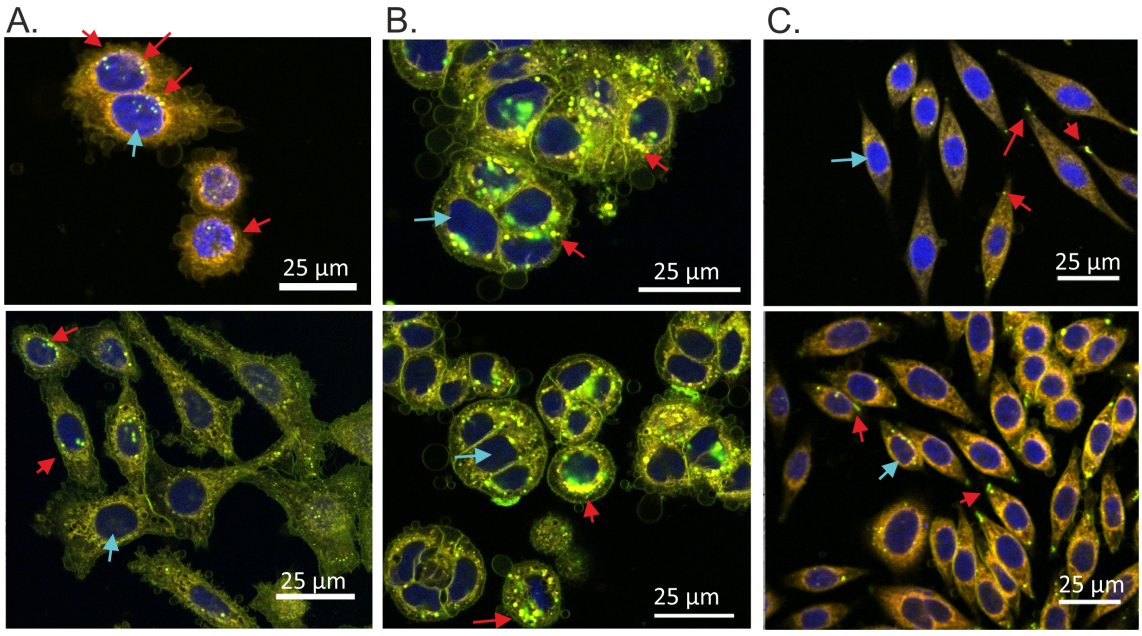


Figure 4.2: Hoechst DNA staining (blue) and Nile red lipid staining (yellow and green) for **A.** SW480, **B.** HT29 and **C.** SW620 fixed cells grown in a laminin coated surface. As a visual aid, some lipid droplets have been labelled with red arrows, and some nuclei have been labelled with cyan arrows. Protocol followed caused the appearance of some blebbing on the SW480 and HT29 cells, likely due to long wait between DPBS washing and PFA incubation, that are stained in the Nile Red channel.

were acquired to differentiate between more apolar environments, like triglyceride-rich lipid droplets, and more polar ones like phospholipid-rich organelle membranes.[244] The second Nile red channel had a lot of crosstalk from the first channel, so the polar lipids contributions always appeared orange compared to the apolar ones, that showed as bright green. SW480 cells were adhered to the glass with epithelial morphology, with multiple apolar vesicles in the cytosol and around or on the top/bottom of the nucleus. HT29 cells grew in isles with a great amount of these lipid droplets, mainly around the nucleus. SW620 cells showed a more fibroblast-like morphology, with very polarized cells that showed lipid droplets mainly in the poles and sometimes in the perinuclear area.

4.2 Raman on Live Single-Cells of Different Cell Lines

Cells were prepared and single cell Raman signal was acquired as described in section 3.3 (page 64). The experiment was repeated for each cell line to increase the number of cells analysed, where each experiment was done with cells from different passages. Data from multiple experiments were combined showing minimum variability between experiments, and thus good reproducibility. For all the analysis, data from different experimental days were treated equally, with a random choice of data from multiple days from training.

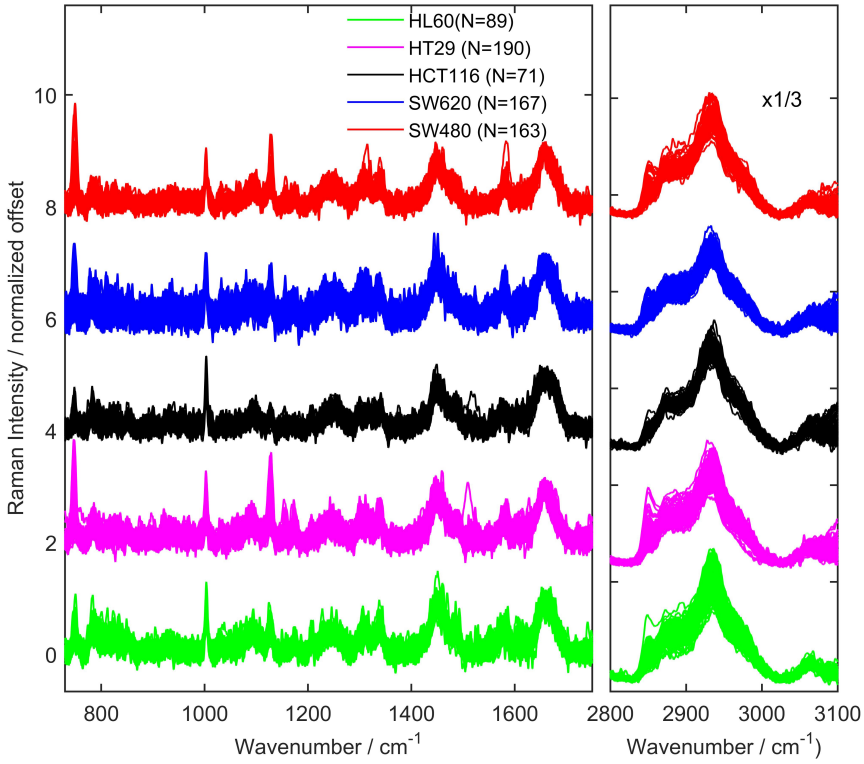


Figure 4.3: Raman Pre-processed data from the observations of all cell lines (HL60, HCT116, SW480, HT29 and SW620) for excitation with the 532 nm laser, after alignment to the calibration peak, cosmic ray subtraction, background correction, baseline subtraction and smoothing.

Figure 4.3 shows all the Raman data used for the classification of the 5 cell lines after the pre-processing steps (spectrum alignment, cosmic ray subtraction, background correction, baseline subtraction, smoothing, etc.), shown with an offset for clarity. All cell lines showed

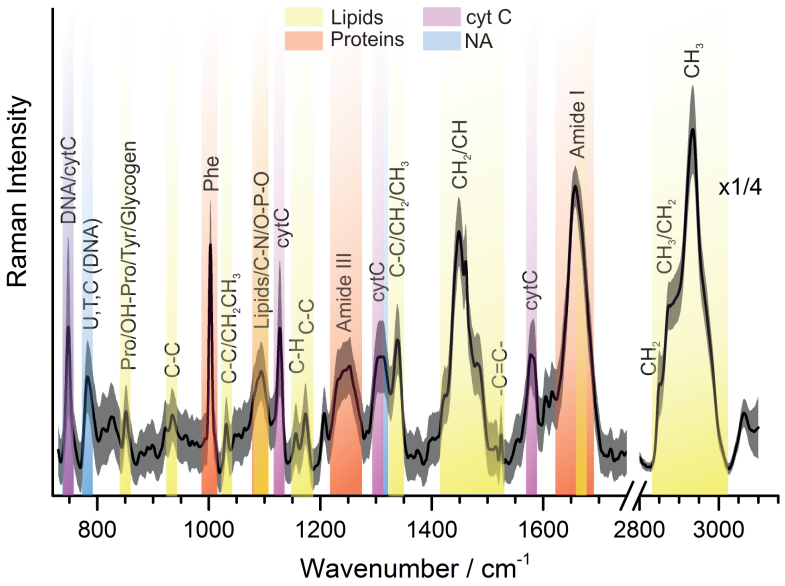


Figure 4.4: Peak labelling of the average Raman spectrum of all cells measured from all the cell lines (N=680), where the area around the curve corresponds to the standard deviation. Single cell spectra were acquired with the visible 532 nm laser (220 mW) with 50% defocusing. As a visual guideline, colours were used to indicate the main contributions for each band: proteins, lipids, nucleic acids or cyt C.

the presence of similar bands, with small changes in their intensity. There was variability between cells of the same cell line even when taking noise into consideration, meaning each of the cell lines showed a certain heterogeneity.

4.2.1 Spectral Averaging

Figure 4.4 shows the total average spectra, normalised to the Amide I peak, and the standard deviation for all measured cells (N=680). The most prominent peaks have been identified in accord with the established literature and are given in Table 4.1.[32, 43, 50, 245, 246]

At a first glance, the fingerprint region is dominated by the sharp phenylalanine peak (Phe) at 1002 cm^{-1} and by the two broad bands at 1455 cm^{-1} (CH_2/CH vibrations, associated with lipids) and the Amide I band at 1657 cm^{-1}. Between the Phe peak and the 1455 cm^{-1} band a series of smaller bands appear, with the Amide III band at around 1230 cm^{-1} and other carbon and carbon-hydrogen related vibrations. At lower wavenumbers than the Phe peak, three strong bands are worth mentioning, with the cyt C[1] band at 747 cm^{-1} overlapping with a DNA band, the DNA band at 782 cm^{-1} and the proline/hydroxyproline band at 815 cm^{-1}. In the CH-stretching region, a very strong CH_3 stretching band is present, with shoulder contributions of CH_2 stretches.

The SW480 and SW620 pair has exactly the same genetic background and thus was expected to be the most similar pair. These cells were compared to begin with, and different multivariate methods were tested to separate them in order to choose the ideal multivariate method for this particular classification problem. Figure 4.5 shows the average of SW480 and SW620 cells with tentative labelling of the main Raman bands, where the difference spectrum of the averages (SW480-SW620) was plotted underneath.

CH stretching The $-CH_2$ and $-CH_3$ stretching contributions in the region of 2800-3200 cm^{-1} (figure 4.5 and figure 4.6) showed higher overall intensity for SW480 cells and a greater

[1] Cytochrome C

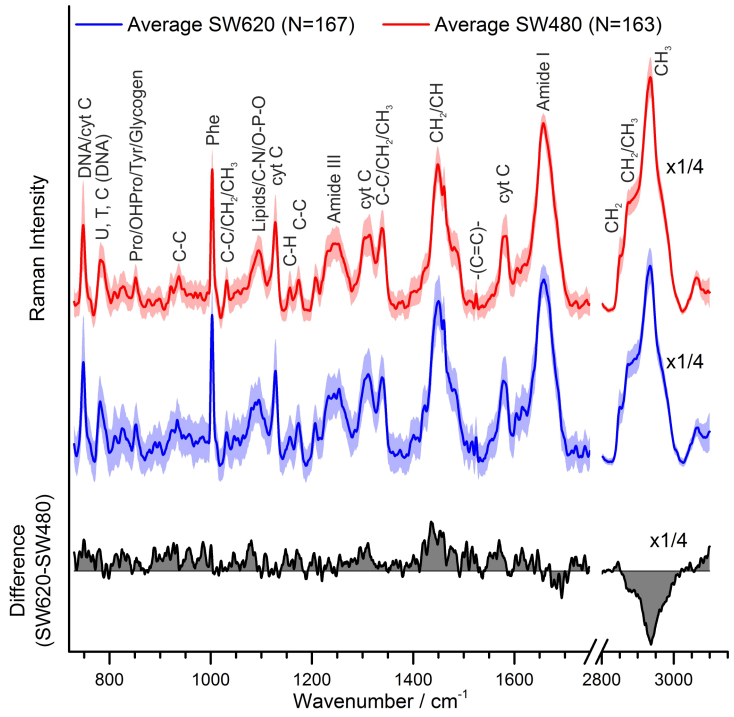


Figure 4.5: Average single-cell spectra and difference spectrum, for primary (SW480) and secondary (SW620) cells. The error around the average shows one standard deviation. The region around 2900 cm^{-1} is shown reduced by a factor of 4 to enhance the details in the fingerprint region. Tentative labelling of the most prominent peaks has been done based on the literature values as per Table 4.1.

Table 4.1: Tentative peak assignment for live single-cell Raman spectrum based in previous literature assignments[32, 43, 50, 245, 246] where ν=stretching, δ=bending, ρ=rocking, ω=wagging, τ=twisting, *as* = asymmetric and *s*=symmetric.

Peak position (cm^{-1})	Assignment (Position)[Reference]
747	DNA (748) [32], cyt C RR[1] (750) [50]
782	U, T, C (ring breathing modes in the DNA/nucleic acid bases) (782) [32]
1003	Phe (1000-1008) [32]
1095	Lipids (1095) [32], C-N (1093) [32] and O-P-O of nucleic acids (1087-1096) [32]
1128	cyt C RR (1127) [50]
1156	C-C vibrations (1152-1158) [32]
1175	C-H vibrations (1176) [32]
1207	OH-Pro (1206-8), Tyr (1204-6), Trp (1208-9), Phe (1209), A, T ring breathing modes (1208) [32]
1230	Amide III - β-sheet (1236-1259) [43]
1243	Amide III - disordered (1240-1245) [43]
1254	Amide III - $\alpha + \beta$ (1258-1286) [43]
1263	Amide III - α-helix (1268-1286) [43]
1311	cyt C RR (1314) [50]
1339	Trp, CH_2/CH_3 ω and τ (1337-9) [32]
1448	CH_2 δ, C-H deformation (1444-1454) [32]
1585	cyt C RR (1585) [50]
1653	Amide I - α- helix (1660-1668) [43], Fatty acids (1657) [32]
1667	Amide I - β- sheet (1667-1679) [43]
2848	CH_2 ν_s (2846) [245]
2873	CH_3 ν_s (FR) (2871) [245]
2901	CH_2 ν_as (2918) [246]
2934	CH_3 ν_s (2938) [245] (2937) [246]
2973	CH_3 ν_as (2965) [246]
3065	-CH ν_as (3015) [32]

CH_2:CH_3 ratio for SW620 cells, indicating differences in lipid composition between the two cell lines with higher lipid content for the larger size cells SW480 (SW480 diameter = 16.9 ± 0.4 μm c.f. SW620 diameter = 14.4 ± 0.3 μm) and in agreement with previous reports on fixed SW480/SW620 cells [97]. The fits of these peaks for all cell lines are shown in figure 4.6A4 and A5, where fittings were done in Origin using a Gaussian-Lorentzian, with position and width kept constant between all cell lines. The positions and widths of the peaks were allowed certain flexibility to ensure good fitting was obtained, with final positions indicated in figure 4.6A4 and A5. The bar plots for the amplitudes for these fittings are shown in figure 4.7C, showing in general higher intensities for SW480 cells than for SW620 cells, except the band at 2848 that showed higher intensity for SW620 cells, corroborating that SW620 cells showed higher CH_2:CH_3 ratio than SW480 cells.

Protein The amide III band (1230-1300 cm^{-1}) and the amide I band (1600-1690 cm^{-1}) are widely used for studying the protein secondary structure. Peaks fitted to the amide III β-sheet, α-helix, $\alpha + \beta$ and disordered structures showed that disordered structure was higher for the SW480 cells and β-sheet was higher for SW620 cells, whilst the ratios of α/β indicated that SW620 cells had more α-helix to β-sheet content ratio than SW480 as shown in figure 4.6B4 and B5. Other protein related peaks associated with hydroxyproline, proline and phenylalanine all showed higher intensity for the SW620 cells. The Amide I bands showed a similar trend to the Amide III fitting for the variation in the β-sheet and α-helix content. Figure 4.6C shows a similar fitting for the Amide I band - however, as the cell lines were normalised to the Amide I, this band should be analysed in terms of shape and not the overall intensity. The bar plots for the amplitudes for these fittings are shown in figure 4.7A and B. As a note, it may look contradictory that on average the α-helix band is higher than the β-sheet band for Amide III but this looks inverted for Amide I. However, the band intensity at each wavenumber is proportional to the cross-section, that changes for each vibration along the spectrum. That is, the β-sheet mode might have a stronger band in the Amide I vibration but a weaker band in the Amide III vibration, and thus only comparisons within the same band and between cell lines should be done. To better understand this, correlation analysis could be used in the future.

Nucleic acids The 782 cm^{-1} nucleic acid peaks and the 810 cm^{-1} peak usually associated with bonded phosphates or phosphodiester bonds showed a larger contribution for SW620 than of SW480 cells, indicating higher nucleic acid to protein ratio. The 1338 cm^{-1} band with mixed contributions of DNA and CH vibrations showed this same trend. This is consistent with the SW620 having larger RNA content [247] and nuclear area [248] than SW480.

Saccharides Most of the peaks associated with saccharide contributions showed a higher contribution in the SW620 spectra. This could be explained by higher concentrations of glycolysis intermediates such as acetate or lactate [249] and increased secretion of pericellular hyaluronan in SW620 compared to SW480 cells.[250] Peaks associated with phosphates also showed a higher contribution for the SW620, which is in agreement with previous reports that showed an increase of the phosphorylated status of these cells [96].

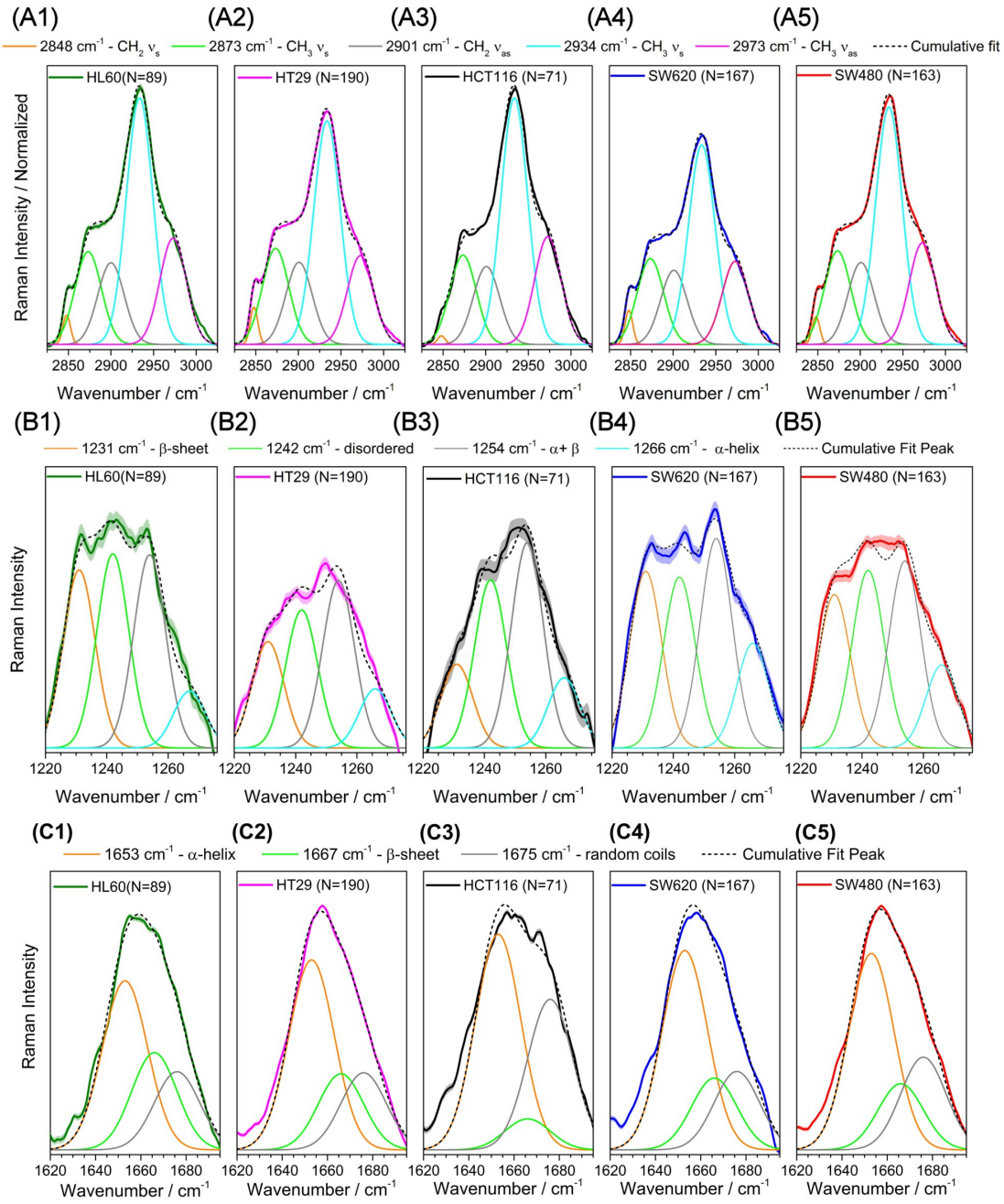


Figure 4.6: Averaged normalised Raman intensity for the CH-stretching region, the Amide III and the Amide I, where the broad faint area around the curve corresponds to the standard error. **A1-5.** CH-stretching Raman intensity for each cell line fitted with 5 Gaussian peaks with tentative labelling. **B1-5.** Amide III region fitted with 4 Gaussian peaks with tentative labelling. **C1-5.** Amide I region fitted with 3 Gaussian peaks with tentative labelling. Peak position and width have been fixed between the cell lines.

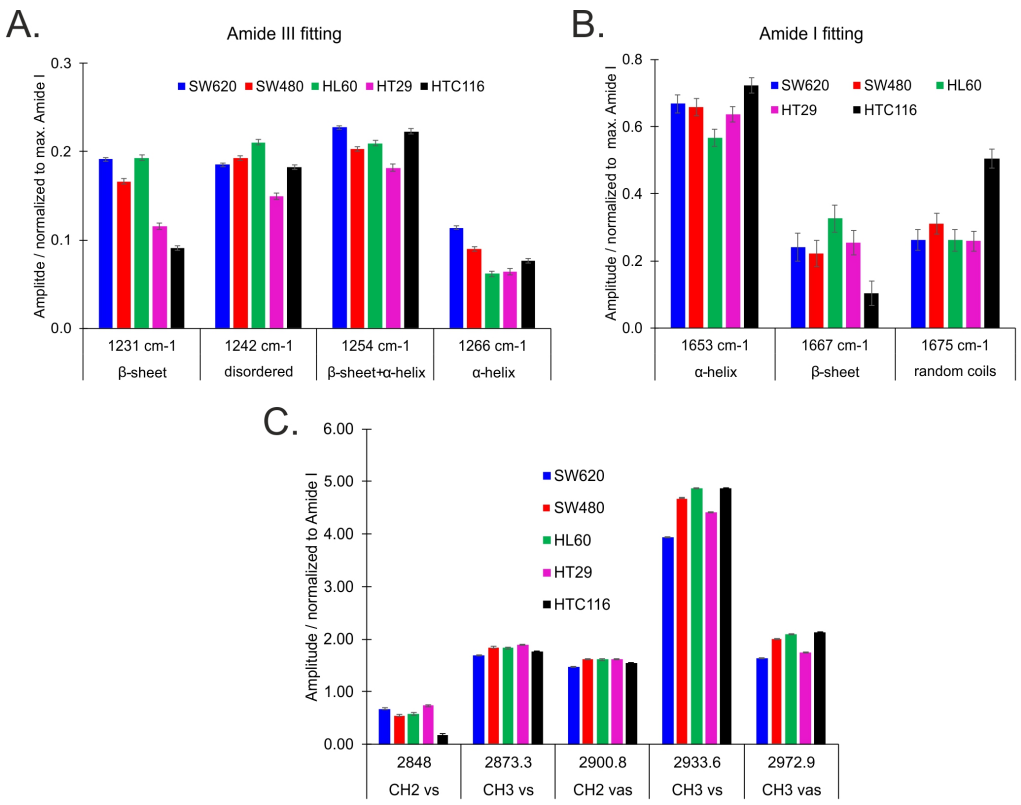


Figure 4.7: Bar plots for the fitting amplitudes for **A.** Amide III, **B.** Amide I and **C.** CH stretching, with standard error for each fitting.

Other peaks Peaks at around 1128, 1310 and 1585 cm^{-1} have previously been labelled as cyt C RR [50, 251] and have been reported to drop as an early sign of apoptosis.[251] Peaks at 1157, 1517, 1525 and 1620 cm^{-1} reveal higher contributions of double bonds to the SW620 normalised spectra, [32] and have previously been reported as cancer biomarkers in different biological samples, assigning them to carotenoids or porphyrins [60, 80, 97].

In summary, when normalising to the Amide I band, SW620 cells showed a larger contribution of α-helix proteins, saccharides, nucleic acids and double bonds related bands, whereas SW480 cells showed the larger contribution of lipids, β-sheet and disordered structure proteins.

4.2.2 Correlation Analysis

It is common to see different assignments of the same band values across the Raman spectroscopy literature. The position of a certain peak with the same label sometimes varies in ± 2 cm^{-1} between different literature reports, maybe due to differences in the calibration state of the equipment, with multiple assignments in the same region of the spectra. This makes it sometimes difficult to know what molecule is the major contributor to a certain band, even more challenging when the sample composition is as rich as the composition found in a cell.

To aid peak assignments and help track cell state we used the p-value filtered correlation matrix of the pre-processed data (figure 4.8), calculated as described in section 3.4 (page 71). Figure 4.9 shows the correlation between the lines discussed here. Only correlations with an absolute value higher than 0.3 were considered for this analysis. It is expected that

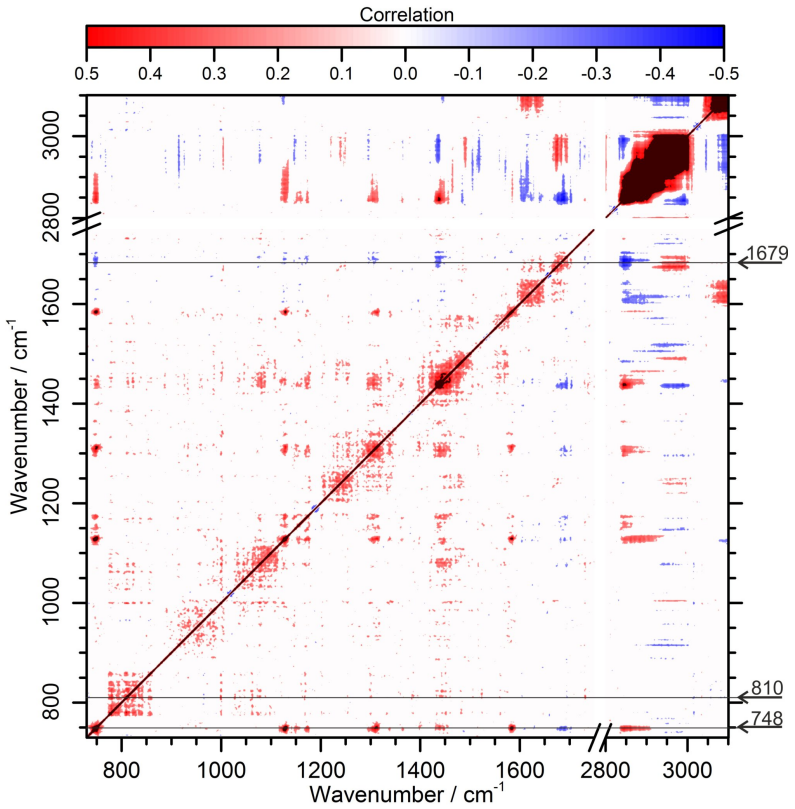


Figure 4.8: Correlation matrix of the different bands for all cells, where the points with p-values $> 10^{-4}$ were considered not significant and set to 0 to simplify the plot.

peaks associated to the same vibrational mode will show a high correlation which can help to understand the source of each peak and avoid uncertainty in the assignment, but this can also happen when two vibrational modes are often found together – e.g.: in the same organelle – or share a common regulation – e.g.: nuclear contents and membrane both increase with cell division. Additionally, in order to ensure that the correlations found are significant, a high number of observations need to be done, and thus this analysis is not always possible if the datasets are too small.

A series of strongly correlated peaks associated with cyt C were found at 748, 1128, 1156, 1175, 1310, 1431, 1438, 1448, 1585, 2845-67 cm^{-1} which had a strong negative correlation with the Amide I peaks at 1682 and 1690 cm^{-1} (see figure 4.8). A similar peak was observed at 752 cm^{-1} by Tsikritsis et al. 2015 [97] and labelled as porphyrins when exciting with the 785 nm laser. Here the intensity observed in this peak was stronger and it included peaks previously identified as cyt C RR. Okada et al. 2012 [50] labelled the bands seen at 750, 1127, 1314 and 1585 cm^{-1} as cyt C resonance, using 532 nm laser excitation, associating its drop with a sign of apoptosis on HeLa cells. cyt C contains a haem group that absorbs at 520 nm and thus can be excited at 532 nm. This would explain why its signature was assigned to porphyrins. Other peaks of this series had multiple assignments including CH_2 and CH_3 vibrations (1431, 1438, 1448, 2845-67). The peaks at 1156 and 1175 could have been due to C-C and C-H vibrations respectively, and the correlation with the 2845-67 cm^{-1} peaks could have been showing a correlation with the lipids signal may be due to the association of cyt C to the mitochondrial membrane or due to apoptotic changes that both change the cyt C signal and the shape of the main CH stretching bands.

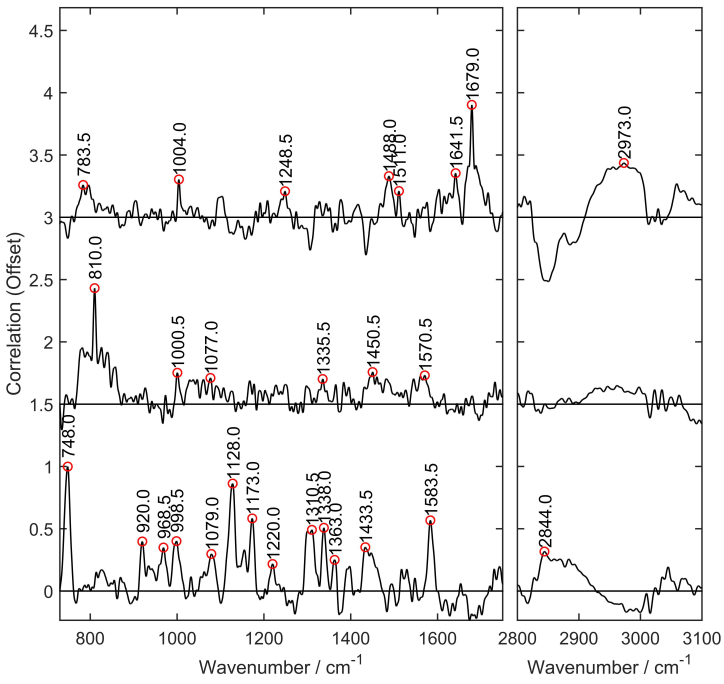


Figure 4.9: Correlation lines for the arrows shown in figure 4.8 showing the correlation series of 1679, 810 and 742 cm^{-1} from top to bottom respectively, with an offset for clarity. For each series, the horizontal line represents the value of zero correlation.

Other highly correlated peaks in the spectra was the 810 cm^{-1} series that positively correlated with 781, 828 and 1732 cm^{-1}. The 810 cm^{-1} is usually labelled as being due to phosphodiester or phosphate vibrations, with the 781 cm^{-1} peak associated with the pyrimidine bases ring breathing mode and the 828 cm^{-1} peak due to phosphates. Overall, this indicated that this series was related to nucleic acid vibrations.

Another notable correlation found was the series of 1679 cm^{-1}, which showed positive correlation along the Amide I peaks at 1642, 1671, 1687, 1689 and 1697 cm^{-1}. These bands are related to Amide I β-sheet (1679 and 1671), α-helix (1642) and disordered (1687) structures.

4.3 Multivariate Analysis

The individual cell spectra were used to classify cells by three different methods: with PCA[1], with DT[2] – both using an individual DT and using the C5.0 algorithm [228] – and with linear kernel SVM[3]. First, we consider this for the potentially more challenging case of SW480 and SW620 cell lines, which are of the same genetic origin and grown under the same conditions. Once optimised, we then extended this to other cell lines.

4.3.1 Principal Component Analysis

Figure 4.10A shows the first four principal coefficients or loadings. Using PC1-3 was enough to separate the two cell lines, by plotting the scores of the first 2 PCs (figure 4.10B). PC1 showed mainly lipid-related contributions and accounted for 26% of the variability, with

[1] Principal Component Analysis
[2] Discrimination Trees
[3] Support Vector Machines

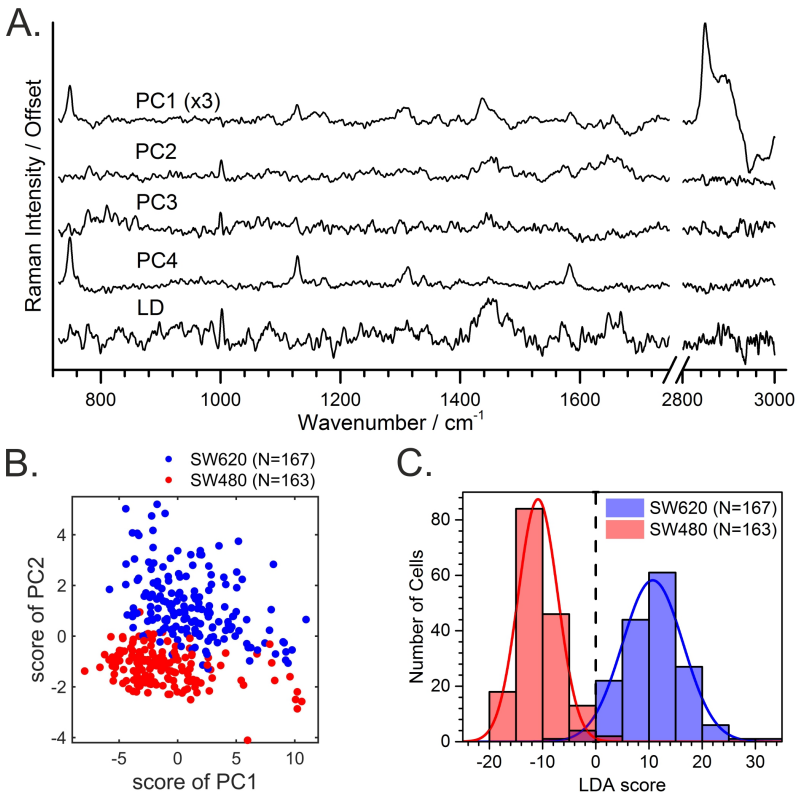


Figure 4.10: PCA/LDA results. (A) The shape of the PCs 1 to 4 and of the LD. PC1 is shown multiplied by a factor of 3 so the features in the fingerprint region are visible. (B) The 2D plot of the scores for the first two PCs. (D) Histogram of the individual cell scores when projecting the cell data onto the LD from (A) with a vertical dashed line at the point of best separation.

also contributions in cyt C[1] peaks. PC2 and 3 showed mixed RNA and protein related contributions and contributed to 5.1% and 3.3% of the variability, respectively. PC4 showed a strong contribution of the cyt C resonance. PCA showed that the SW620 cells are more heterogeneous than the SW480 cells, indicating greater within-class variability. PC2 was the component that better separated the two cell lines and showed two sharp peaks at 781 cm^{-1} (DNA) and 1001 cm^{-1} (phenylalanine), and broader peaks around 1455 (CH_2 vibrations), 1573 (Carboxylic group or nucleic acids) and 1647 cm^{-1} (Amide I). This component seems to be accounting for mixed contributions to proteins, lipids and nucleic acids. Interestingly, PC4 did not show different contributions between the SW620 and SW480 cells but seemed to be related to the within-class heterogeneity of the cells. The histograms of the scores are given in figure 4.11. The variance explained and the cumulative variance explained for the first 25 PCs is shown in figure 4.12A as a measure of how effective PCA is in dimensionality reduction and how important each of the PCs is in order to explain the total variability of the system. In this case, 25 PCs explain around 60% of the total system variability, with PC1 accounting for around 25% of it.

4.3.2 Principal Component Analysis and Linear Discriminant Analysis

Figure 4.10A shows an example of a Linear Discriminant (LD) when using 25PCs that provides a good classification of the two cell lines (correct rate of 98.7±0.3%), when running the model onto a random sample of the cells (10 fold validation). This LD was dominated by the PC2 contribution and showed positive values for SW620 cells and negative values for SW480 cells. The shape of the LD showed the enrichment in CH_2 *vs* of the SW620

[1] Cytochrome C

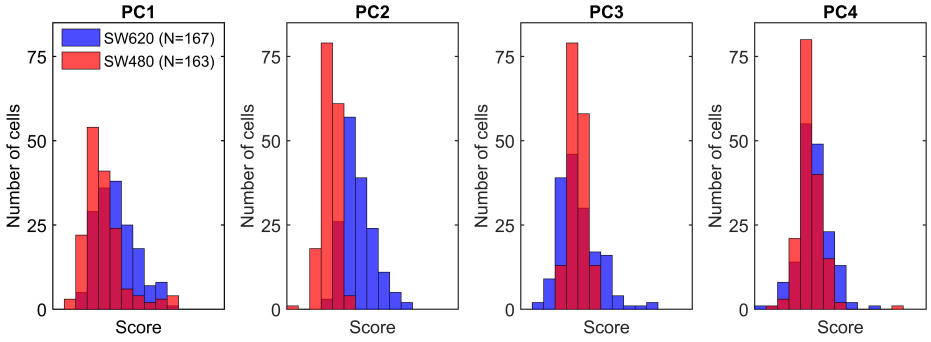


Figure 4.11: Histograms for the scores of components 1 to 4 for the SW620 and SW480 cells. The areas where the histograms overlap as shown with the overlapped colour (deep red). PC1 and PC2 showed different average scores for each cell line, where PC2 showed the best separation. PC3 and PC4 showed similar average scores for each cell line, showing mainly within-sample variability.

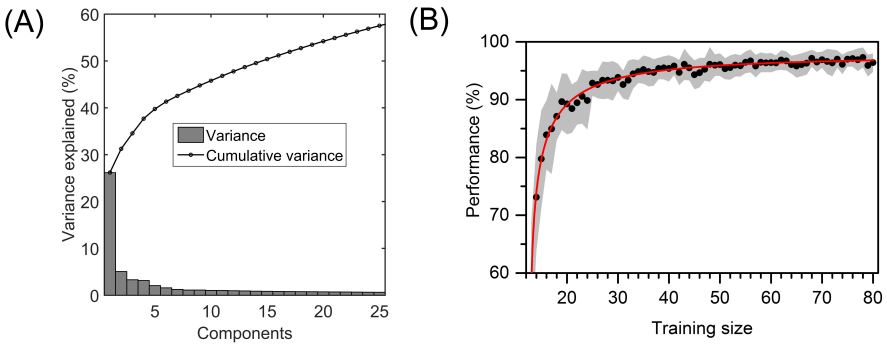


Figure 4.12: (A) Scree plot showing the variance explained and cumulative variance explained of the first 25 PCs. (B) Performance of the LDA model when using 25 PCs and different numbers of cells in the training group, with a test group of 50 cells. The area shows the standard error and a Sigmoidal Weibull function was fitted to the data with a saturation value of 97.4±0.3%.

and the increased content in CH_3 stretching vibrations of SW480 cells. The cyt C associated peaks were absent, indicating that the differences found here are not artefacts caused due to differences in viability from different samples. Modes related to phosphates were in general of negative sign, whilst the amino acid-related peaks like phenylalanine, tyrosine or hydroxyproline and the Amide III band showed positive contributions.

The performance of the LDA model for different training test sizes was also tested as shown in figure 4.12B, where the model plateaued at around 40 cells (around 20 cells from each cell type chosen randomly), with a saturation value of 97.4±0.3%.

In summary, the LDA[1]/PCA results confirmed that the metastatic SW620 cells had a higher CH_2/CH_3 ratio as well as larger contributions from amino acids, phosphates and proteins than the SW480 cells at a single-cell level and that these were good biomarkers to classify the cells. The scores for the LD are shown in figure 4.10C.

4.3.3 Discrimination Trees

A binary classification tree is based on choosing the best variables to classify the data by using multiple binary decisions. An example of this tree when using k-fold validation (k=10) is shown in figure 4.13A, where 90% of the 330 cell spectra available were used for the training (145 SW620 and 152 SW480 cells) – equivalent to one of the splits of 10-fold cross-validation. The cells were classified by using 10 decisions. To avoid over-fitting of the model, the test set was used to assess the performance. To choose the best size for the training dataset, trees were fitted to random training subsets of different sizes, and the performance was calculated by applying each of the models obtained onto a random testing subset of 50 cells from the testing set (see Figure 4.13B). When fitting a model with 5 repetitions of 10-fold validation the average performance obtained was of 86±1%. It was common that lipid-related peaks such as 2938 or 1455 cm^{-1} were chosen. The peaks chosen by any given tree varied slightly for different trees depending on the training subset. Thus using an algorithm for optimisation of the trees by fitting multiple trees is advantageous.

The data were then fitted to a C5.0 model [228], where multiple trees are used to optimise the final tree used for classification. Figure 4.13C shows the single frequencies most frequently chosen by the models as vertical lines, where the boldness of the lines is proportional to their frequency of appearance in the models, and the average spectra of the cells have been included as a reference. It can be seen that the Amide I region and the CH stretch regions were often chosen as classifiers. The band at 1455 cm^{-1} was chosen 99.1% of the time (CH_2 – CH bending assignments) whilst that at 1423 cm^{-1} (usually assigned as N-H in-plane deformation) was chosen 93.9% of the times. The second most commonly used band was at 2937 cm^{-1} (98.8%, CH_2 asymmetric and CH_3 symmetric stretching) surrounded by other frequent bands all over the CH_3 stretching region, all chosen by more than 80% of the trees. This underlines the importance of the CH_3 contributions for the discrimination between the cell lines. Interestingly, the 890 cm^{-1} peak was chosen 87.6% of the time (often labelled as saccharide band or methylene group). Amide I bands were chosen less often (<75%).

[1] Linear Discriminant Analysis

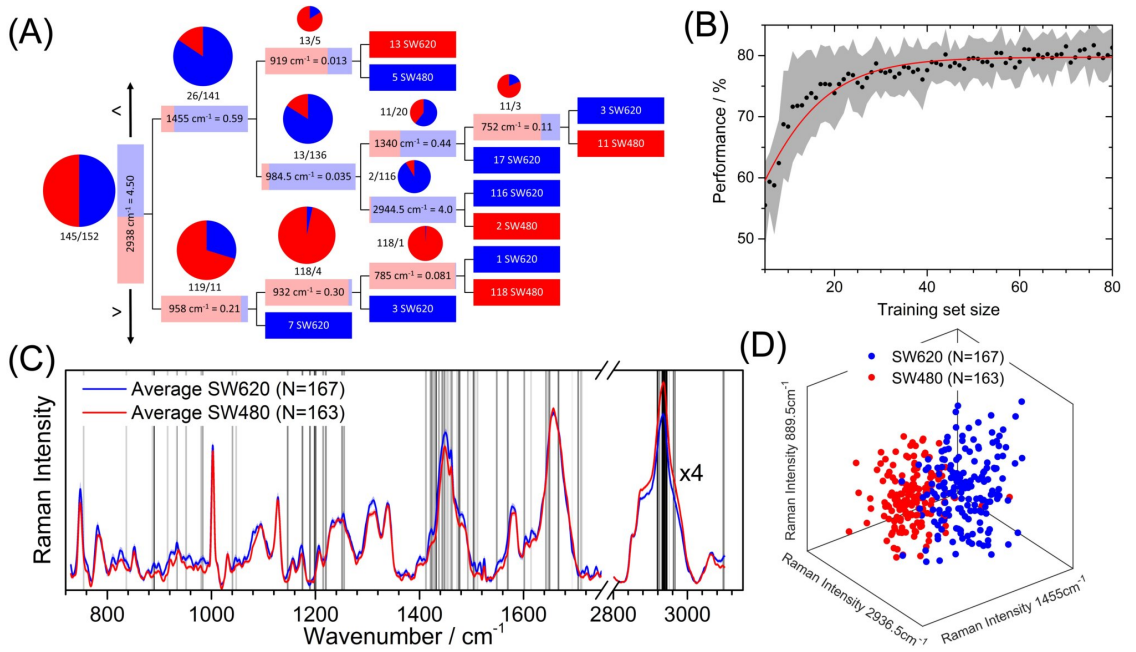


Figure 4.13: **(A)** Example of one of the trees fitted when using k-fold validation (performance of 83.3%). For each node, the cells whose intensities were lower or higher than the value of the node are sent to the upper or lower branch respectively. Pie charts indicate the proportion of cells in each node (Red=SW480; blue=SW620). **(B)** Changes in the performance of a DT with the training set size, with the broad line around the curve showing the standard error. **(C)** Bands chosen by the C5.0 trees shown as vertical lines, where the line intensity is proportional to the frequency at which the bands were chosen. The averages of the two cell lines have been shown as a reference (broad line around the curves showing 95% confidence interval). **(D)** The 3D plot of the 3 most frequent bands obtained from the C5.0 algorithm.

Proline and hydroxyproline related bands were also chosen for the final classifier (938 cm^{-1} - 37% or 920 cm^{-1} - 23%). The intensities of the three most frequent bands were 3D plotted in Figure 4.13D, showing good clustering of the two cell lines.

4.3.4 Support Vector Machines

Whilst DT models are relatively interpretable, they are often outperformed by other machine learning methods. A prominent disadvantage of SVMs, however, is that it is difficult to extract useful knowledge from the trained models. In our experiments, SVMs (linear kernel) achieved of an accuracy of 98.2±0.4%, significantly higher than both the DT and the DT C5.0. The choice of kernel made no significant difference to the accuracy, suggesting that the problem is linearly separable, a conclusion that is also supported by the high accuracies achieved using LDA.

4.3.5 Model Comparison and Final Performance Values

The performance of all the multivariate methods is shown in figure 4.14A. The final performance values obtained were of 98.7±0.3% for the PCA/LDA classifier, 86±1% for the simple DT, 94.0±0.9% for the C5.0 DT and 98.1±0.4% for linear kernel SVM. Multivariate methods often balance between intuitive results and good performance [216]. The PCA/LDA

has the advantage that the LD shows the component of best separation and it is easier to relate the variance of specific spectral features and hence to relate it to the underlying biology. The simple DT and C5.0 DT output are of single bands, which is the simplest and most intuitive output to relate with the spectral changes from the ones reviewed here but also gives a less powerful classifier.

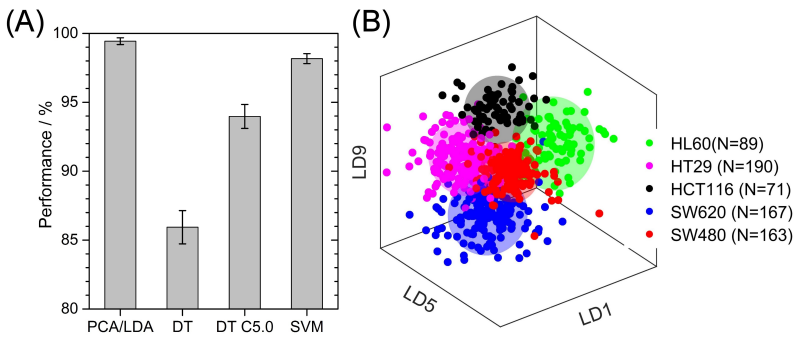


Figure 4.14: **(A)** Performance of the four classification methods when applied to the SW620 and SW480 datasets. **(B)** 3D-plot of chosen LDs of the different cell lines, where the spheres are centred on the average values and have a radius of two standard deviations.

4.3.5.1 Average Analysis of All Cell Lines

Figure 4.15 shows the average signal of the cells with the standard deviation of the signal for all the cell lines used.

The HL60 cell line showed lower intensity in the 749 cm^{-1} band but not in other cytochrome related bands, probably indicating lower DNA content than the adherent cell lines, but with higher intensity in the 782 cm^{-1} band associated with nucleic acids, which could be showing a higher RNA content. When looking at the 782 cm^{-1} band and the 810 cm^{-1} bonded phosphates band, the normalised intensity followed the trend HL60>HCT116>SW620>HT29>SW480. Interestingly, the modal number of the cell lines according to the literature shows the inverse trend HT29 (68-72)>SW480 (58)>SW620 (50) [252]>HL60 (46)>HTC116 (45) [253, 254]. As the peaks are normalised to the Amide I, this could be showing that the protein content was strongly correlated with the DNA content.

Previous studies of xenographs of HT29, HCT116 and SW620 cells showed that the most common metabolites were amino acids and lactate [255], indicating that the 1725 cm^{-1} peak associated with ν C = O and the 885 and 898 cm^{-1} peaks probably had a strong contribution from lactate. These peaks showed the trend HCT116>SW620>HT29>SW480≈HL60, which also agrees with previous magnetic resonance spectroscopy results [249, 255]. This can be attributed to the Warburg effect, due to which highly proliferative cancerous cells have increased lactate contents; HCT116, SW620 and HT29 are known to have lower doubling times than SW480 and HL60 cells [249, 256–258]. For the carcinoma cell lines, the lactate contribution appeared to be correlated with the cancer stage.

In general, HL60 had higher phosphate than the CRC[1] cell lines. For the CRC cell lines, differences between the 810 and 828 cm^{-1} peaks and the 1095 cm^{-1} peak could be indicating that HCT116, HT29 and SW620 cells showed more bonded phosphates than SW480 cells and that HCT116 cells had lower free phosphate concentration than the other cell lines.

[1] Colorectal Cancer

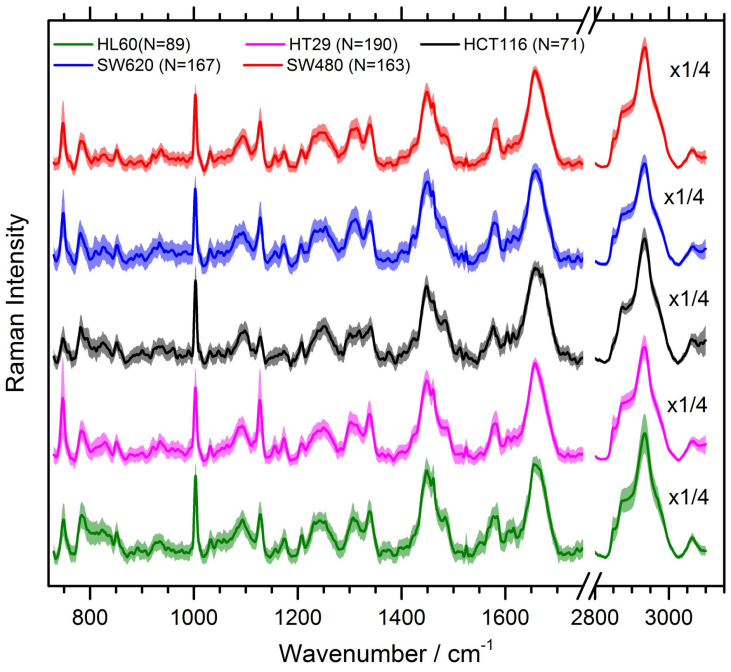


Figure 4.15: Raman intensity average signal for each cell line used, where the data was previously normalised to the Amide I band. The broad faint area around the line corresponds to one standard deviation. The silent region of the spectrum has been omitted and the CH stretching region has been multiplied by a factor of 1/4 to enhance the visibility of the fingerprint region.

Spectral regions around Amide III were fitted with Gaussian peaks as shown in figure 4.6B and figure 4.7B. The Amide III band has very different shapes for the different cell lines. Both HL60 and SW620 cells showed high contributions for both β-sheet, disordered and $\alpha + \beta$ secondary structure, with a lower contribution of α-helix structure. In contrast, HCT116, HT29 and SW480 cells showed reduced β-sheet peak height with higher disordered and $\alpha+\beta$ contributions, suggesting that increased ratio of $\alpha+\beta/\beta$-sheet could be a signature of primary colorectal cancer. This would merit further investigation. SW620 and SW480 cells showed higher α-helix contribution than the other cell lines. Amide I fitting showed a similar trend to the one seen in Amide III within fitting error, where HCT116 showed a significantly higher contribution of random coils. In terms of amino-acids content, the phenylalanine peak had a lower contribution for the adenocarcinoma primary cell lines (HT29 and SW480) followed by SW620 cells, with a higher contribution for the HCT116 cells and the HL60 cells.

Fittings were also done to the Amide I band, shown in figure 4.6C and figure 4.7B. However, the spectra is normalised to this band and there are also CH-stretching contributions affecting it, meaning that this band is perhaps harder to fit than the Amide III band. Overall, the Amide I band showed similar trends to the Amide III bands, with higher random coil and lower β-sheet contributions of HCT116 cells and higher β-sheet for HL60 and SW620 cells.

Fitting to the CH stretching region, figure 4.6A and figure 4.7C, showed SW480, HL60 and HCT116 to have higher contributions above 2930 cm^{-1}. HCT116 cells had a very low contribution in the 2848 cm^{-1} CH_2 symmetric band compared to the other cells, showing higher fatty acids levels for HCT116 cells than for SW620 cells [93, 255]. Whilst for the adenocarcinoma cell lines, the contributions above 2900 cm^{-1} appear to be dependent on the cancer stage (SW480>HT29>SW620), a promising biomarker that would need to be confirmed in further experiments with additional cell lines and ultimately in patient samples.

In summary, when normalising to the Amide I, the results suggest that HL60 cells showed low DNA, lactate, β-sheet content and high bonded phosphates, lipids, disordered and $\alpha+\beta$

secondary protein structure, clearly separating it from the CRC cell lines. HCT116 cells showed lower cyt C peaks, β-sheet content, free phosphates and CH_2 symmetric stretching band, and higher lactate, disordered and $\alpha + \beta$ contributions, all possible signatures of colorectal carcinoma compared to adenocarcinoma. For the colorectal adenocarcinoma cell lines, SW620 cells showed lower phenylalanine peak and lower $\alpha+\beta/\beta$-sheet ratio and SW480 showed lower bonded phosphates. The lactate contribution measured using the 1725 cm^{-1} peak seems to be proportional to the cancer stage, whereas the CH stretching contributions above 2900 cm^{-1} were inversely proportional to the cancer stage. This would indicate that more malignant cells would tend to increase their lactate/protein ratio – due to the Warburg effect – while decreasing their lipid CH_3/protein contents. However, from imaging data (see figure 4.2), we know HT29 and SW620 cells show more lipid droplets in their cytosol, explaining why they have increased CH_2/CH_3 ratio.

4.3.5.2 Multivariate Analysis of All Cell Lines

In general, the differences between the cell lines are subtle when looking at the average spectra, but are clear when applying the PCA/LDA model. The LDA model, when applied to all cell lines data, consisted of 10 LDs, each of them maximizing the separation between a pair of the 5 cell lines. Figure 4.14B shows a 3D plot of three selected LDs that showed the best separation where the average and two standard deviations of each cell population has been shown as a sphere. All cell lines showed very clear clustering separated from each other. HL60 clustered further from the other cell lines in LD1 as the only non-adherent cell line. HCT116 showed clear separation with the other CRC cell lines, underlying the ability of Raman spectroscopy to separate between different cancer types even within the same organ. SW480 and HT29 laid very close to each other and showed the worst separation as expected given that they both originated from primary colorectal adenocarcinoma. The PCA/LDA model using a 10-fold cross validation showed an accuracy of 92.4±0.4%.

Table 4.2: Confusion Matrix of PCA/LDA, where each column corresponds to the predicted members of a class, and each row corresponds to the actual membership of the class.

	HL60	HT29	HCT116	SW620	SW480
HL60	**91 ± 1**	3.5 ± 0.9	0.4 ± 0.3	1.3 ± 0.5	3.9 ± 0.9
HT29		**90 ± 1**	0.6 ± 0.2	0.5 ± 0.2	9 ± 1
HCT116		1.4 ± 0.6	**97.0 ± 0.9**		1.6 ± 0.7
SW620	0.6 ± 0.2	1.9 ± 0.5	0.6 ± 0.3	**96.2 ± 0.7**	0.8 ± 0.3
SW480	0.3 ± 0.2	7.8 ± 0.9		0.2 ± 0.2	**91.7 ± 1**

When running a 10-fold cross validation PCA/LDA on the data, the training group is of 68 cells, and the testing group is of 612 cells. The confusion matrix for the model obtained from calculating the 50 confusion matrices of the 5 repeats of 10-fold validations is shown in Table 4.2, where each row was divided by all the members of the row in each fold, and then the average and standard error was calculated (so the sum of each row adds to 100). This is the equivalent to a table where 100 observations from each class are available and allows calculating a standard error. HCT116 and SW620 cells showed very good separation from the other cell lines, with 97% and 96% correct rate respectively. HL60 cells showed moderate

misclassification to HT29 and SW480 cells (3.5 and 4%). The model often misclassified SW480 as HT29 cells and vice-versa, with around 8% of SW480s being classified as HT29 cells, and 9% of HT29 cells being classified as SW480 cells. Even though HT29 cells belong to a more developed stage of CRC, it is still a cell isolated from the primary tumour, similarly to the SW480 cells, and showed a more challenging classification – contrary to the initial hypothesis that the SW480/SW620 model would be the most challenging one as they have exactly the same genetic background. This highlights the ability of RS[1] to detect different cell phenotypes independent of genetic background.

4.3.6 Partial Least Squares Regression

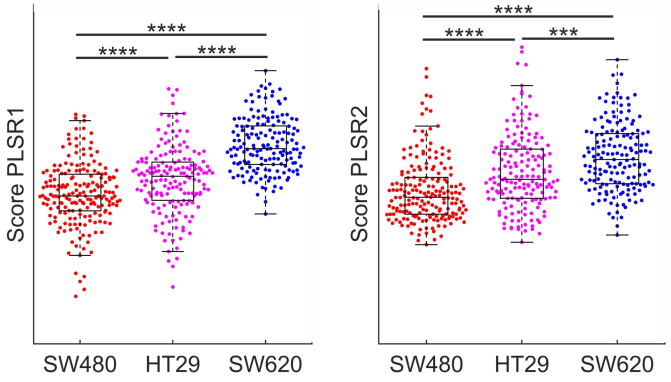


Figure 4.16: Composite Box plots / bee-swarm plots for the scores of each cell line in the PLSR coefficients 1 and 2, showing a linear trend with disease stage. The p-values for each pair (t-test) are indicated, where (*** = $p<0.001$) and (**** = $p<0.0001$).

In order to find possible biomarkers for disease stage in colorectal adenocarcinoma, the spectra of the SW480/HT29/SW620 cell lines were fitted to a PLSR[2] model looking for spectral features that changed linearly with disease stage. Only components 1 and 2 that showed a significant increase with cancer stage ($p<0.01$) were considered for analysis, and the cell scores for each cell line are plotted in figure 4.16 as beeswarm plots. These two PLRS loadings are plotted in figure 4.17. Among other potential biomarkers noticed in the average analysis, results showed a decrease of the stretching peak at 2850 cm^{-1}, decrease of the DNA peak at 787 cm^{-1}, increase of the 1438 cm^{-1} peak with a decrease around 1465-1490 cm^{-1} and a decrease in the Amide I contribution above 1675 cm^{-1}. An increase at 810-813 cm^{-1} (bonded phosphates and phosphodiester) may be linked with increased phosphorylated status with cancer stage [96] and/or increased nucleic acid content. The 1556 cm^{-1} peak related with double bonds and previously reported to increase in SW620 cells compared to SW480 cells [97], showed an increase with cancer stage when considering HT29 cells. PLSR analysis also seemed to indicate shifting to higher wavenumbers of the phenylalanine peak at 1002 cm^{-1} and the 1174 cm^{-1} peak from the cyt C series, and shifting to lower wavenumbers of the 747 and the 1227 cm^{-1} cyt C peaks with advancing adenocarcinoma stage.

[1] Raman spectroscopy
[2] Partial Least Squares Regression

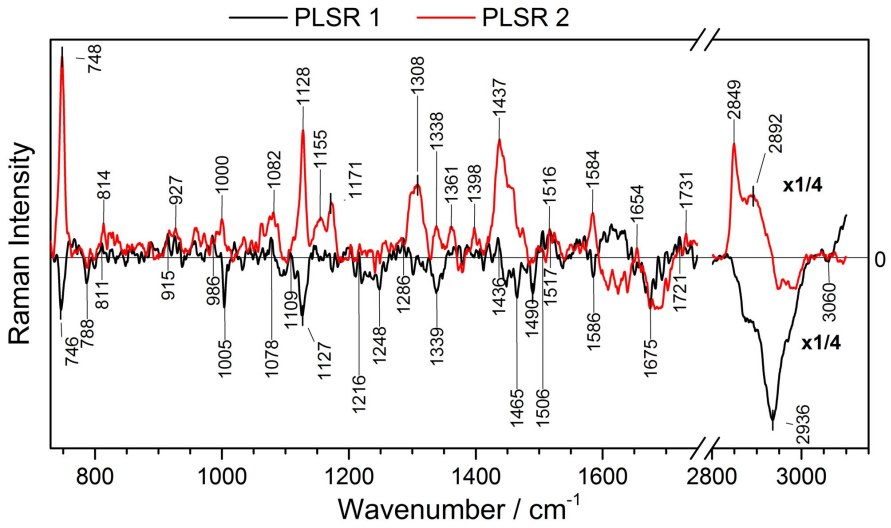


Figure 4.17: Shape of the two PLSR components that achieved separation between the SW480/HT29/SW620 populations with $p<0.01$. Main peaks have been labelled.

4.4 Final Conclusions

In this chapter it has been shown that RS[1] can be used for the classification of different cancer stages by doing spectroscopy of hundreds of live cells, including cells with identical genetic background. This shows that RS is sensitive to cell properties such as proliferation or metastatic potential. For the metastatic model system, it was found that when normalising to the Amide I peak, secondary tumour cells (SW620) displayed higher saccharides, phosphates, nucleic acid content, α-helix, β-sheet and $\alpha+\beta$ secondary structure, increased ratio of α/β secondary structure and increased ratio of $CH_2 : CH_3$ stretching bands. The SW480 cells displayed a higher proportion of disordered structure and increased overall CH stretching intensity. Correlation analysis was used as an aid for band assignment and to better understand the datasets. Little is understood on the link between Raman spectra variation and the underlying biochemistry, and Correlation analysis is a valuable tool to this end. PCA discrimination indicated that the cyt C[2] peaks accounted for most of the within sample variability whilst the protein, nucleic acids and lipid-associated peaks gave the largest variability between cell lines.

Supervised multivariate methods like LDA[3]/PCA[4] and SVM[5] results yielded >98% correct rate in classification between the SW620/SW480 cell lines compared to DT[6]s and C5.0 DTs, that gave a good but lower performance, though they allowed obtaining single peak biomarkers. HCT116 and SW620 cells showed the best classification, whereas HT29 cells showed the worst, with frequent misclassification of HT29 cells and SW480 cells.

When comparing multiple CRC[7] cell lines it was found that the primary CRC cell lines

[1] Raman spectroscopy
[2] Cytochrome C
[3] Linear Discriminant Analysis
[4] Principal Component Analysis
[5] Support Vector Machines
[6] Discrimination Trees
[7] Colorectal Cancer

(SW480, HT29 and HCT116) showed increased $\alpha + \beta/\beta$-sheet ratio in the Amide III band compared to the HL60 and SW620 cells. The carcinoma cell line HCT116 showed lower cyt C, CH_2 symmetric stretching and free phosphates, and higher lactate contributions compared to the adenocarcinoma cell lines. The analysis of the average and PLSR[1] analysis with the colorectal adenocarcinoma stage showed an increase in the lactate contribution at 1725 cm^{-1}, the 810-813 cm^{-1} peak associated with bonded phosphates and phosphodiester, and the 1556 cm^{-1} peak related with double bonds, and a decrease on the contributions above 2900 cm^{-1}, the DNA peak at 787 cm^{-1} and the Amide I contribution above 1675 cm^{-1} among others. These changes are likely indicating that more advanced cancer stages are more metabolically anaerobic cells with a higher phosphorylated status and with differences in their lipid metabolism and composition. Their possible applications as biomarkers deserve further study. Overall, the PCA/LDA correct rate for the separation of different cancer types was 92.4±0.4%, with even higher classification rates for HCT116 cells and SW620 cells, showing the potential of Raman spectroscopy for diagnosis of cancerous cells.

[1] Partial Least Squares Regression

Chapter 5

Raman for detection of photodynamic agents on single colorectal cancer cells

The previous chapter presented the first Raman results on live cells on multiple colorectal adenocarcinoma cell lines SW480/SW620/HT29, representing different stages of CRC[1] progression and compared these to a carcinoma cell lines HCT116 and a non-colorectal cell line HL60 (leukaemia). The results showed the successful classification of these cells using PCA[2]/LDA[3] with high accuracy and identified potential biomarkers that require further investigation in clinical samples.

In this chapter, Raman detection of different drugs and their effects on live single cells was investigated, beginning with a classical photodynamic system, in which the cells are incubated with 5-ALA[4] to induce the overproduction of PPIX[5], that can then be activated using laser light. A novel photosensitiser, DC473, was also explored, mapping its location within fixed cells and subsequently quantifying its accumulation in live cells. Finally, RS[6] was used to track the effects of DC473 PDT[7] on live single cells over time.

5.1 Classic Photodynamic Agent - Protoporphyrin IX

Protoporphyrins are tetrapyrroles containing four methyls, two propionic acids, and two vinyl side chains. PPIX[8] is a highly conserved molecule between all living organisms and is the final intermediate in the synthesis of the heme group (see figure 5.3 inset on page 114). [147] The number IX indicates the position of these side chains. PPIX is a hydrophobic molecule that can chelate metallic ions to form metalloporphyrins such as in haemoglobin, cyt C[9] or cytochrome P450 in the liver. [147] PPIX is regulated to be at low concentrations in cells and to be converted very efficiently to heme by the ferrochelatase enzyme. [147]

[1] Colorectal Cancer
[2] Principal Component Analysis
[3] Linear Discriminant Analysis
[4] 5-aminolevulinic acid
[5] Protoporphyrin IX
[6] Raman spectroscopy
[7] Photodynamic Therapy
[8] Protoporphyrin IX
[9] Cytochrome C

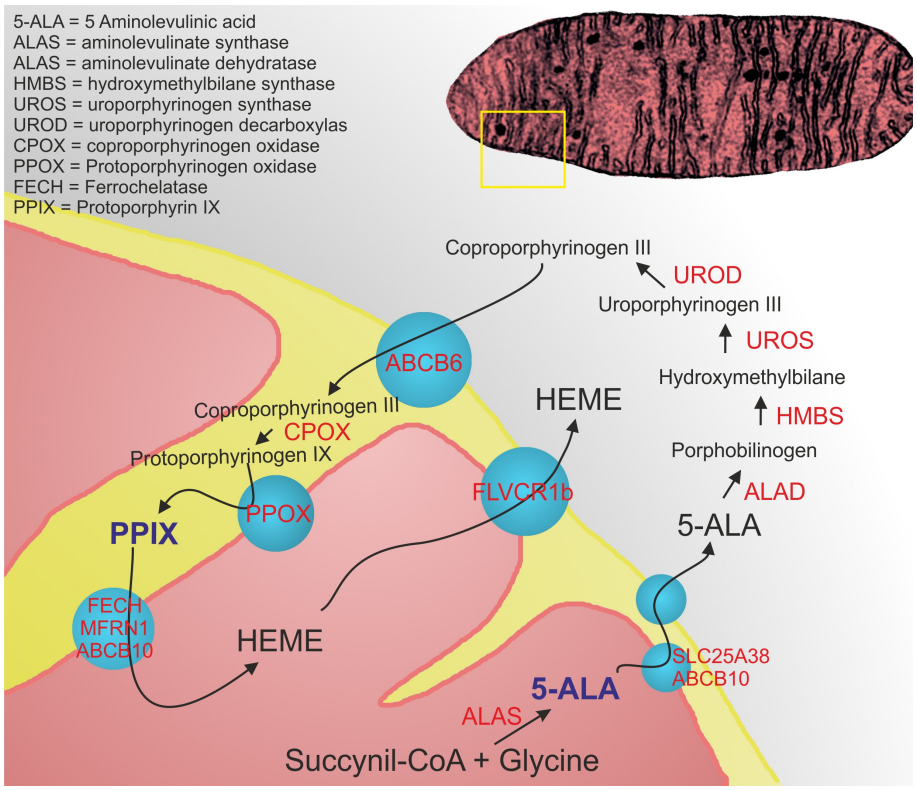


Figure 5.1: Schematics of the synthesis of PPIX from 5-ALA, adapted from Chiabrando et al. [150].

Accumulations of PPIX can cause skin photosensitivity, biliary stones, hepatobiliary damage and even liver failure, as in X-linked protoporphyria, where the function gain of ALAS[1]2 causes PPIX accumulation. [147]

PPIX synthesis occurs partially in mitochondria and cytoplasm as depicted in figure 5.1. PPIX in the cell can either be used for the synthesis of heme – requiring the presence of iron – or be excreted. The main regulation point in this pathway is the ALAS enzyme, with heme-dependent negative regulations at the gene, mRNA and protein level, and with additional metabolic and iron-dependent regulations. [147] For this reason, the accumulation of PPIX from 5-ALA[2] can be cell cycle-dependent and seeding concentration-dependent when working with cell lines.

5-ALA is not a common metabolite readily available to the cells, and thus when overfeeding 5-ALA the main cell regulation point is bypassed and intermediate products accumulate, including PPIX. This leads to excess PPIX being pumped out to the cytoplasm and excreted to the extracellular space, generally via the ABC[3] transporters. PPIX accumulation is photosensitizing due to absorption of light, reaction with oxygen and generation of ROS[4] that can then react with surrounding macromolecules. Cancer cells accumulate higher amounts of PPIX than normal cells due to reduced activity of ferrochelatase and slightly higher activity of the deaminase (HMBS[5]). [152]

[1] Aminolevulinic acid synthase
[2] 5-aminolevulinic acid
[3] ATP binding cassette
[4] Reactive oxygen species
[5] hydroxymethylbilane synthase

Previous experiments with CRC[1] cell lines and 5-ALA incubation using PPIX extraction and fluorescence spectroscopy showed that confluent cultures accumulate more PPIX than subconfluent ones when working with SW480 and HT29 cells, with values ranging between 150-200 to 20-50 ng of PPIX per mg of protein respectively when using concentrations of 5-ALA of up to 1.2 mM for 3 h. [153] When looking at the subconfluent results, SW480 cells showed higher uptake of 5-ALA than HT29 cells, but lower HMBS activity, resulting in an overall higher accumulation of PPIX for HT29 cells, that ranged around 30-50 ng of PPIX per mg of protein. [153] Gederaas *et al.* [154] obtained a similar PPIX concentration of around 30-40 ng per mg of protein for SW480 cells when incubated with 2 mM 5-ALA for 3.5 h. Brunner *et al.* [259] confirmed a similar trend for the two cell lines, that showed around 150 ng of PPIX per mg of protein for SW480 cells and 260 ng of PPIX per mg of protein for HT29 when incubating with 0.6 mM of 5-ALA for 3 h, but obtained higher accumulation of PPIX. Overall, these results suggest a strong dependence of PPIX accumulation with the metabolic state of the cells, showing an order of magnitude difference in accumulation of PPIX when using low or high confluent cultures. When using subconfluent cultures, the accumulation of PPIX in these CRC cell lines is of around 30-50 ng per mg of protein, resulting in an estimate of 0.003-0.005% weight concentration of PPIX in a cell, undetectable by Raman unless using RR[2].

5.1.1 Standards Raman Spectra

Stocks of 5-ALA, PPIX and cyt C were measured to obtain the Raman fingerprint of each of the molecules. cyt C and PPIX are expected to have very similar spectra, as they both contain the same coordination complex with or without the iron respectively, and both are expected to show strong RR when using the 532 nm laser. All standards were measured as detailed in subsection 3.3.3 (page 66).

Figure 5.2 shows the Raman spectrum of a 5-ALA HCl crystal (Sigma Aldrich), that is fully consistent with the Raman certificate provided by Sigma. However, Raman detection of 5-ALA is not expected on the cell samples as it is at a weight concentration of around 0.03% weight/volume (2 mM), and further washed after incubation, too low to detect with RS[3].

The vibrations of octaethylporphyrin were first assigned by Abe *et al.* (1978) [261] and, in general, the literature follows its original band assignment, where frequencies are referred to as ν_i. To include the effects of the vinyl groups, the work from Choi *et al.* [262–264] was also considered. Figure 5.3 shows the *fingerprint* spectrum of PPIX and figure 5.4 shows the *fingerprint* spectrum obtained from the cyt C standard. For cyt C band identification, the bonding of the porphyrin ring to the protein must also be considered, with peak assignment based on Smith *et al.*. [265] A table reproducing the assignments from the literature for each of the bands is shown in Table 5.1, where δ stands for in-plane bending, γ stands for out of plane modes and ν stands for stretching. The atoms referred to in each vibration are indicated in the insets in figure 5.3 and figure 5.4, with α atoms, β atoms and *meso* atoms

[1] Colorectal Cancer
[2] Resonant Raman
[3] Raman spectroscopy

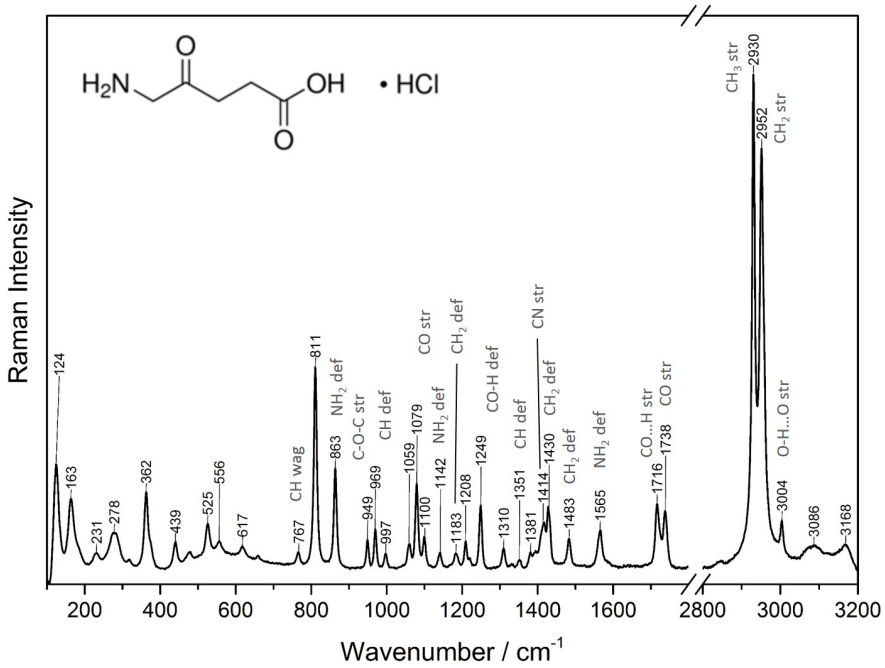


Figure 5.2: Raman spectrum of 5-ALA crystals obtained with the 532 nm laser with main peak identification based on Chen *et al.* [260].

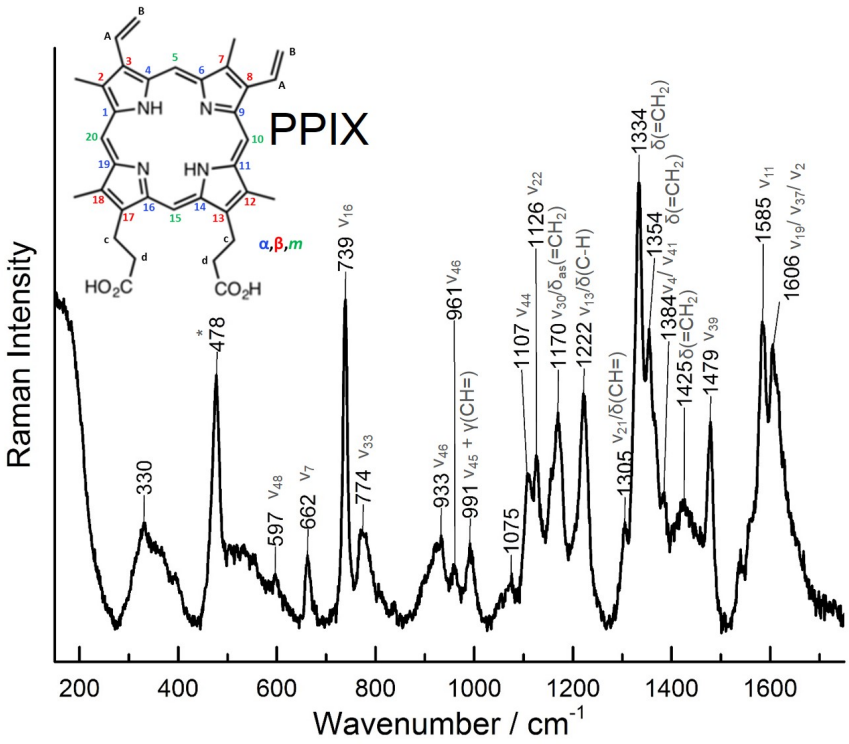


Figure 5.3: Raman spectrum of PPIX (inset) in solution obtained with the 532 nm laser with main peak identification based on Abe *et al.* (1978), Choi *et al.* (1982), Brunner *et al.* (1973) and Sarkar *et al.* (1986) [261–264]. The peak at 478 cm^{-1} labelled as * was caused by the glass substrate.

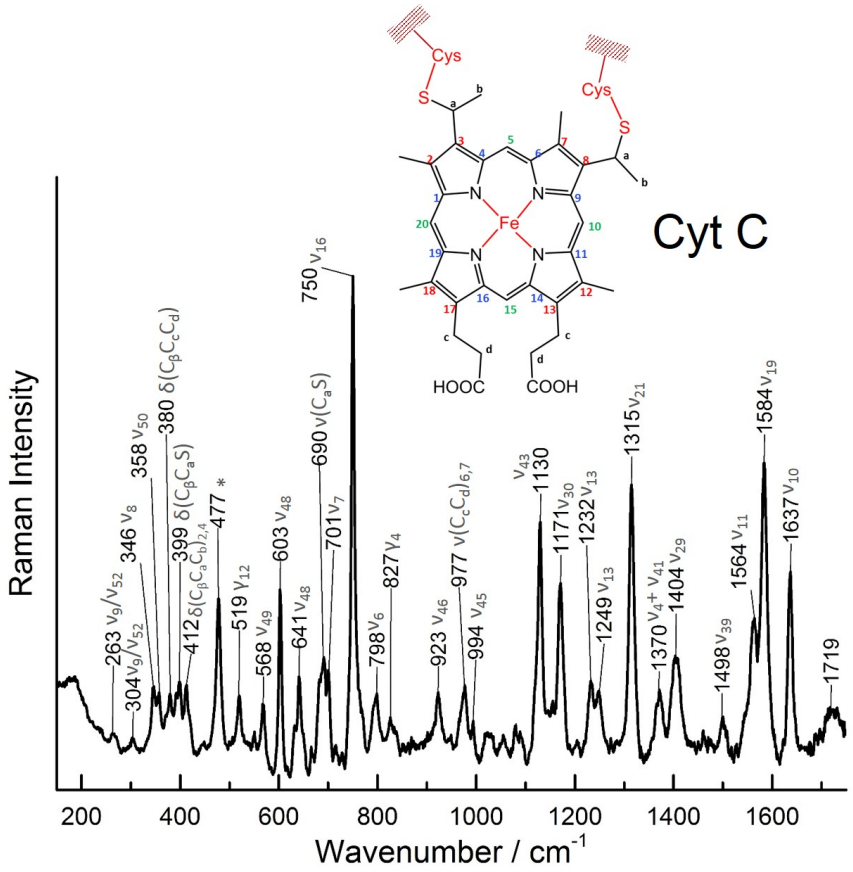


Figure 5.4: Raman spectrum of cyt C in DPBS[1] obtained with the 532 nm laser with main peak identification based on Abe *et al.* (1978), Choi *et al.* (1982) and Smith *et al.* (2005) [265]. The peak at 477 cm^{-1} labelled as * was caused by the glass substrate.

that belong to the ring structure (color coded); **A** and **B** for the nomenclature of the vinyls; **c** and **d** for the carbons in the propionyls, and **a** and **b** for the carbons in the bonding of heme with the Cys. *Et* stands for the peripheral ethyl group for the octaethylporphyrin, and ν' and δ' indicate antisymmetric stretching and deformations with respect to the C_2 axis of of the pyrrole ring.

Table 5.1: Peak labelling of cyt C and PPIX standard spectra based the original nomenclature of Abe *et al.* (1978) for the modes from the ring atoms and completed with the assignments from Choi *et al.* (1982) and Smidth *et al.* (2005) for the vinyl and the cytochrome specific modes respectively. [261, 262, 265]

Vibrational Mode	Assignment (and Intensity %)	Wavenumber (obs/calc) cm^{-1}	Obs. PPIX cm^{-1}	Obs. cyt C cm^{-1}	Ref.
ν_9	$\delta(C_\beta - Et)23, \nu(C_\alpha C_m)16$	226/230		263	[261]
ν_{52}	$\delta(C_\beta - Et)43, \nu(MN)29$	287/264		263, 304	[261]
ν_8	$\delta(C_\beta - Et)56, \nu(C_\alpha C_m)11$	344/326		346	[261]
ν_{50}	$\delta(C_\beta - Et)28, \nu(C_\alpha C_m)15$	.../333		358	[261, 265]
	$\delta(C_\beta C_c C_d)$			380	[265]
	$\delta(C_\beta C_a S)$			399	[265]
	$\delta(C_\beta C_a C_b)_{2,4}$			412	[265]
γ_{12}	pyr swivel	520		519	[265]
ν_{49}	$\delta'(C_\alpha C_\beta C_\beta)29, \delta(C_\alpha C_m C_\alpha)14$	550/546		568	[261]
ν_{48}	$\delta'(C_\beta - Et)29, \nu(C_\alpha C_m)13$	605/620	597	603, 641	[261, 265]
	$\nu(C_a S)$			690	[265]
ν_7	$\delta(C_\beta C_\alpha N)20, \nu(C_\alpha C_\beta)19$	674/655	662	701	[261]
ν_{16}	$\nu(C_\alpha N C_\alpha)14, \nu(C_\beta Et)14$	751/741	739	750	[261]
ν_{32}	$\delta'(C_\beta - Et)50, \delta'(C_\alpha - C_m)22$	785/789	774		[261]
ν_6	$\delta(C_\alpha C_m C_\alpha)36, \nu(C_\alpha N)27$	806/809		798	[261]
γ_4	fl(C_mH)	827		827	[265]
ν_{46}	$\nu'(C_\beta - Et)20, \delta'(C_\alpha - C_m)17$	924/895	933, 961	923	[261]
	$\nu(C_c C_d)$	975		977	[265]
ν_{45}	$\nu'(C_\alpha N)19, \nu'(C_\alpha C_m)16$	993/1007	991	994	[261]
	$\gamma(CH=)$	999	991		[262]
ν_{44}	$\nu(C_\beta - Et)29, \nu(C_\alpha C_\beta)26$	1113/1097	1107		[261]
ν_{22}	$\nu'(C_\alpha N)37, \nu'(C_\beta Et)26$	1121/1118	1126		[261]
ν_{43}	$\nu'(C_\beta - Et)38, \nu'(C_\alpha N)24$	1148/1139		1130	[261]
ν_{30}	$\nu'(C_\beta - Et)49, \nu'(C_\alpha N)28$	1159/1157	1170	1171	[261]
	$\nu(C_B C_A)$	1167	1170		[262]
ν_{13}	$\nu(C_\alpha N)63, \nu(C_\beta C_\beta)13$	1220/1262	1222	1232, 1249	[261]
ν_{21}	$\delta'(C_m H)53, \nu'(C_\alpha C_\beta)18$	1308/1281	1305	1315	[261]
	$\delta(CH_2)$	1343	1334, 1354		[262]
ν_{41}	$\nu(C_\alpha N)50, \delta(C_\alpha C_m)11$	1389/1662	1384	1370	[261]
ν_4	$\nu(C_\alpha N)53, \delta(C_\alpha - C_m)21$	1383/1386	1384	1370	[261]
ν_{29}	$\nu'(C_\alpha C_\beta)47, \nu'(C_\beta - Et)26$	1409/1409		1404	[261]
ν_{40}	$\nu'(C_\alpha C_\beta)40, \nu'(C_\beta - Et)14$	1443/1454	1425		[261]
	$\delta(=CH_2)$	1434	1425		[262]
ν_{39}	$\nu'(C_\alpha C_m)39, \nu'(C_\alpha N)16$	1487/1498	1479	1498	[261]
ν_{11}	$\nu(C_\beta C_\beta)57, \nu(C_\beta - Et)16$	1579/1587	1585	1564	[261]
ν_{37}	$\nu(C_\alpha C_m)36, \nu'(C_\alpha C_m)24$	1604/1633	1606		[261]
ν_{19}	$\nu'(C_\alpha C_m)67, \nu'(C_\alpha C_\beta)18$	1603/1600	1606		[261]
ν_2	$\nu(C_\beta C_\beta)60, \nu(C_\beta - Et)19$	1602/1591	1606		[261]
ν_{10}	$\nu(C_\alpha C_m)49, \nu(C_\alpha - C_\beta)17$	1655/1656		1637	[261]

cyt C contains a heme group – i.e., the PPIX with an iron atom chelated at the centre – and binds to the protein via the two methylene groups. Some peaks are expected to be the same for PPIX, e.g. the general ring modes, but some methylene peaks are expected to disappear and some other peaks are expected to shift due to a change in the environment when bonded to the protein. In fact, cyt C is an example of a protoporphyrin for which the protein distorts the porphyrin significantly. [266] Some peaks common to both spectra

showed shifts to higher wavenumbers for cyt C compared to PPIX, such as the ν_{16} vibration at 750/739 cm^{-1} peak or the ν_{21} 1315/1305 cm^{-1} (cytC/PPIX). The $=CH_2$ bending at 1334 for PPIX disappears for cyt C. The out of plane ring deformation at 1584/1585 cm^{-1} (cyt C/PPIX) is exhibited by both compounds in more or less the same location, but with higher intensity for cyt C. Both the PPIX and the cyt C solutions showed a strong fluorescent baseline when using the visible laser, so higher signal to noise ratios could not be achieved, and the contributions to the CH stretching regions of the spectra were omitted from the comparison.

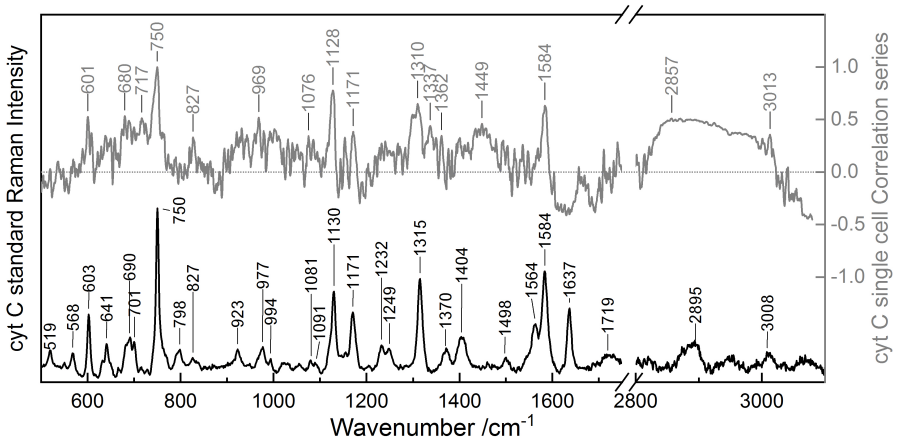


Figure 5.5: Comparison of the cyt C standard and cyt C live single cell correlation series from the data from chapter 4.

To verify the assignment of the cyt C peaks in the single cell spectra, both the cyt C signal and the 749 cm^{-1} correlation series from the live single-cell data obtained in Chapter 4 are shown in figure 5.5. Most of the stronger peaks showed an equivalent in the correlation spectra with almost complete overlap. The 1637 cm^{-1} band is shifted to 1657 cm^{-1}, likely due to the negative correlation of cyt C with the Amide I band, giving a strong negative contribution in this region that shifts the band to higher wavenumbers. Positive correlation with the 1449 cm^{-1} CH stretching band is also present in the correlation spectra.

The spectrum of cyt C and PPIX have significant similarity, making their independent detection in cell samples very challenging. To illustrate the expected spectral changes when both the cyt C and the PPIX are present, figure 5.6 is showing both spectra added together with the same intensity in the ν_{16} 750/739 cm^{-1} peak. Some of the stronger candidates for tracking PPIX such as the 739 cm^{-1} or the 1584 cm^{-1} peaks might prove difficult to track within a cell due to almost complete overlap with cyt C bands. Some other candidates are also highlighted like the 1222 cm^{-1} band - that may overlap with the side of the Amide III band - or the 1479 cm^{-1} band - that will fall on the side of the strong CH stretching band at around 1455 cm^{-1}. As some of these bands are really close to each other, it was opted to avoid smoothing the spectra in the pre-processing steps to avoid losing any resolution in the position and width of the bands.

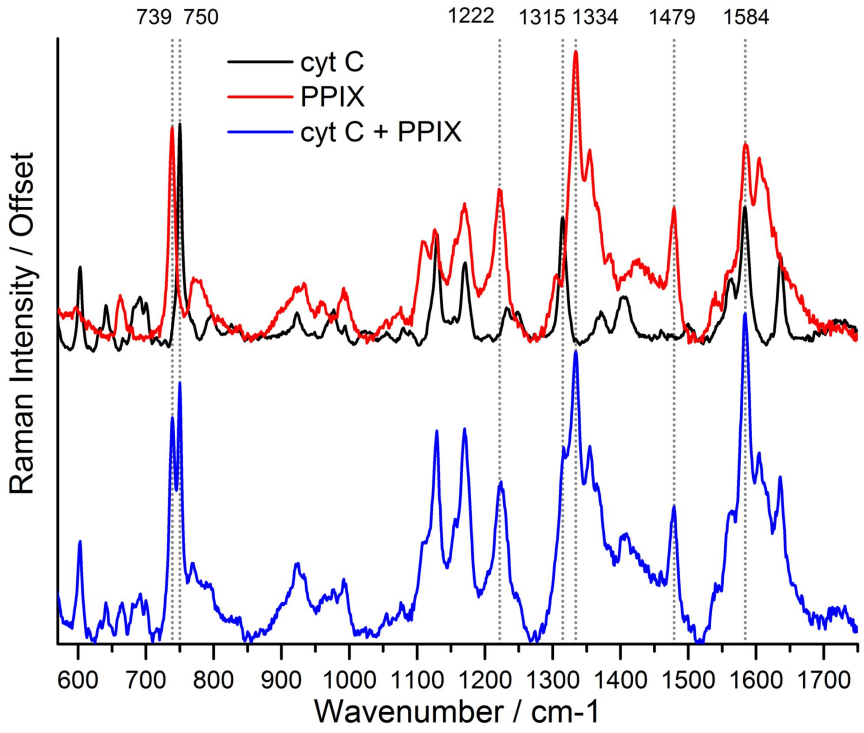


Figure 5.6: Overlap of cyt C and PPIX spectra, where the spectra of each compound has been shown with an offset, and the simulated spectrum obtained when adding these spectra with equal weighting is shown.

5.1.2 Raman on PPIX Expressing Single Adenocarcinoma Cells

When measuring the 5-ALA treated cells, some cells spontaneously developed blebbing after being exposed to the 532 nm laser light, indicating that not only was the Raman measuring the cell but also potentially activating the PPIX photodynamic activity. This did not happen for control cells and did not always happen for all cells measured, which could indicate heterogeneous accumulation of PPIX. Treated cells showed a fluorescent baseline of the Raman spectrum, that although could be baseline subtracted, added considerable noise to the treated data. Figure 5.7A shows the averages for the datasets obtained from multiple experiments, where all the datasets contained more than 100 single-cells spectra. A first observation of the spectrum showed that no obvious peaks could indicate the presence of the accumulated PPIX. The difference spectrum is shown in figure 5.7B. For both figures, the main peaks highlighted as promising for PPIX quantification are shown as vertical lines.

For SW480 cells, multiple peaks showed changes along the spectrum following 5-ALA incubation. Related with the cyt C and PPIX spectrum, a positive peak is present at 747 cm^{-1}. There was a positive contribution at 1218 and 1231 cm^{-1}, potentially showing a increase in the 1222 cm^{-1} PPIX band. A broad band around 1310 cm^{-1} showed an increase, overlapping with the 1315 cm^{-1} band. An increase in the 1451 cm^{-1} broad band with a side peak at 1479 cm^{-1} could be showing an increase in the 1479 cm^{-1} PPIX band.

For HT29 cells, a negative band at 747 cm^{-1} is also present, indicating a drop in cyt C. This cell line also showed a drop in the 1128 cm^{-1} peak, a slight drop in the 1222 cm^{-1} peak, and negative contribution at 1315 cm^{-1}, but positive contributions at 1479 cm^{-1} - with a drop

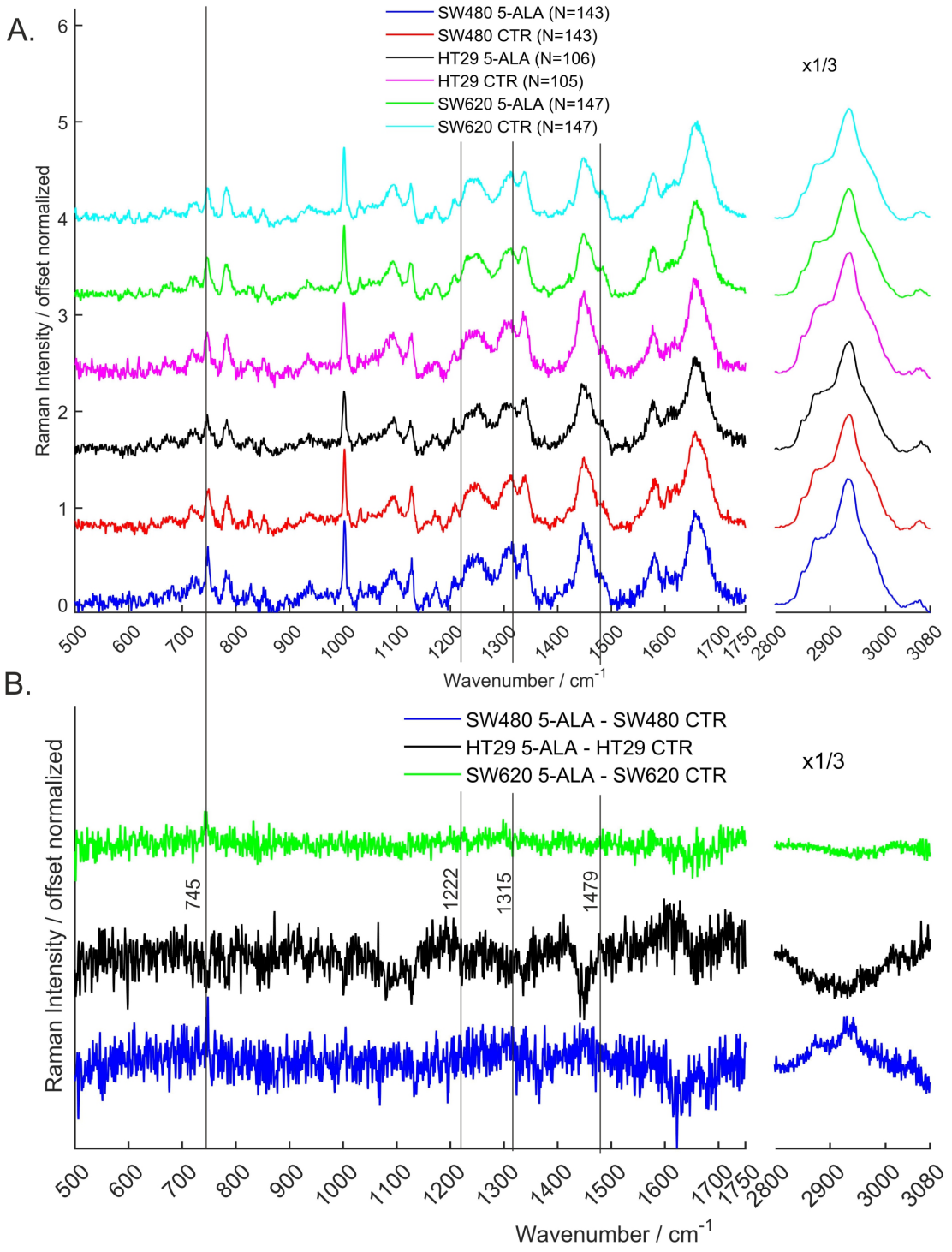


Figure 5.7: **A.** Average of the 5-ALA treated cells and controls for the three cell lines for the fingerprint region and the CH stretching region. No clear peaks can be observed as major indicators of PPIX presence. **B.** Difference spectra for the 5-ALA treated cells and control cells for each cell line. Main peaks of interest highlighted in figure 5.6 are indicated with vertical lines.

in the 1451 cm^{-1} band. Other interesting peaks were the negative Phe peak at 1002 cm^{-1}, the positive contributions on the side of the Amide I band around 1607 and 1630 cm^{-1} and the drop in CH stretching contributions at high wavenumbers.

For SW620 cells, the changes were smaller when normalising to the Amide I band, with a clear positive peak at 745 cm^{-1}, and positive peaks at 1224 and 1300 but neutral contributions at 1479 cm^{-1}.

Overall, the average analysis was unclear, with no striking differences between the control and the 5-ALA treated cells. This might suggest that single-cell analysis is essential for understanding the population dynamics in PPIX accumulation. To understand the cyt C and PPIX peaks present in the sample at a single cell level, with their actual positions for the single-cell measurements, all the strong peaks from the spectra of the standard for cyt C and PPIX were analysed for correlation.

5.1.3 Correlation Analysis for Separation of Cytochrome C and Protoporphyrin IX Contributions

When calculating the correlation matrix for the single cell acquired data – calculated as in Chapter 4 –, two distinct correlation series were identified, and they are plotted in figure 5.8 corresponding to the series for the 601.5 cm^{-1} peak – assigned as the 603 cm^{-1} cyt C standard peak – and the 660.5 cm^{-1} peak – assigned as the 662 cm^{-1} PPIX standard peak. Both peaks appeared shifted to lower wavenumber in the single-cell correlation spectra compared to their positions in the standard sample spectra. The average single cell spectra showed a distinct peak at 601 cm^{-1}, but the 661 cm^{-1} peak only appeared as a shoulder of the stronger 669 cm^{-1} peak. Both peaks showed correlation series with similar bands, but either peak was absent for the other peak's correlation series, indicating that they were not correlated to each other and that they corresponded to different independent spectra overlapped. Some of the correlation bands showed a slight shift between the two correlation series, like the 1123/1128 cm^{-1} or the 1315/1310 cm^{-1}. However, due to the noisy nature of the spectra, it was difficult to separate their contributions. This suggested that even though the signal is very weak, the spectra showed signs of the presence of PPIX.

5.1.4 PCA Analysis of the 5-Aminolevulinic Acid Data

PCA[1] was run on the single cell data from all cell lines to obtain the axes of maximum variability. PCA may identify the PPIX accumulation markers among other variability sources such as artefacts due to variability in sample preparation or changes in background contributions. However, PCA is not a good technique when peaks from the same molecule shift due to changes in environment, as it considers all wavenumbers as independent, and could confuse the interpretation. When running PCA on the data, the first 15 coefficients only explained 24% of the total variability of the system, showing contributions in the main cyt C

[1] Principal Component Analysis

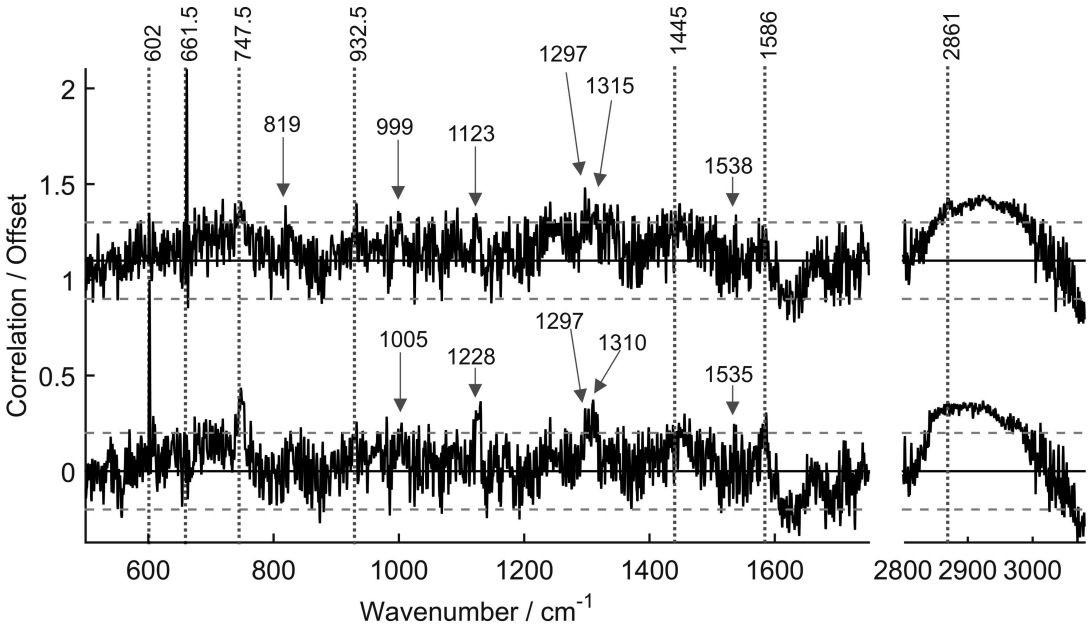


Figure 5.8: Correlation series for the single-cell data for peaks at 601.5 cm^{-1} and at 660.5 cm^{-1} assigned as the 603 cm^{-1} cyt C standard peak and the 662 cm^{-1} PPIX standard peak respectively.

bands with slight shifts to higher or lower wavenumber, indicating changes in bands position and shape.

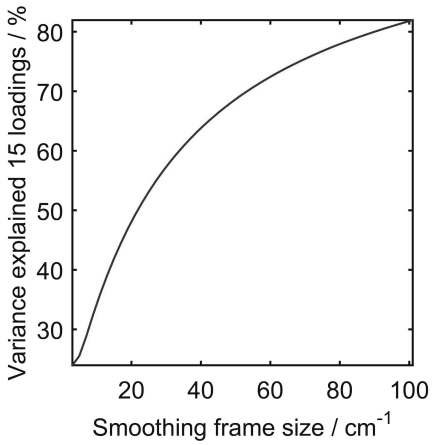


Figure 5.9: PCA variance explained for 15 PCs when using different smoothing frame sizes in the data.

An option to minimise these effects for quantification purposes is to smooth the spectra before running PCA, obtaining broader bands and losing resolution, but reducing sensitivity to noise. Figure 5.9 shows the variance explained when using different smoothing frame sizes with a Savitzky-Golay filter with a quadratic polynomial. A gentle smoothing of the spectrum (quadratic SG filter with a frame of 8.5 cm^{-1}) increased the variance explained for 15 coefficients from 24% to 45%, with the first 5 coefficients explaining 33% of the variance. Variance explained increased rapidly for frames for up to 25 cm^{-1} and then kept increasing at a slower pace for broader frame sizes.

For consistency with data from chapter 4, the frame size was set to 8.5 cm^{-1}, and PCA analysis was repeated, where the loadings are shown in figure 5.10. PC1 showed the classic shape for CRC cell lines, with a similar contribution in the CH stretching region to the data obtained on chapter 4 for figure 4.10A on page 101. PC2 had a similar shape to PC1, with

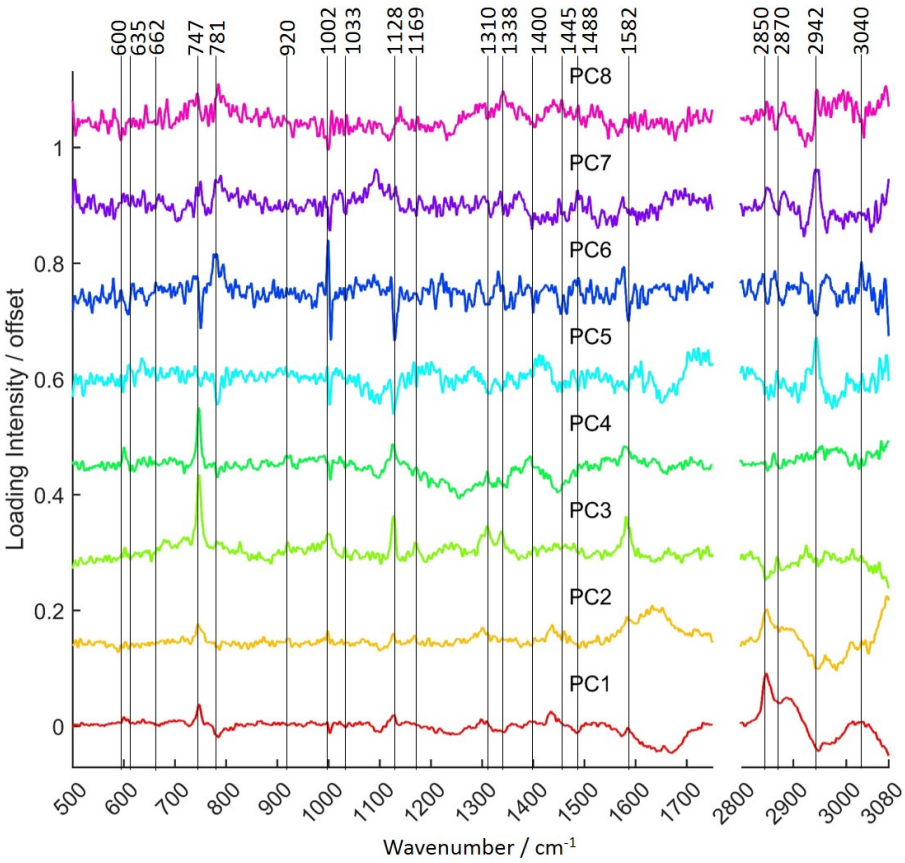


Figure 5.10: Loadings for the first 10 PCs when analysing the 5-ALA and control samples data for the three cell lines after smoothing the data. Even though multiple PCs showed peaks in similar positions, the effect of small bands shifts is minimized, allowing a more representative dimensionality reduction when analysing 10 PCs.

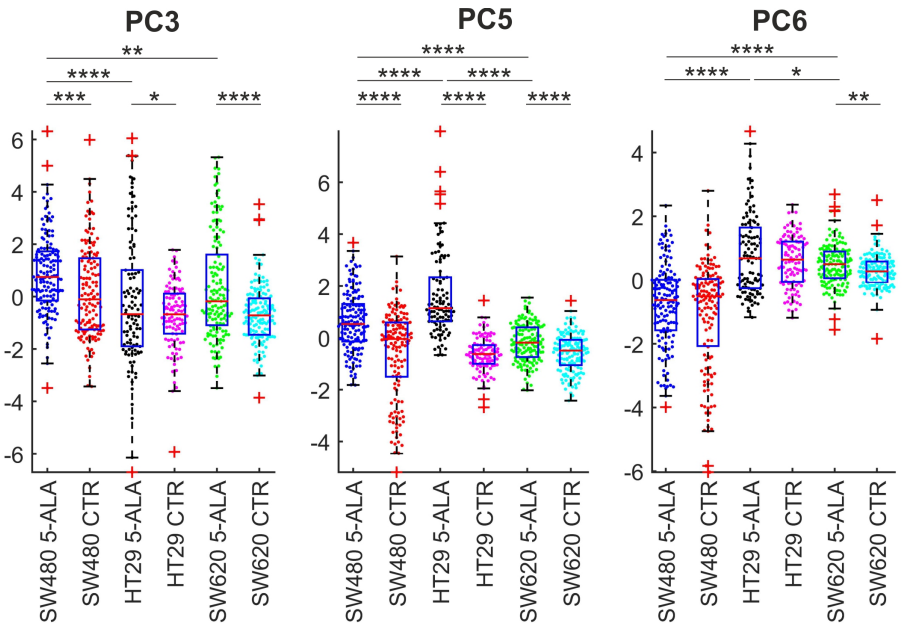


Figure 5.11: Beeswarm and boxplot of the PCA scores for the most relevant PCs for PPIX detection. Significance level between classes has been indicated. PC3 and specially PC5 shows significantly higher scores for the treated cells compared to the controls for all cancer stages. (*:p< 0.05,**:p<0.01,***:p< 10^{-3},****:p< 10^{-4}).

just a few differences such as a broad positive contribution around the Amide I area. PC3 and 4 showed mainly strong cyt C peaks, with differences in the Phe peak in the region around 1400 cm^{-1}. PC5 showed mainly negative contributions on cyt C related peaks, with positive peaks for the bands at 739 and 1478 cm^{-1}, but negative peaks at 1127, 1223 and 1585 cm^{-1}. Other bands from this component showed a strong negative peak at 783, 1002 cm^{-1}, associated with DNA and Phe, and a strong peak at 2944 cm^{-1} in the $-CH_3$ stretching region. The only component that showed the presence of the 662 cm^{-1} peak was PC6, with positive to negative sharp contributions on most cyt C peaks, indicating a shift to lower wavenumbers.

When looking at the scatter plots of PC1, PC2 and PC4 (data not shown), no separation between groups can be observed. These components were seen to account mainly for within-sample variability. The scores for PC3, PC5 and PC6 have been plotted on figure 5.11. These three components showed overall higher scores for the 5-ALA samples compared to the control samples. The significance of these differences was calculated using a two-sample t-test, considering only each cell line control/5-ALA pair and differences between 5-ALA treated samples between different cell lines. Only PC3 and PC5 showed significant differences between 5-ALA and control samples for all the cell lines. This could indicate an overall drop of cyt C levels in these cells due to photodynamic activation with the Raman laser excitation line, with very difficult to detect contribution of PPIX indicated with subtle peaks at 739 and 1478 cm^{-1} that degraded during the measurement. The assignment for the contribution at 2942 cm^{-1} is unknown, but seeing that both the PPIX and the cyt C correlation series (figure 5.8, page 121) showed a strong correlation with the CH stretching region – with a shifted contribution towards higher wavenumbers for the PPIX – this could be a CH stretching contribution of the PPIX molecule, that could not be measured in the standards due to the high fluorescence baseline of the concentrated sample. This band is also present in PC6, that also showed higher scores for 5-ALA treated cells than for control cells, although the difference was not significant for SW480 and SW620 cells.

5.1.5 Protoporphyrin IX Detection Using Fluorescence

After each experiment, confocal fluorescence images from adhered cells prepared in parallel to the measured cells were acquired to confirm PPIX expression. PPIX excitation spectrum is well known, with a Soret excitation band at 405 nm and multiple Q bands at 505 nm, 540 nm, 575 nm and 630 nm. [267] The second Q band is the one responsible for the RR effect when using the 532 nm laser. Confocal images for the three different cell lines are shown in figure 5.12. PPIX shows a very weak and noisy fluorescence on the sample but stronger excitation laser intensities could activate the photosensitising effect. In general, HT29 cells showed stronger fluorescence than SW620 and SW480 cells. Control cells showed no fluorescence using the same acquisition settings, confirming the accumulation of PPIX when inducing its synthesis using 5-ALA.

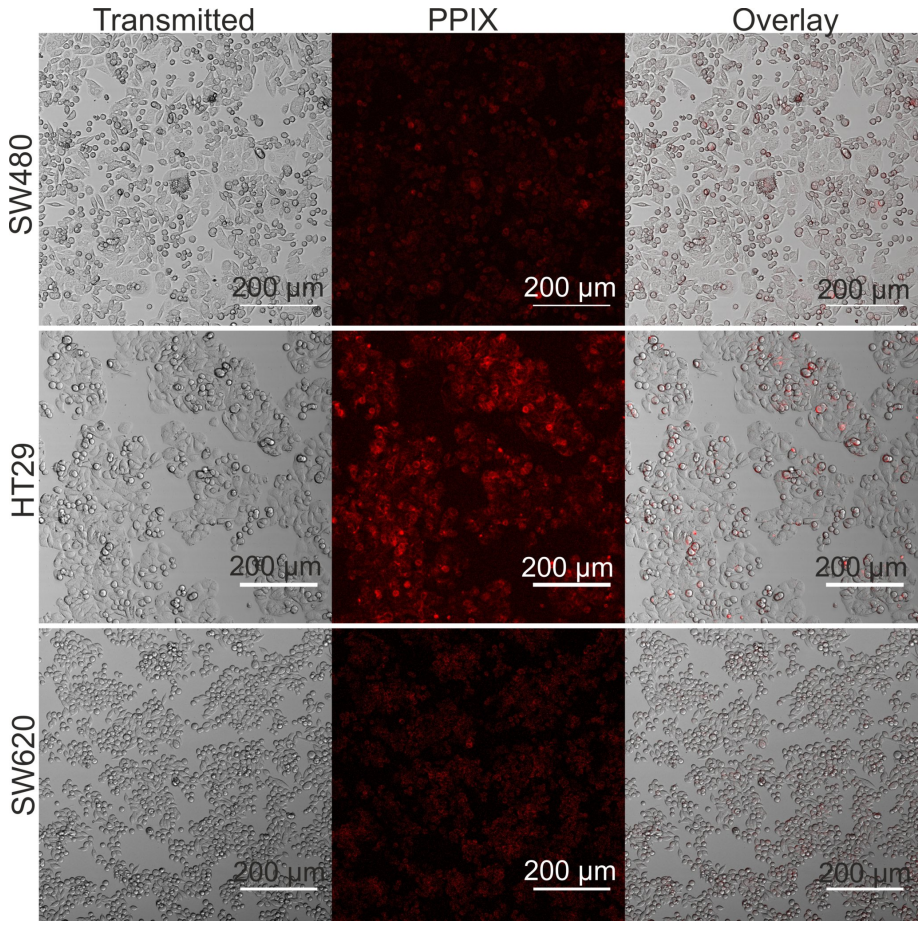


Figure 5.12: Fluorescence of PPIX in colorectal cancer cell lines SW480, HT29 and SW620. Fluorescence has been enhanced equally in all images to increase visibility in the printed version.

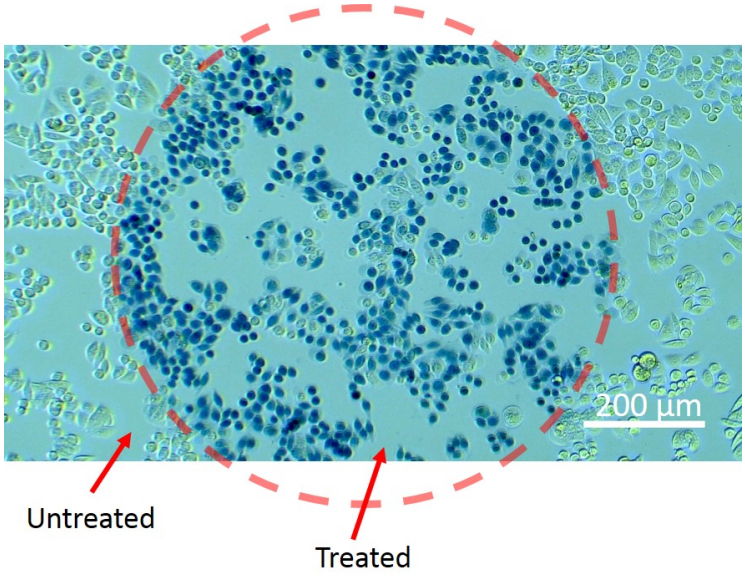


Figure 5.13: SW480 cells death after PDT with PPIX measured with Trypan Blue, where PPIX caused the cells to permeate trypan blue, measuring the loss of viability in terms of increased membrane permeability. Blue cells indicated dead cells, and bright cells indicated viable cells. The blue circle in the control well could not be found - presumably due to minimal or low loss of viability.

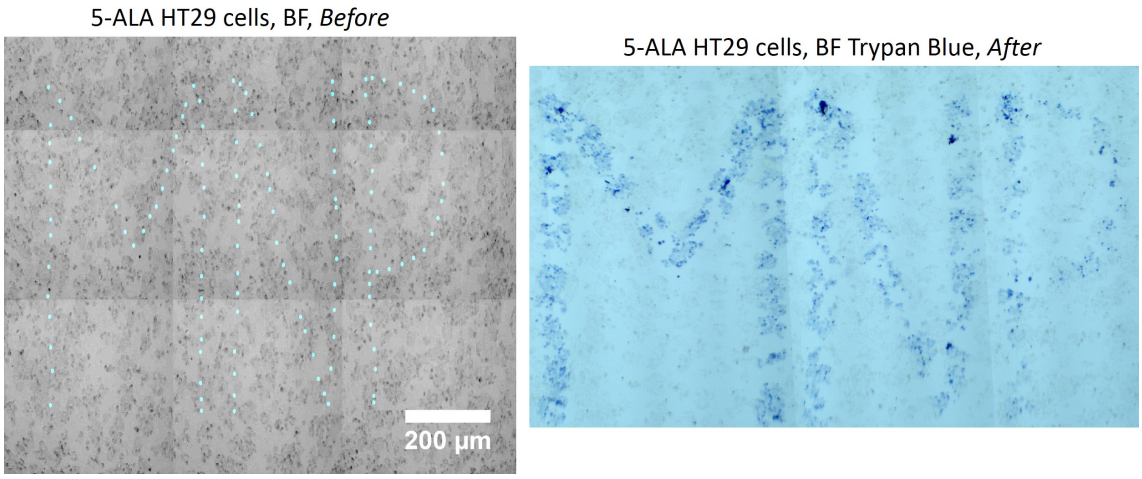


Figure 5.14: PPIX-death patterning with UV light on HT29 cells, where each point in the bright field image (in light blue) was exposed to a dose of ≈4.4 J/cm^2 and a spot size of 95 μm diameter. Cells were subsequently incubated with Trypan Blue and washed with DPBS to show the patterned dead cells.

5.1.6 Photodynamic Activation of Protoporphyrin IX

Cells were prepared as detailed in subsection 3.6.4 on page 76, and one of the wells was not detached and was imaged and treated with UV[1] light (340-380 nm, Intensity of ≈219 mW/cm^2, total dose of ≈13 J/cm^2). Cells were then incubated with trypan blue, that was then washed with DPBS, allowing measuring viability. Figure 5.13 shows a circle of treated cells compared to the untreated cells around it. A control sample without 5-ALA was exposed for the same amount of time and treated with Trypan Blue, but the circle of dead cells could not be found in the microscope field, presumably due to no membrane integrity damage at this UV dose. In another example (see figure 5.14), HT29 cells were killed following a pattern spelling *MNP* (Molecular and Nanoscale Physics Group), by spending 20 s in each point (340-380 nm, total dose of ≈4.4 J/cm^2).

Additionally, time effects in the morphology of the cells were tracked in the confocal fluorescence microscope. Cells were imaged before and after 1 min of 100% intensity light with the 10× objective (340-380 nm, Intensity of ≈219 mW/cm^2, total dose of ≈20 J/cm^2). Images of the cell morphology changes can be found in figure 5.15.

5.1.7 Conclusions and Challenges

Microscopy results, including fluorescence of treated cells and Trypan Blue staining showing photosensitivity of 5-ALA treated cells compared to the controls, confirmed the expression of PPIX by the model system cell lines. Even though data, in general, showed changes between the 5-ALA treated and untreated cells, it showed no clear bands for detection of the pure PPIX signal. This joined with the morphology changes seen for some of the cells caused by the measuring laser could be indicating that the laser was activating the porphyrin photodynamic effect and thus detecting it for just the initial part of the total acquisition time, lowering its contribution. Overall, 5-ALA samples showed significantly higher contributions on PC5 with

[1] Ultraviolet

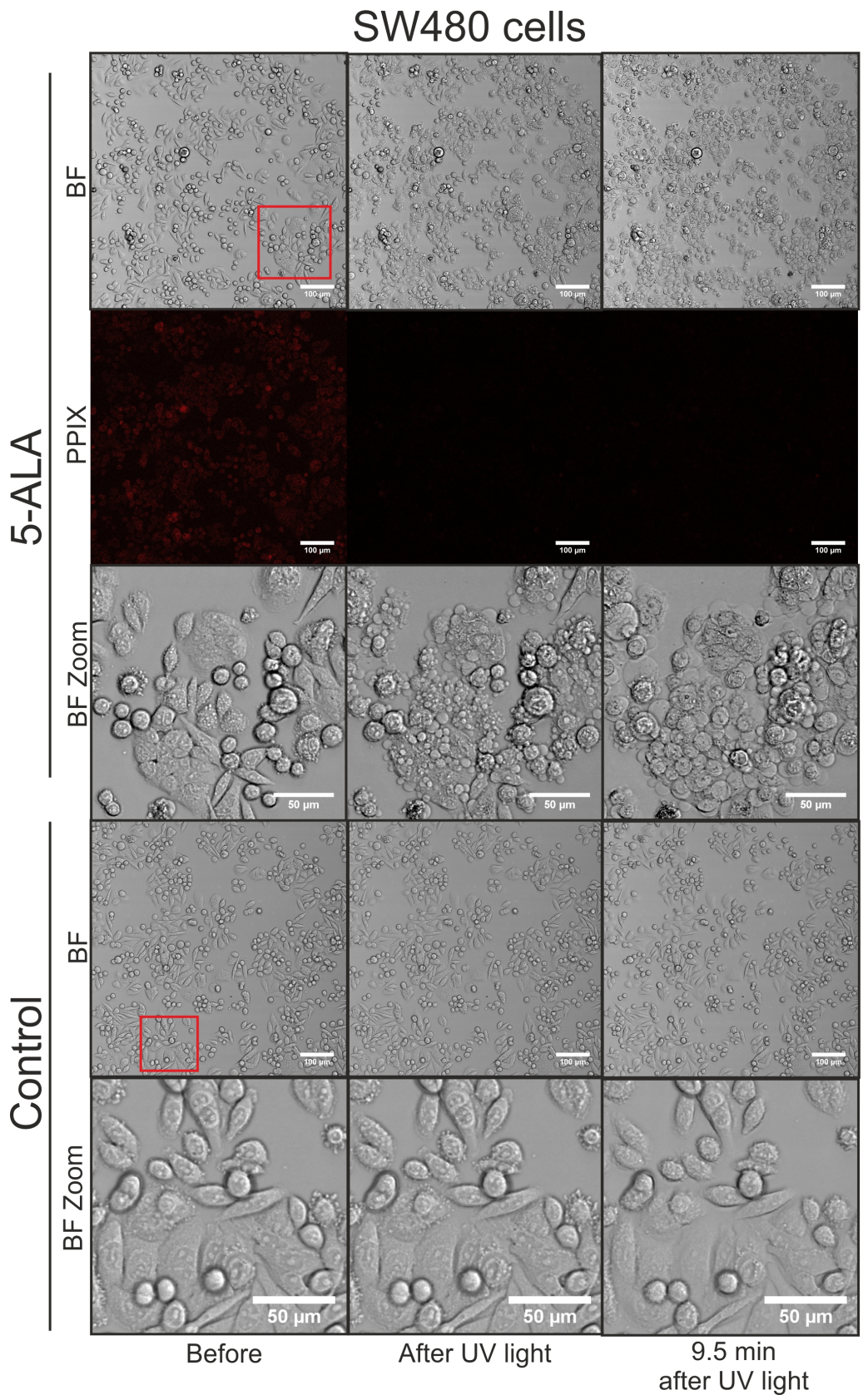


Figure 5.15: Effects of PPIX activation with UV light (148 J/cm^2) of 5-ALA incubated or control cells over time. Fluorescence was lost for 5-ALA cells after UV light exposure, with strong morphology changes already visible right after UV light, where the zoom showed that previously attached cells appear rounded and covered in small membrane blebs, that became bigger after 9.5 min, with clear cell damage. Control cells exposed to the same amount of UV light showed no visible changes after 9.5 min.

HT29 > SW480 > SW620 differences between sample and control, and overall a drop in cyt C bands in 5-ALA treated cells, that were likely undergoing either apoptosis or necrosis due to the photodynamic treatment. Results indicated a potentially higher accumulation of PPIX in HT29 cells compared to SW480 and SW620 cells, evidenced by the significant differences in PC5 scores. This agrees with previous observations by Brunner *et al.* and Kriegg *et al.* with respect to SW480 and HT29 cells [153, 259], showing additionally that SW620 cells showed the lowest accumulation. However, these scores are not normalised to the overall protein contents of the cells and are difficult to relate to the overall accumulation.

5.2 Novel Photodynamic Agent DC473 Accumulation

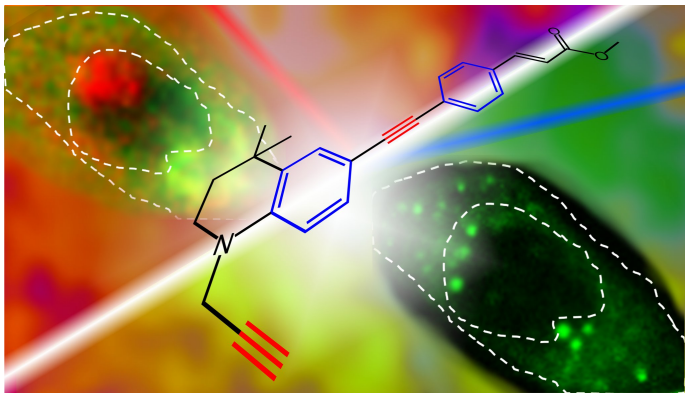


Figure 5.16: Table of contents figure for Raman and fluorescence confocal mapping of DC473 within a single cell, Gala de Pablo *et al.* in *The Analyst* [2], published in the SPEC2018 collection.

Imaging and tracking bioactive compounds in vivo is essential to understanding their mechanism of action. However, tagging them while retaining their inherent biological activity and function is very difficult, in particular when large fluorescent molecules such as the 238 residue GFP[1] are employed as labels. [268] An inherently fluorescent bioactive molecule that does not require labelling is, therefore, highly desirable. However, there are other issues to consider when using fluorescence for quantitative molecular tracking, including photobleaching during the experiment, [268] solvatochromatic or fluorogenic effects in different locations within the cell and non-linear responses from the fluorophore due to localised concentrations or self-absorption. [243] For this reason, novel and complementary techniques for the detection and characterisation of a bioactive molecule within a cellular environment are needed, where RS[2] could be a great asset. However, the Raman signal intensity is several orders of magnitude weaker than fluorescence or elastic scattering signals making the technique inherently slow and lacking sensitivity. [111]

To overcome these drawbacks, and to obtain orthogonal information provided by Raman measurements, multimodal imaging systems can be used. [111] Fluorescence is usually a problem in Raman spectroscopy, as it can be excited when using visible lasers and cause significant background signal. Much effort has, therefore, gone into the design of Raman probes to reduce interference from fluorescence. [111] Since 2003, some groups have adopted the multimodal approach for increasing the tissue-diagnosis sensitivity of Raman probes by combining Raman with tissue auto-fluorescence. [111, 269, 270] Fluorescence tends to be used

[1] Green Fluorescent Protein
[2] Raman spectroscopy

either as a confirmatory technique for the Raman mapping or as a quick mapping technique to choose the areas of interest prior to Raman imaging in a large sample. [111] Some researchers have used both techniques in a complementary manner, such as the work of Carney *et al.* on extracellular vesicles using LTRS[1] and fluorescence labelling with CD9+ antibodies. [271] Raman and fluorescence have also been shown to provide complementary information for studying pH effects of doxorubicin uptake; [272] for tissue imaging using SERS[2] and fluorescence-based endoscopy, [75] and for dual drug release studies, using both SERRS[3] and energy transfer to probe length scales. [273] Researchers that have combined fluorescence and Raman include Zeng *et al.*, Manen *et al.* and Uzunbazakava *et al.*. In the work from Zeng *et al.*, they tracked autofluorescence lifetime and SRS[4] of naturally occurring lignin molecules in poplar cells. [274] Manen *et al.* tracked both the RR[5] signal of flavocytochrome b558 and the fluorescence of labelled quantum dots in single live neutrophils using the same excitation laser. [275] Uzunbazakava *et al.* investigated DNA using spontaneous RS and two-photon fluorescence of the DNA[6] dye Hoechst 33342 [270].

Designing a molecular probe that can be imaged by both tandem fluorescence and Raman scattering could give additional information in the location and environment of the probe. Following this approach, Li *et al.* designed a mitochondrial probe for both SRS and aggregation induced emission fluorescence in HeLa cells. [276] However, to our knowledge, designing a potential chemotherapeutic compound that can be imaged by both fluorescence and specific spontaneous RS in tandem (fluoRaman) has not been shown previously.

The vibrational signals associated with chemotherapeutic drugs often occur within the so-called *fingerprint* region of the spectrum, that also contains the vibrational information in both cells and tissue. [123] Consequently, the reliable and confident analysis of drug distribution in cells can be extremely challenging due to overlap with other cell molecules. Designing bioactive compounds with Raman signatures in the cell-silent region (1800-2800 cm^{-1}) can enable detection of the compound more accurately in cellular samples. [121] Alkynes are one such example, and typically exhibit a vibrational mode of ca. 2125 cm^{-1}, with intrinsically sharp peaks (FWHM[7] 14 cm^{-1}). [277] Alkyne labels (C $\equiv$ C) for RS were introduced in the last decade by Yamakoshi *et al.* [124, 125] and alkyne palettes have been developed by Chen et al. using C^{13} and C^{12} isotope combinations [277] and by Hu *et al.* with their Carbow. [122] These probes show significant potential for RS diagnostic in cancer therapy.

PDT[8] utilises PS[9]s to induce the production of large quantities of ROS[10] for the targeted destruction of diseased tissues and has been used to treat a variety of cancers. [278] However, current PSs approved for therapeutic use suffer from a wide variety of issues, including poor pharmacokinetic properties due to their high molecular weight, often porphyrin structures, and ineffective targeting leading to the death of healthy tissue. [278–280] DC473

[1] Laser tweezers Raman spectroscopy
[2] Surface enhanced Raman spectroscopy
[3] Surface-enhanced resonance Raman scattering
[4] Stimulated Raman scattering
[5] Resonant Raman
[6] Deoxyribonucleic acid
[7] Full-width half maximum
[8] Photodynamic Therapy
[9] Photosensitiser
[10] Reactive oxygen species

is a new low molecular weight, small PS, that elicits cell death through the stimulated production of ROS when activated by UV[1]-A, violet or corresponding near-IR[2] two-photon wavelengths, whose patent is commercialized by LightOx Ltd. [159–161] DC473 is a substituted diphenylacetylene with a highly lipophilic, electron rich tetrahydroquinoline donor and a methyl cinnamate acceptor.[160]⋆ DC473 exhibits strong solvatochromatic fluorescence and possesses two alkynes with Raman vibrational peaks in the silent region of the spectrum, ideal for RS of single cells. To improve this photosensitising activity, a precise understanding of the compound's cellular localisation is necessary.[160]⋆

In this section, the cellular localisation of DC473 in fixed SW480 CRC[3] cells was investigated using confocal fluorescence and Raman (fluoRaman) microscopy, proving co-location of the compound signal with both techniques, and obtaining the Raman signature of the organelles where it localises within the cell. The accumulation of this compound was studied in live cells in the model system of cell lines SW480, HT29 and SW620. The changes in Raman signals of the compound after photoactivation was then studied in live cells, in an attempt to elucidate its mechanism of action from a vibrational spectroscopy perspective.

5.2.1 DC473, EC23 and DC271 Raman Spectra

Three different compounds were provided by Lightox Ltd. for Raman analysis: EC23, DC271 and DC473. EC23 is a synthetic retinoid derived from natural retinoid ATRA[4] with enhanced photostability [281], that has been shown to induce neural differentiation and to have competitive binding to RAR[5] and RXR[6] receptors [282] and its effects on stem cells were analysed with IR [283]. DC271 is an active synthetic retinoid with additional fluorescent properties, that showed localization in the nucleus [161]. DC473, however, does not show active retinoid properties, with no binding to RAR or RXR, but shows solvatochromatic properties with potential as a PS. As all chemicals were provided in DMSO[7], a spectrum of DMSO was acquired (see figure 5.17) and peaks were labelled based on Martens *et al.*, with labels on table 5.2. [284]

Figure 5.18 shows the 785 nm Raman spectra of the crystal forms when dried onto quartz of EC23, DC271 and DC473. The three molecules share an overall similar chemical structure, with a tolan central group (diphenylacetylene) with different substitutions. As the molecules are very similar, obtaining the spectra of all the molecules can not only help with the choice of the molecule of interest but also help assign the main bands related with the tolan groups and the band related with the different substitutions. The positions of the main bands and their assignments are shown in table 5.3. The alkyne showed a very distinct band around 2200 cm^{-1}. The band was measured at 2197 cm^{-1} for DC473, at 2195 cm^{-1} for DC271 and at 2203 cm^{-1} for EC23. Alkyne peaks usually appear at ca. 2125 cm^{-1} [277]; however, the electron-rich tetrahydroquinoline donor group causes a shift to 2197 cm^{-1}. [122] N-Propargyl

[1] Ultraviolet
[2] Infrared
[3] Colorectal Cancer
[4] All trans-retinoic acid
[5] Retinoic acid receptor
[6] Retinoic X receptor
[7] Dimethyl sulfoxide

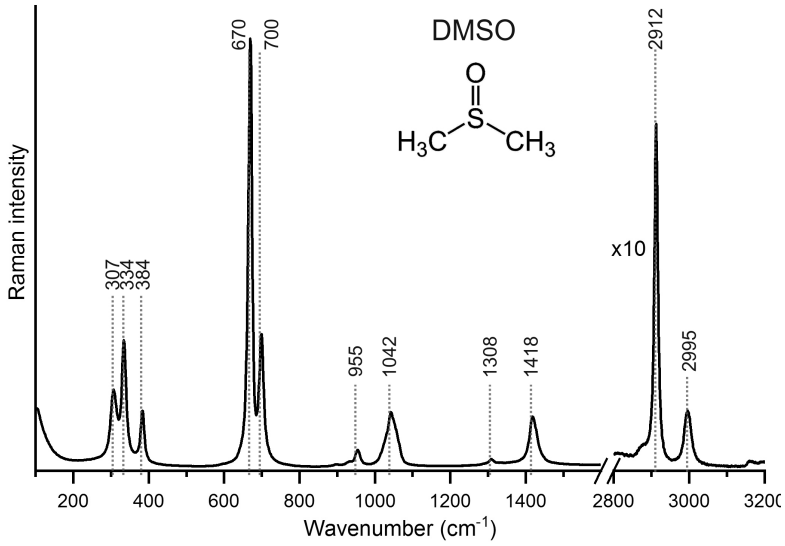


Figure 5.17: Spectrum of DMSO when exciting with the 785 nm laser line, where main bands have been labelled.

Table 5.2: Peak labelling of DMSO spectrum based on of Martens *et al.* (2001). [284]

Assignment (and Intensity %)	Wavenumber cm^{-1}	Literature [284] cm^{-1}
C-S-C out-of-plane bend	307	309
C-S=O out-of-plane bend	334	333
C-S=O in-plane-rock	384	382
C-S symmetric stretch	670	667
C-S antisymmetric stretch	700	698
H-C-H rocking	955	953
S=O antisymmetric stretch of dimer	1042	1042
H-C-H deformation	1308	1307
H-C-H deformation	1418	1417
C-H symmetric stretching	2912	2913
C-H antisymmetric stretching	2995	2994

alkynes are typically much less intense and of shorter wavenumber than their diphenylacetylene counterparts and it is, therefore, likely that this alkyne is present as the weaker peaks found at 2234/2237/2238 cm^{-1} for DC473/DC271/EC23. The strongest bands of the spectrum were assigned to phenyl ring vibrations at 1597/1598/1600 cm^{-1}, at 1141/1145/1153 cm^{-1} and at 1181/1189/1179 cm^{-1} for DC473/DC271/EC23 respectively. Additionally, EC23 and DC271 showed a common strong band at 1127 cm^{-1} likely caused by the phenyl-carboxylic acid group vibration. Even though the samples were dried before measuring, some of the DMSO bands in the spectrum were present at ca. 305, 334, 388, 673, 702, 1042, 1412 and 2916 cm^{-1}.

Additionally, the spectrum of the provided DC473 (1 mM) in DMSO solution was acquired using the 785 nm laser with the 40× objective in a quartz slide-coverslip setup (15 accumulations, dose of 1.13 mJ, standard confocality) – see figure 5.19, confirming that the phenyl and the alkyne main peak positions in solution were approximately the same as in crystal form. The DMSO signal was overwhelming compared to the DC473 signal, but considering that 1 mM is equivalent to approximately 0.03% w/w, the fact that RS can measure the compound proves that it possesses a strong Raman signal.

Table 5.3: Bands positions (P), intensities (I) and assignments for DC473, DC271 and EC23 for the crystal spectra acquired with the 785 nm laser shown in figure 5.18. Intensity is relative to the maximum intensity of the spectrum, that was the Phe peak at around 1197 cm^{-1} for the three compounds.

DC473		DC271		EC23		Assignment
P / cm^{-1}	I / %	P / cm^{-1}	I / %	P / cm^{-1}	I / %	
305	10	305	10			DMSO
334	12	335	11	337	1	DMSO
		383	8	388	2	DMSO
411	12	424	7	407	2	Tolan/phenyl
453	11	461	10	461	3	Tolan/phenyl
491	11					Ester
		525	4	525	3	Phenyl/carboxylic acid
608	2	610	4	622	3	
674	18	673	33			DMSO
702	5			690	5	DMSO
		773	3	777	2	Carboxylic acid
829	2	828	3	826	2	
		892	2	891	1	Carboxylic acid
954	1	952	2	955	1	DMSO
		1018	6	1017	3	Carboxylic acid
1035	2	1085	1	1064	12	DMSO
				1104	6	
		1127	29	1127	33	Phenyl/carboxylic acid
1141	37	1145	33	1153	27	Phenyl
1181	11	1189	8	1179	15	Phenyl
1204	5	1223	1	1211	2	Phenyl/Tolan
1263	3	1278	6	1277	7	Phenyl/Tolan
1330	7	1310	5	1314	3	Phenyl/Tolan
1367	5	1374	2	1384	1	Phenyl/Tolan
				1404	1	
1412	9	1418	5			DMSO
1448	10					Ester
				1487	5	
1521	5	1511	1	1509	2	Phenyl/Tolan
1555	6	1550	2	1548	3	Phenyl/Tolan
1597	100	1598	100	1600	100	Phenyl/Tolan
1634	20					Ester
2197	27	2195	49	2203	56	Alkyne
2234	4	2237	2	2238	7	Alkyne
2874	4	2874	1			Tertiary amine
2916	5	2915	4			DMSO

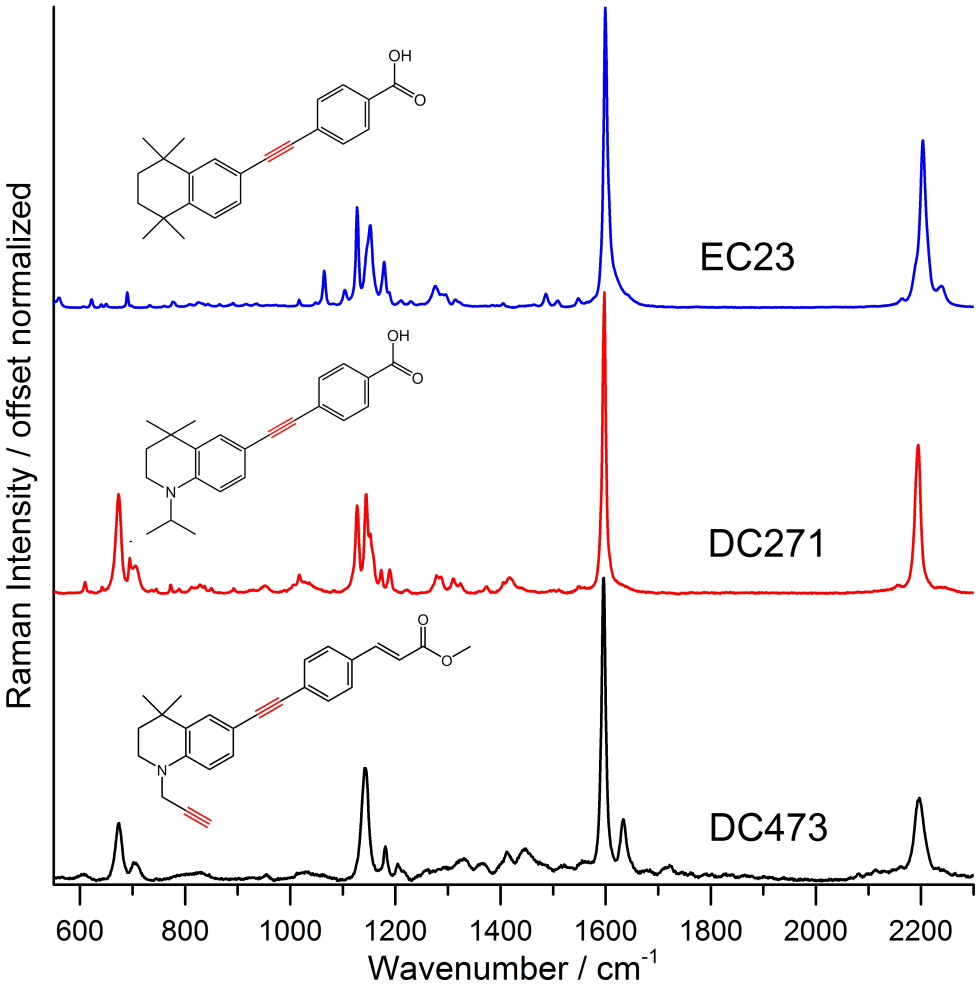


Figure 5.18: Normalised spectra of crystaline akyne compounds DC273, DC271 and EC23 when exciting with the 785 nm laser line. Compounds were provided by Lightox Ltd. Spectra has been offset for clarity.

The photophysical behaviour of Tolan has been widely studied, and the compound's conjugated structure has been extensively modified to tailor its properties towards a variety of applications. [281–283, 285–291] By appending strong π-donor (e.g. the tertiary amine NR_2) and π-acceptor (e.g. CO_2H, CO_2R) moieties, the absorption and emission characteristics of these compounds can be significantly bathochromically shifted due to the formation of intramolecular charge transfer excited states. [292, 293] Tolans are also known to exhibit triple bond stretching vibrations of particularly strong intensity in Raman spectroscopy in comparison to other alkynes. [125]

For the present work, the photodynamic activity of DC473 was investigated for its potential as a cancer treatment and the strong Raman signal and fluorescence allowed tandem imaging with Raman and confocal fluorescence microscopy. DC271, however, showed promising applications given its retinoid activity and orthogonal properties as Raman and fluorescent active molecule and will be studied in more detail in the future.

Figure 5.20 shows the normalised UV-Vis and fluorescence spectrum of DC473 in different solvents and table 5.4 gives the main spectroscopic parameters for this figure (provided by our collaborators in Julius-Maximilians-Universität). The highly lipophilic, electron rich

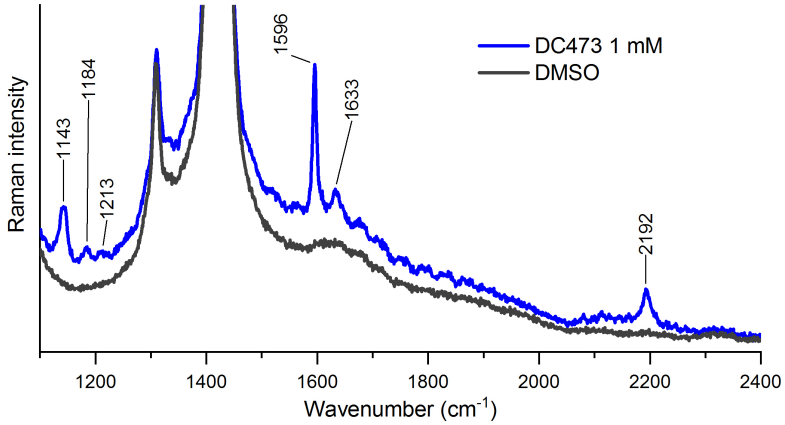


Figure 5.19: Spectra of DC473 (1 mM) in DMSO when exciting with the 785 nm laser line, where main DC473 bands have been labelled, and the spectra of the pure DMSO is shown as a reference.

tetrahydroquinoline donor and a methyl cinnamate acceptor on DC473 made the compound exhibit a highly solvatochromatic fluorescence emission with quantum yields up to 1.0 in non-polar solvents like Toluene, that dropped to 0.01 in more polar solvents such as Ethanol. In non-polar solvents, DC473 exhibits strong emission in the blue/green region (400-500 nm for methylcyclohexane, 450-650 nm for Toluene), while in more polar solvents the emission is significantly bathochromically shifted due to stabilisation of the intramolecular charge transfer excited state (450-750 nm for Ethanol, 500-700 nm for dichloromethane). DC473 exhibits two strong absorption bands in the UV and violet regions.

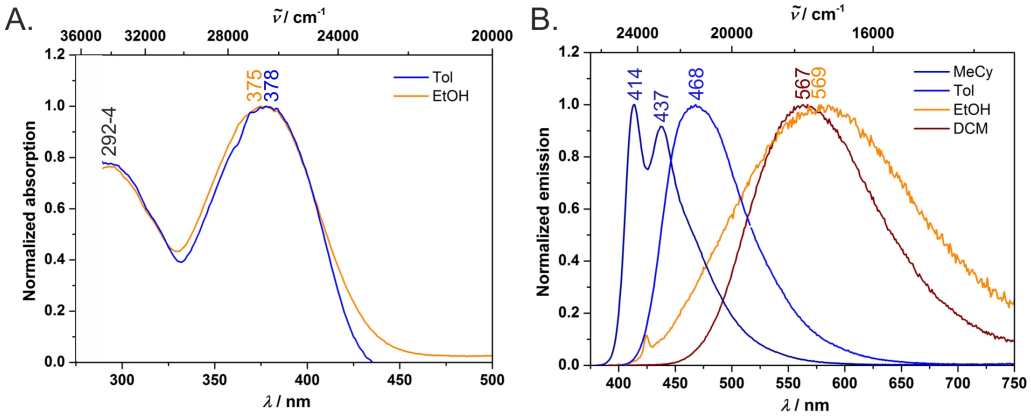


Figure 5.20: A. UV-Vis absorption and **B.** emission Fluorescence spectra of DC473 in different solvents, provided by our collaborators in Julius-Maximilians-Universität (Tol=Toluene, EtOH=Ethanol, MeCy=Methylcyclohexane and DCM=Dichloromethane).

Table 5.4: Spectroscopy values for DC473 in different solvents, provided by our collaborators in Julius-Maximilians-Universität (λ=wavelength, ϕ=quantum yield, τ=fluorescence lifetime).

Solvent	λ_{abs}(max)/nm	λ_{em}(max)/nm	ϕ	τ (ns)
Methylcyclohexane	-	414, 437	-	-
Toluene	378, 292	468	1.00	1.6
Ethanol	375, 294	569	0.01	-
Dichloromethane	-	567	-	-

Accordingly, in a cellular context, previous experiments with fluorescence microscopy indicated that DC473 is clearly present in non-polar, membrane-rich environments. However, while the compound was not visible in the cytosol by fluorescence, it would be unwise to infer that the compound was completely absent from this environment.[160]⋆ For this, both Raman

and fluorescence were undertaken in the same cells, and both the phenyl vibrations at 1597 cm^{-1} and the 2197 cm^{-1} alkyne peak were used for the detection of the compound via RS in single cells.

5.2.2 Confocal Fluorescence and Single Point Raman Detection of DC473 within Fixed Single-Cells

For an initial determination of the compound location within the cells, the primary adenocarcinoma CRC cell line, SW480, was used. Raman spectra were acquired with the 785 nm laser at individual points in fixed SW480 cells that had been incubated with DC473 and compared with the confocal fluorescence images of the same cells. Figs. 5.21A and 5.21B show the bright field and confocal fluorescence (ex. 405 nm) images of an SW480 cell following 4 h incubation with DC473. Figure 5.21C shows raw Raman spectra taken from different parts of a fixed cell. DC473 accumulation was detected in vesicles and in membranes, but not in the nucleus. The fluorescence image in figure 5.21B showed accumulation of the compound mainly in vesicles and dimly in the cytosol, but not in the nucleus. Raman measurements at the indicated points (V1-3) confirmed the presence of DC473 in particles which correlate with the strong 1455 cm^{-1} band associated with lipids, indicating that they were likely lipid droplets. [97, 294] Cytosol measurements showed weak signals of DC473, in agreement with the fluorescence characterisation. Single point measurements in the nucleus showed only a very weak signal from the compound.

For single-point Raman acquisition with the 532 nm laser, Figs. 5.21D and 5.21E show the bright field and confocal fluorescence images of an SW480 cell following incubation with DC473. Figure 5.21F shows individual points probed with RS with the 532 nm laser in the same cell in vesicles (V), the nucleus (N), cytosol (C) and filopodia (F). Similar to the results obtained with the near-infrared excitation, DC473 was detected strongly in vesicles and weakly in the cytosol. The examined cytosol regions exhibited the characteristic 749, 1128, 1310 and 1582 cm^{-1} cyt C peaks, indicating that the DC473 in this compartment could be localized near, or inside, the mitochondria. Mitochondria are an important target for PDT agents and it is, therefore, important to understand the behaviour of DC473 with respect to co-localisation with the mitochondria signal. Ideally, higher resolution studies should be undertaken to confirm this co-localisation. However, given the large membranous area associated with the mitochondria and the ability of DC473 to be accumulated in lipid-rich environments, this seems reasonable. Indeed, given that the DC473 signal is not resonantly enhanced whilst that of the cyt C is, this would imply significant accumulations of the drug in the mitochondria.

5.2.3 Raman Mapping and Confocal Fluorescence Detection of DC473 within Fixed Single-Cells

Raman mapping was undertaken on fixed SW480 cells incubated with the compound (4 h), or with an equivalent concentration of DMSO, as a control (figure 5.22) using 785 nm excitation. This excitation was chosen due to a small but non-zero overlap of the fluorescence

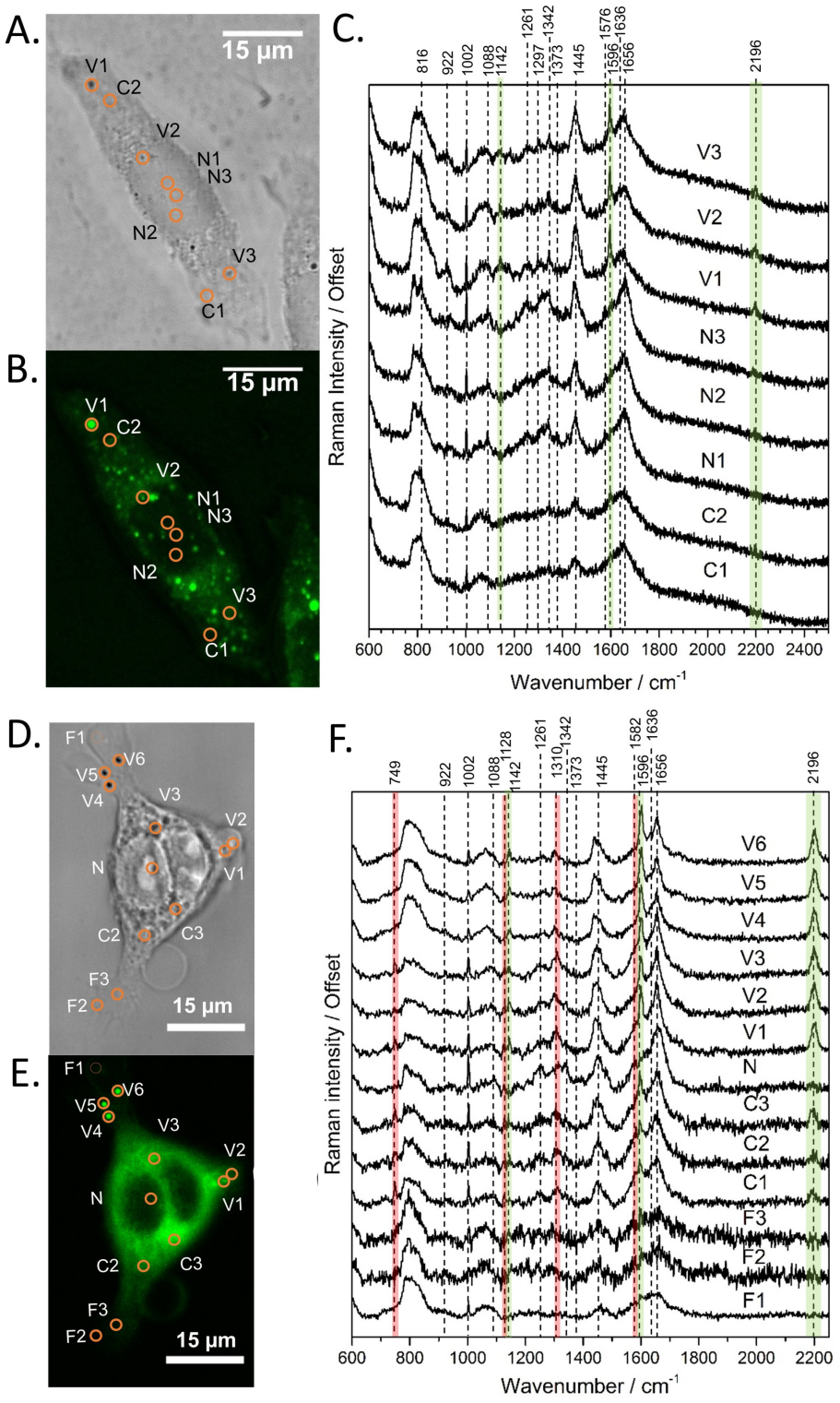


Figure 5.21: **A.** and **D.** show the Bright field and **B.** and **E.** show the confocal fluorescence images of fixed SW480 cells following incubation with DC473 whose point spectra are shown in **C.** and **F.** respectively. (V=vesicle, N=nucleus, F=filopodia, C=cytosol). **C.** Raw Raman 785 nm spectra at selected points of the SW480 cell showed in **A.** and **B.** The green bands show the main peaks associated with the presence of the compound in different cellular locations, with strong presence in vesicles. **F.** 532 nm Raman baseline subtracted spectrum of selected points of the SW480 cell showed in **D.** and **E.** The green bands show the main DC473 peaks, showing similar band positions to figure 5.18. Measurements in the cytosol showed weak signal of DC473 and RR bands of cyt C[1], indicated with red bands.

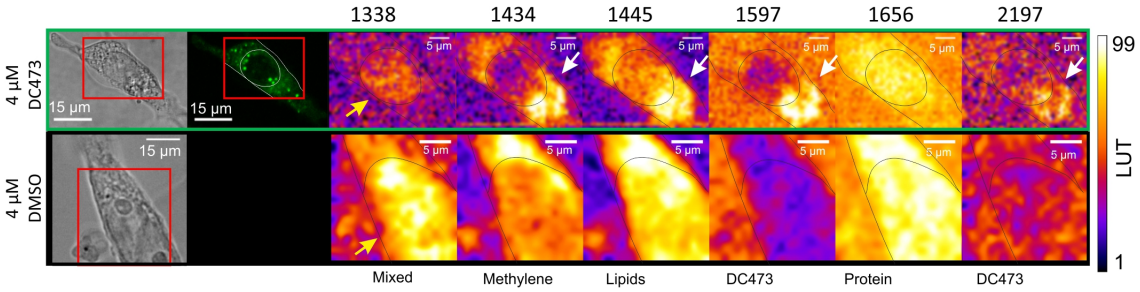


Figure 5.22: Peak integrated intensity maps for SW480 fixed cells either incubated with DC473 or with an equivalent concentration of DMSO as a control. The cell and nuclear areas have been outlined as visual guidance. The 1597 and 2197 cm^{-1} peaks showed the DC473 accumulation mainly in lipid droplets (white arrows), as indicated by the correlation with the 1445 cm^{-1} intensity map mainly due to lipids vibrations. The 1338 cm^{-1} intensity map showed mixed contributions of proteins and nucleic acids (nucleus labelled with a yellow arrow), and the 1656 cm^{-1} map showed Amide I intensity as a map of proteins. The 1455 and 1434 cm^{-1} maps mainly showed contributions of lipids.

excitation spectrum of the molecule with the 532 nm excitation line, that not only gave a small baseline on the spectra but also seemed to bleach the compound by the end of the mapping acquisition. Bright field and fluorescence images for the DC473-treated cell are shown together with the Raman maps associated with the main spectral peaks. The 1338 cm^{-1} band, usually assigned as a mixed contribution of CH_2 wagging, $C-O$ stretching, phenyl stretching and $N-H$ deformation, [32, 295, 296] here appeared most strongly in the nucleus. The 1434 cm^{-1} band associated with methylene stretching [297, 298] showed a strong contribution in lipid droplets rich regions and exhibits a strong correlation with the 1455 cm^{-1} band associated with lipids and fatty acids [297] as well as the 1597 cm^{-1} phenyl peak associated with DC473. The 1656 cm^{-1} Amide I band [297] due to proteins showed intensity over the whole cell area.

The two main DC473 peaks, 1597 cm^{-1} (phenyl) and 2197 cm^{-1} (alkyne), showed a strong correlation with the lipid peaks and a weaker contribution in the cytosol, indicating localisation of the drug in lipid-rich environments. Whilst the 1597 cm^{-1} DC473 peak was very strong and provided good contrast in the Raman images, its presence on the side of the Amide I peak made it difficult to separate from that of protein. In contrast, the 2197 cm^{-1} band, whilst weaker, occurs in the 'silent' region of the spectrum giving unequivocal detection of the compound. Interestingly, a high concentration of the compound was found in a group of perinuclear vesicles while other vesicles did not show detectable drug uptake using RS, but showed strong fluorescence. The population on the top left side of the nucleus does show a signal at 1445 cm^{-1}, but weaker, and no detection of DC473, while the population in the low right side (white arrows) showed DC473 presence. It could be that due to the extremely high quantum yield of the compound when in an apolar environment, fluorescence saturates in intensity, whereas Raman differentiates these two populations of vesicles as containing different amounts of DC473 – detectable and undetectable. A possible explanation for this is either changes in the polarity of the environment of the compound or aggregation.

PCA analysis of one of the maps of a cell incubated with DC473 is shown in figure 5.23A, along with the PCA[1] loadings of these components in figure 5.23B. PC1 showed cell signal

[1] Principal Component Analysis

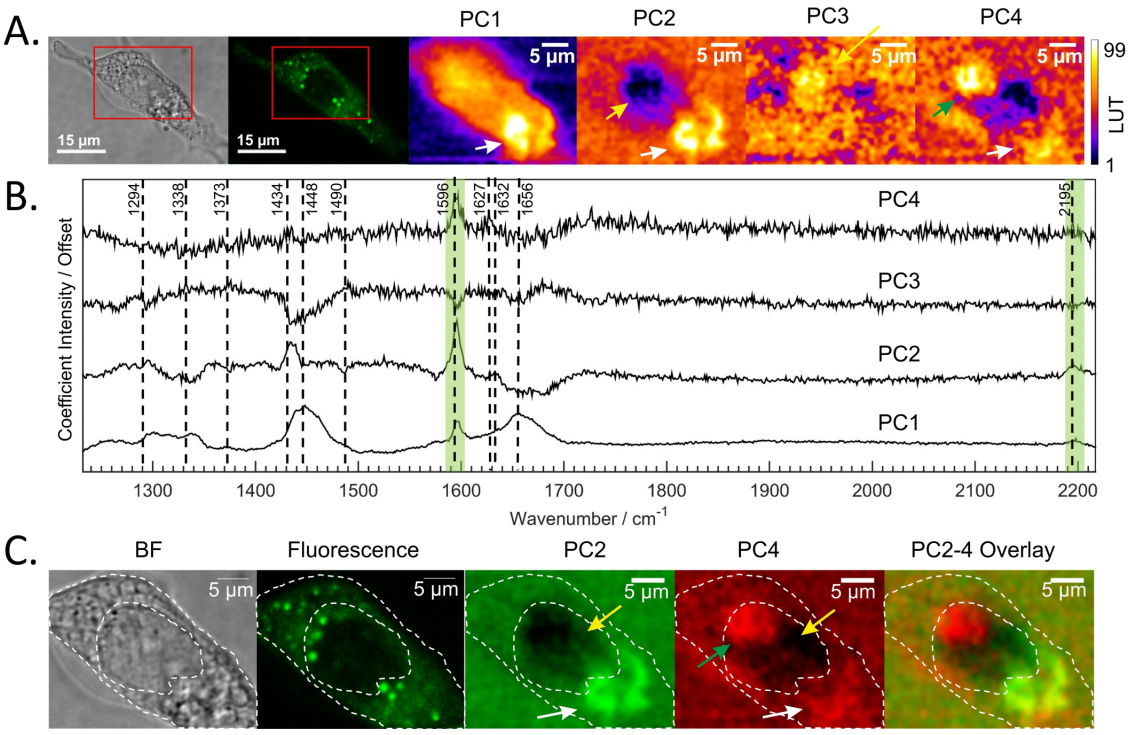


Figure 5.23: **A.** Bright field and confocal fluorescence images of the fixed SW480 cell following incubation with DC473. The maps of each of the scores of the 4 main PCs are shown, with white arrows indicating the DC473 loaded lipid droplets, yellow arrows indicating the nuclear area and a green arrow for the DC473 accumulation in the region overlapping with the nucleus. **B.** Coefficients for the first 4 PCs whose scores are mapped in A. **C.** Bright field, Fluorescence and PCA coefficients with DC473 presence. The cell main features have been overlaid with the images as a visual aid. The intensity of PC4 suggests that Raman could be picking up the presence of DC473 in a region in the nucleus that does not show in the fluorescence image (green arrow) with additional accumulation in the lipid droplets area (white arrows).

with drug contribution, where the whole cell showed higher scores than the background and thus gave similar information to the fluorescence image. PC3 intensity was highest in the nuclear area (indicated with a yellow arrow), with negative contributions for both drug and the broad lipid band around 1448 cm^{-1}. PC2 and PC4 showed the main drug-specific components. PC2 showed very strong DC473 presence and positive 1434 cm^{-1} methylene band indicating accumulation in the elongated structures in the region below the nucleus (white arrow) in figure 5.23A/C. PC4 shared some similarities with PC2, but with lower DC473 bands intensity and with a stronger contribution of a band at 1627 cm^{-1}, and absence of the 2195 cm^{-1} – either undetectable due to noise, or detecting a metabolite of the compound without the alkyne. The scores for this component did not correlate with the fluorescence signal, as shown in figure 5.23. The PC4 component indicated the presence of DC473 in a compact area near or inside the nucleus (green arrow), and dim contributions in areas in the lower right region of the cell (white arrow), suggesting that a low concentration of the compound could be present inside the nucleolus. The absence of fluorescence in this region might be explained by a solvatochromatic reduction in quantum yield when the compound is localised in a more polar environment.

In conclusion, the location of DC473 is mainly in organelles, potentially lipid droplets, and in the cytosol, potentially in mitochondria, correlating with the fluorescence information. However, not all lipid droplets accumulated compound. The compound also showed a weak signal in a compact region inside the nucleus, that could be the nucleolus. More research is needed to understand its exact location within the cell and its effects in each of these locations, where the location in the mitochondrial membrane and inside the nucleus would increase its damaging effects as a photosensitiser when exposed to UV light and are thus desired.

5.3 Detection of DC473 in Live Single-Cells

After the characterization of the compound within fixed single cells using Raman and fluorescence, the next step was to detect the compound using RS[1] in live single cells from the adenocarcinoma progression model system SW480/HT29/SW620. Cells were prepared in a similar fashion to previous experiments, with a 4h incubation with 4 μM concentration of the photosensitiser DC473 dissolved in DMSO, whereas the control cells got added an equivalent volume of DMSO. Around 135 cells were measured for each cell line in two independent experiments for different passage numbers, using the defocused laser beam to obtain the average cell spectra for every single cell. Results showed good reproducibility between experiments. The 532 nm laser was used for this experiment to ensure higher quality Raman signal.

Figure 5.24 shows the average spectra for each of the cell lines comparing cells following incubation with DC473 or with DMSO[2] (control), normalised to the Amide I band. The five strongest peaks from the spectrum of pure DC473 and from the DMSO from figures 5.18 and figure 5.17 (page 5.18 and 130 respectively) are indicated: 674 cm^{-1} (DMSO),

[1] Raman spectroscopy
[2] Dimethyl sulfoxide

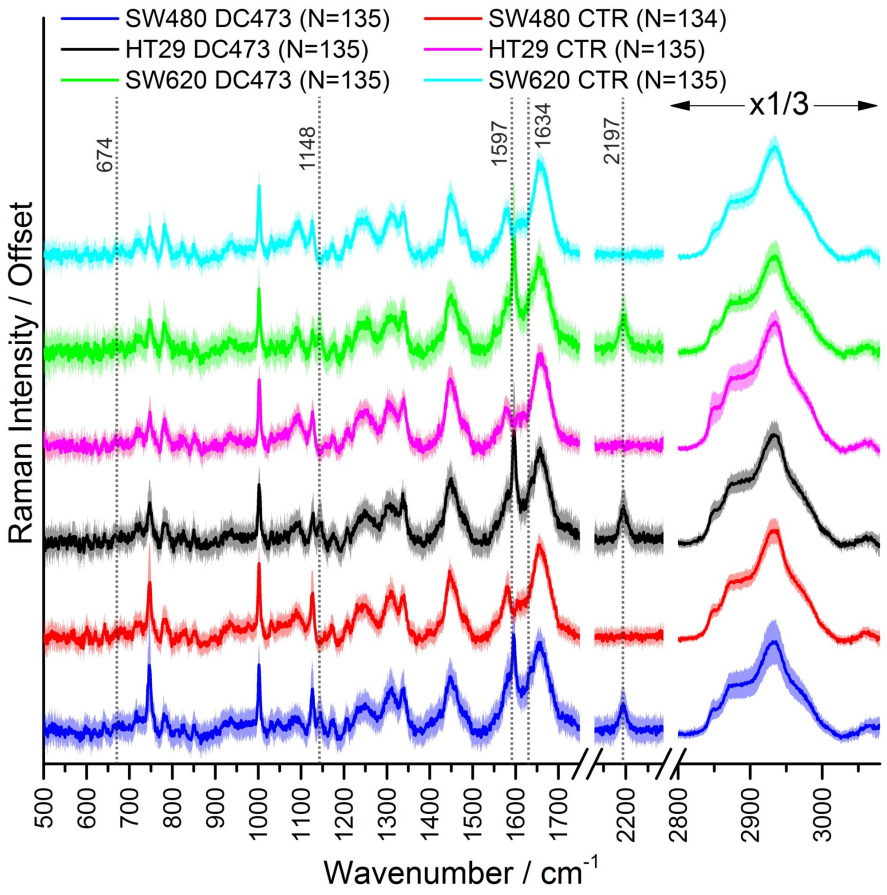


Figure 5.24: Average spectra of ≈135 live cells per cell line either treated with 4 μM DC473 or with an equivalent volume of DMSO (Control) for the three cell lines in the adenocarcinoma model system. The shaded area around the curve shows one standard deviation. The five stronger peaks for the spectrum of pure DC473 are indicated with a dashed line and described in the text.

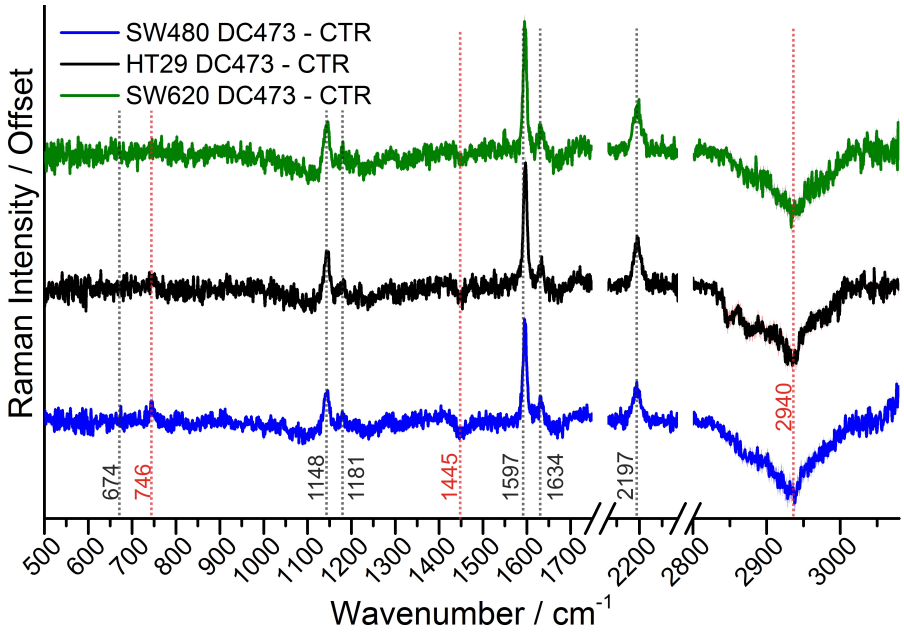


Figure 5.25: Difference spectra (DC473 - Control) of the figure 5.24 average spectra either treated with 4 μM DC473 or with an equivalent volume of DMSO (Control) for the three cell lines in the adenocarcinoma model system. Labels in black indicate DC473 bands, whereas labels in red indicate other cell associated bands.

1141 cm^{-1} (37% intensity, phenyl), 1597 cm^{-1} (100% intensity, phenyl), 1634 cm^{-1} (20% intensity, Ester) and 2197 cm^{-1} (27% intensity, alkyne). Averages show unequivocal uptake of the drug for all cell lines, with higher DC473 content (per Amide I) for HT29 and SW620 cells. The 1141, 1597 and 2197 cm^{-1} bands are present in all cell lines, and the 1634 cm^{-1} peak shows as a shoulder on the Amide I band. DMSO is non-detectable on the sample, with the 674 cm^{-1} peak not present in the cell spectra. This is confirmed in the correlation series (see figure 5.26), that neither showed the 674 cm^{-1} band.

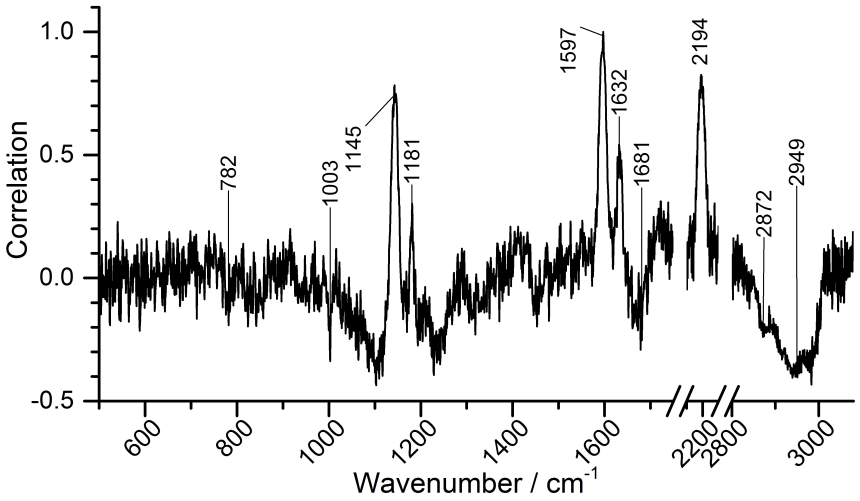


Figure 5.26: DC473 1597 cm^{-1} correlation series for the single-cell spectroscopy data. Data showed the main photosensitiser peaks after uptake in a single-cell, with a negative correlation with CH stretching bands and the phenylalaline peak at 1003 cm^{-1}. Main positive and negative peaks have been labelled.

This can be better appreciated in the difference spectra shown in figure 5.25. The difference spectra not only shows the main DC473 peaks, but also shows a reduction in the CH contribution both at 1445 cm^{-1} and at the stretching band at 2940 cm^{-1} following incubation with DC473, indicating that cell accumulation of DC473 in lipid droplets could interfere with lipid metabolism and/or regulation. The 1181 cm^{-1} peak (phenyl, 10% intensity in the standard spectrum) can be appreciated in the difference spectra, and the 674 cm^{-1} DMSO band is generally absent. Additionally, the difference spectra showed an increase in the 746 cm^{-1} cyt C[1] band and a strong bend at 1585 cm^{-1} after incubation with DC473, strengthening the hypothesis that not only DC473 localized in mitochondria but it may also interfere with it even before PS[2] activation.

5.3.1 Principal Component Analysis of the DC473 Accumulation Data

To further understand the data, PCA[3] was done. PCA on the not-smoothed data gave a variance explained of ≈ 11, 6, 3, 2 and 1 % for the first 5 PCs (total 25% explained). Figure 5.27A shows the first 6 loadings for PCA analysis, where the main DC473 (black) and cell (red) peaks have been labelled. Figure 5.27B shows the scores for those first 6 PCs. PC1 showed mainly drug related peaks, with the exception of a small contribution of cyt C at 746

[1] Cytochrome C
[2] Photosensitiser
[3] Principal Component Analysis

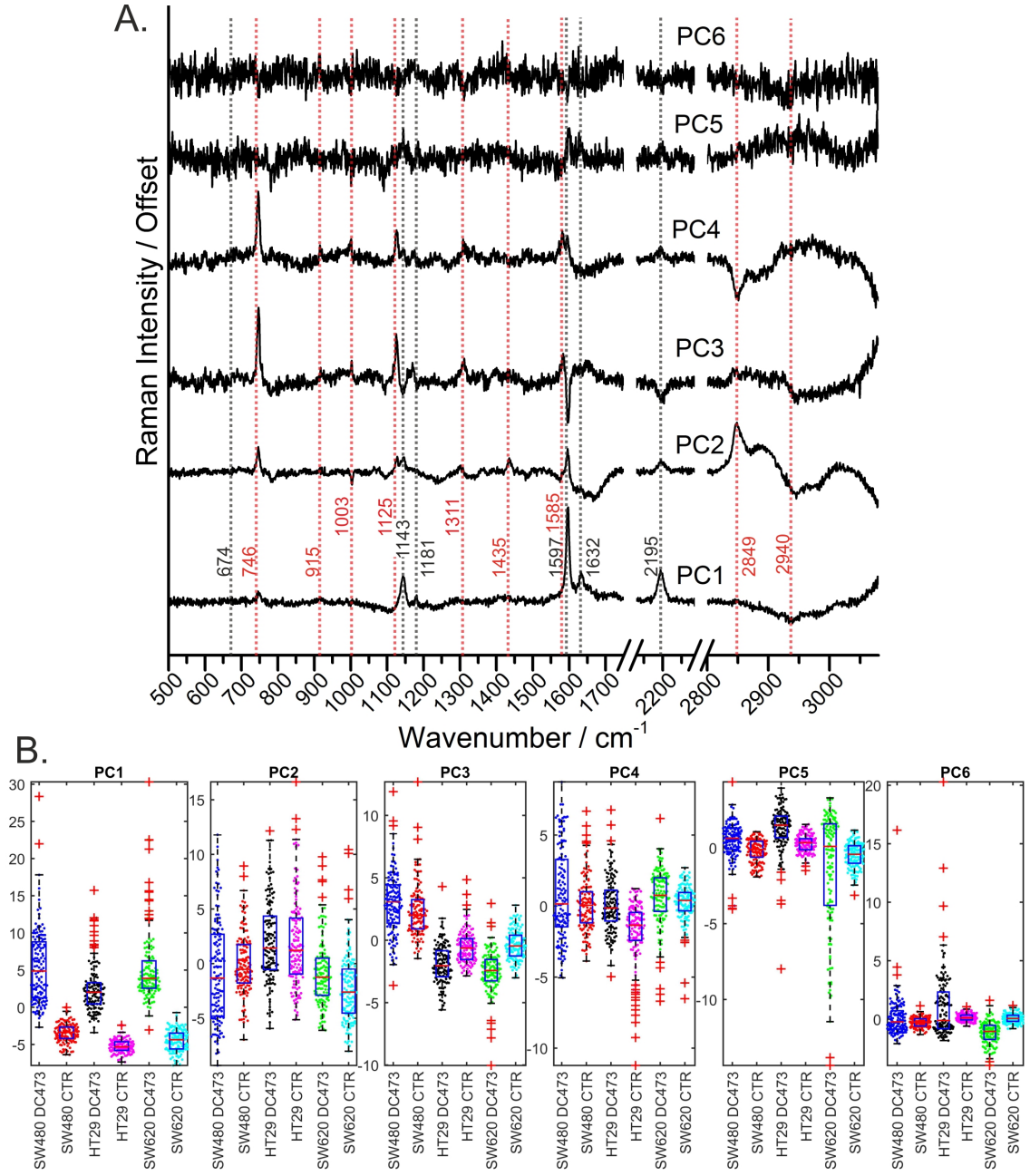


Figure 5.27: **A.** First 6 loadings for PCA analysis of cells following incubation with DC473, where the main DC473 peaks (black) and cell peaks (red) have been indicated. **B.** Boxplots and beeswarm plots for the scores of the first 6 PCs from the PCA analysis of the Raman spectra for the adenocarcinoma model system cells following incubation with DC473 or with DMSO (Control). PC1 was the main axis of drug accumulation, PC2 showed mainly within sample heterogeneity, PC3 showed mainly peaks higher in the control group, PC4 showed decreased lipids contribution in incubated cells and PC5 and 6 accounted for heterogeneous subpopulations in SW620 and HT29 cells incubated with DC473 respectively.

cm^{-1} and negative contribution in the CH_3 stretching band around 2940 cm^{-1}, confirming the trend of the *difference spectra* that cells with higher concentrations of DC473 displayed higher cyt C contributions (746 cm^{-1}) and lower lipid contribution. PC2 showed a similar shape to the one previously observed in the PCA analysis of the different cell lines from chapter 4 (page 91, shape of PC1) with significant contributions of cyt C peaks, showing single-cell within population heterogeneity in figure 5.27B. PC3 showed mainly cyt C bands and negative DC473 bands. PC4 showed a drop on the 2849 cm^{-1} CH_2 stretching band with positive cyt C bands, showing high contributions for DC473 incubated cells than control cells (figure 5.27B), meaning an overall CH stretching drop occurred following incubation with DC473.

PC5 and 6 showed considerable noise and seemed to pick out subpopulations in SW620 and in HT29 respectively (see figure 5.27B). PC5 showed lower intensity contribution for the 1597 cm^{-1} band compared to the 1143, 1632 and 2195 cm^{-1} than what was observed in the spectrum, maybe indicating that part of the drug could be in a different environment or aggregating in these cells. It also showed negative contributions at 781 cm^{-1} (generally DNA band) and at 1091 cm^{-1} (phosphates), showing that this particular subpopulation could be in a different cell cycle stage with lower DNA. PC6 however, is largely noise.

5.3.2 Fitting of the Main DC473 Bands

To compare the DC473 accumulation between different cell lines, we used GLS[1] single cell peak fitting of the spectral peaks at 2195 cm^{-1} and at 1597 cm^{-1}. Figure 5.28 shows an example of the single cell GLS fittings on both peaks. For the 2195 cm^{-1} band, the fitting was straight forward as the DC473 peak was present alone. The peak was allowed a variation in the position of $\pm$ 10 cm^{-1} and a width of 20 $\pm$ 10 cm^{-1} to also be able to track any peak shift occurring in single cells.

For the 1597 cm^{-1} band, the fitting was more challenging due to overlap from the adjacent amide and cyt C bands. Three bands were fitted for the Amide I; namely: α-helix (1648 cm^{-1}), β-sheet (1667 cm^{-1}) and random coils (1680 cm^{-1}), all of them with a width of 40 cm^{-1} and with a width variation of $\pm$ 20 cm^{-1}. Additionally, the 1585 cm^{-1} cyt C peak was also fitted (width 40 cm^{-1}). All peaks were allowed a variation in the peak position of $\pm$ 5 cm^{-1} except the DC473 fitted peak at 1597 cm^{-1}, that was allowed no variation in position to avoid fitting artefacts in noisy spectra. The results from the fitting amplitude are shown in figure 5.29, where the significance between cell types is also indicated (based on t-test with unequal variances). For all cell lines, uptake of DC473 was highly significant compared to control cells. When comparing different cell lines (normalising to the Amide I band), SW620 and HT29 had similar uptake levels and significantly higher accumulation of DC473 than SW480 cells. This suggests that more advanced tumour cells tend to accumulate higher amounts of this photosensitiser.

For the 2195 cm^{-1} peak, as the position and width of the peak were given a range to get the best possible fit, the values for the final position and width of the peak for each

[1] Gaussian-Lorentzian sum

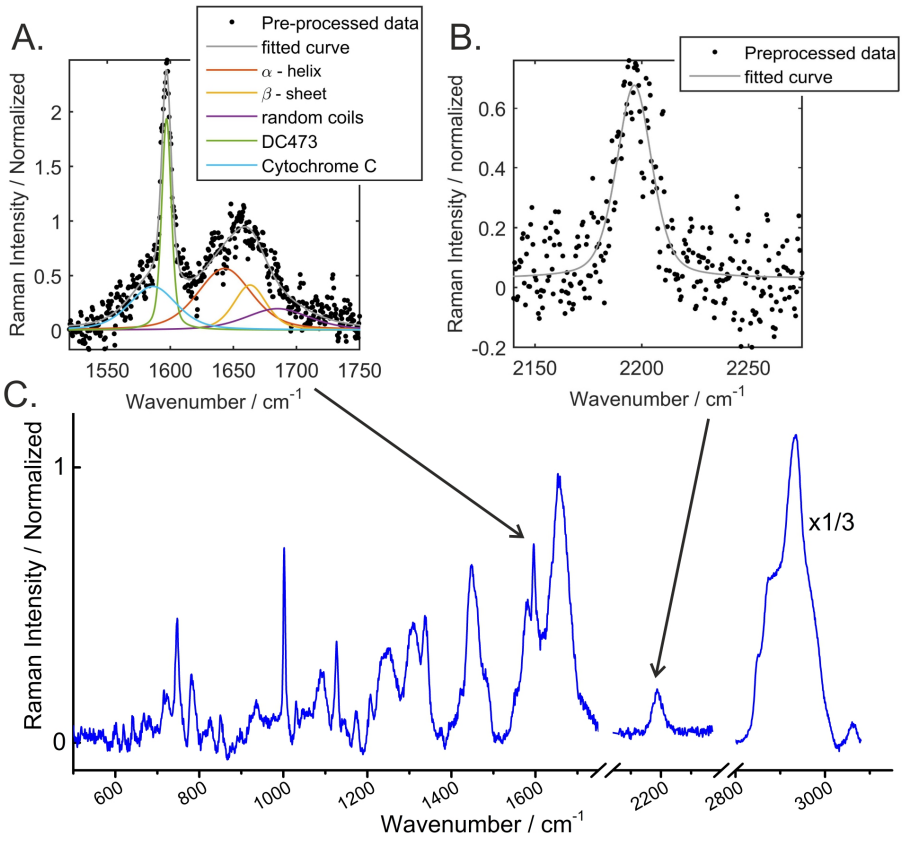


Figure 5.28: Examples of the GSL fittings done for the peaks at **A.** 1597 and at **B.** 2195 cm^{-1} on single cell data for the adenocarcinoma cell lines. **C.** The mean of all data where the fitted peaks have been indicated.

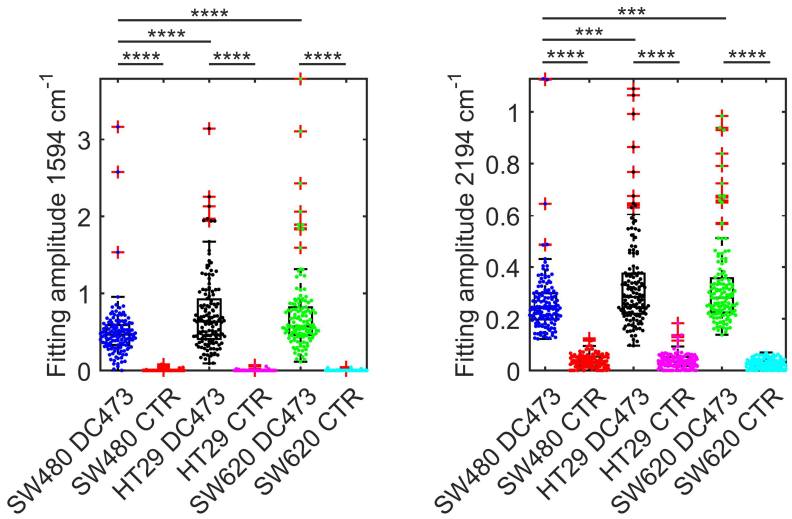


Figure 5.29: Boxplots and beeswarm plots for the peak fitting amplitudes for the peaks at 1597 (*left*) and at 2195 cm^{-1} (*right*). Significance level from a normal distribution t-test are indicated (*:p< 0.05,**:p<0.01,***:p< 10^{-3},****:p< 10^{-4})

cell line could be obtained. When filtering by the amplitude of the peak by removing any fittings with amplitude lower than 0.15, the mean amplitude, width and position obtained (± standard error) are shown in table 5.5. Cell lines with higher accumulation of DC473 showed a small significant shift towards higher wavenumbers and wider peaks. The reason for this is currently unknown, although it could be showing some sort of aggregation of the compound at higher concentrations shifting the peak to higher wavenumbers, causing two populations of the compound to be present, thus broadening the peak.

Table 5.5: Summary of fitting parameters for the alkyne peak at 2197 cm^{-1} for each of the cell lines. The parameters correspond to the GLS curves ± standard error. The number of cells that had an amplitude > 0.15 is shown (N), and were used to calculate the remaining parameters.

	N	Amplitude / norm. to Amide I	Position / cm^{-1}	Width / cm^{-1}
SW480 DC473	121	0.28 ± 0.01	2193.9 ± 0.2	18.8 ± 0.2
HT29 DC473	129	0.34 ± 0.02	2195.2 ± 0.2	19.7 ± 0.4
SW620 DC473	133	0.33 ± 0.01	2194.8 ± 0.2	21.0 ± 0.5

Additionally, to confirm the observations from the comparison of the averages, the single-cell data was integrated between 2800 and 3080 cm^{-1} to observe the single cell behaviour of the treated cells – shown in figure 5.30. Data showed a significant drop in CH-stretching bands intensity for all cell lines, more pronounced in the cell lines that showed more accumulation of the compound.

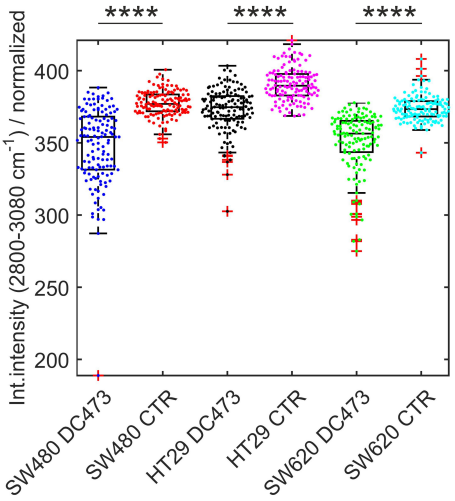


Figure 5.30: Boxplots and beeswarm plots for the integrated intensity of the single-cell data between 2800 and 3080 cm^{-1} (Data normalised to the Amide I band) following incubation with DC473 or DMSO. Significance level from a normal distribution t-test are indicated (*:p< 0.05,**:p<0.01,***:p< 10^{-3},****:p< 10^{-4})

However, current experiments were only in the accumulation of DC473 but said nothing about its activity as a photosensitiser. From the mapping results, it appeared that the accumulation happened mainly in lipid droplets, showing a strong correlation with lipid bands. Additionally, cells incubated with DC473 showed significantly lower intensity of lipid bands (normalised to the Amide I), indicating that DC473 accumulation could be affecting lipid metabolism. However, we would expect that the photosensitiser that accumulates in the

cytosol and mitochondria would generate ROS and cause more damage than the photosensitiser accumulated in lipid droplets, that could be aggregating and thus be in a less active form.

5.4 Tracking Apoptosis after UV Light Activation of DC473

To better understand the photodynamic effect of DC473, its activation with UV[1] light was studied both in solution and in live cells.

5.4.1 Activation of DC473 in solution

Previous results of our collaborators in LightOx showed that DC473 could not be bleached in a cuvette using oxygen saturated solutions and at various concentrations when using a 405 nm excitation light. [160] However, we found that deep UV light ($\sim 360nm$) was effective in bleaching of the fluorescence and loss of the Raman signal of the compound. A 100μM solution of DC473 in DMSO was measured before and after exposure to UV light (340-380 nm, ≈131 mW/cm^2, 40 min, total dose of ≈5.2 J/cm^2), using the 785 nm laser to avoid sample fluorescence (15 accumulations, 1.13 mJ dose, standard confocality). The acquired spectra are shown in figure 5.31, after baseline subtraction and normalisation to the maximum DMSO band. The DC473 contributions to the spectrum were zoomed in – remaining contributions were strong bands from the DMSO[2] solvent. The drug could be detected at this low concentration (0.003%), and UV light caused a significant drop in the main DC473 bands, without the appearance of other secondary bands. This could be indicating that the Tolan structure of DC473 is broken after UV light exposure, and thus the remaining metabolites are too low concentration to detect without the DC473 enhancement. It must be noted that even after such a long UV light exposure, not all the DC473 signal disappeared.

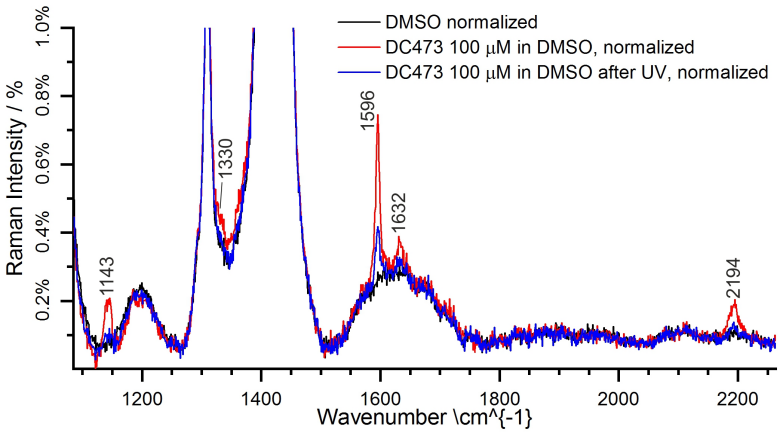


Figure 5.31: Raman spectra of DC473 (100 μM) in DMSO before and after exposure to UV light (340-360 nm, 5.2 J/cm^2). The DMSO spectrum has been included as a reference, and the main DC473 observed peaks have been labelled. Intensity is expressed as a % of the maximum DMSO band to which the spectrum was normalised.

[1] Ultraviolet
[2] Dimethyl sulfoxide

5.4.2 Tracking Apoptosis after DC473 Activation

To study the effects of the active photosensitiser in single-cells, a live time-dependent experiment was designed. Cells were incubated with the compound (4 μM) for 4 h as in previous experiments and detached for Raman measurements in cell dissociation buffer. A Raman spectrum in the fingerprint region was acquired before UV treatments (5s, standard confocality, 20% defocus). Subsequently, cells were treated with 3 seconds of UV light (340-380 nm) with a total dose of ≈ 711 mJ/cm^2. Due to the experimental set up used, the minimal UV light intensity on the sample without changing objective was of ≈ 237 mW, and thus lower doses were difficult to achieve with the current manual exposure. The Raman microscope was then immediately programmed to acquire a spectrum 14 times every 15 s. The total time spent measuring per cell was of $\approx$ 4 min 51 s, allowing 10-11 repeats per hour. The total laser dose on each cell was of 1.65 J with the 532 nm laser, equivalent to 525 kJ/cm^2, almost 6 orders of magnitude higher than the UV light – however, UV light, especially deeper UV, is known to be more phototoxic to cells than green light. The experimental design could only obtain information on 10-11 cells per group, with a total of around 30-35 cells per cell line, however, had good time resolution. Two controls were prepared: cells without DC473 exposed to the UV dose and cells without DC473 only exposed to the Raman laser. It was already known that DC473 did not degrade under the 532 nm laser doses used in this experiment, and thus no control with DC473 and no UV light was undertaken. The total number of cells observed was 94, with a total number of spectra of 1410.

An example of the data obtained in an experiment is shown in figure 5.32. The DC473 bands at 1597 and at 2195 cm^{-1} showed a strong contribution in the initial spectrum (*before*, t=0) and dropped rapidly after PS[1] activation with UV. However, not all the drug was activated, with a small amount of Raman detectable drug remaining in the cell. These data suggests that the drug not only bleached after UV light exposure, as fluorescence experiments by our collaborators showed, but it actually underwent a chemical change, losing its characteristic Raman signal. Spectral changes associated with cells after exposure were subtle.

Interestingly, changes in the morphology of the treated cells following all spectra acquisition were already visible for some cells in particular cell lines, as can be seen in figure 5.33. Cells would undergo blebbing – a sign of apoptosis or necrosis – by the end of the acquisition (indicated with white arrows). For the DC473 PDT cells, bright field pictures before and after the acquisition were taken. For the controls, no morphology changes were observed (only two cells out of 61 showed any changes), and some example pictures were obtained to show this. From the cells measured, 7/11 SW480 cells (64%), 0/12 HT29 cells (0%) and 6/10 SW620 cells (60%) showed morphology changes due to PDT. SW480 cells, the cell line that accumulated less DC473, showed morphological changes due to PDT more often, whereas HT29, that showed a DC473 average intensity around 20% higher, had no morphological response whatsoever. HT29 cells and SW620 cells, with a very similar accumulation, had very different responses to PDT. A better understanding of the responses to PDT of these cell lines needed to be obtained from biochemical tests such as MTT.

[1] Photosensitiser

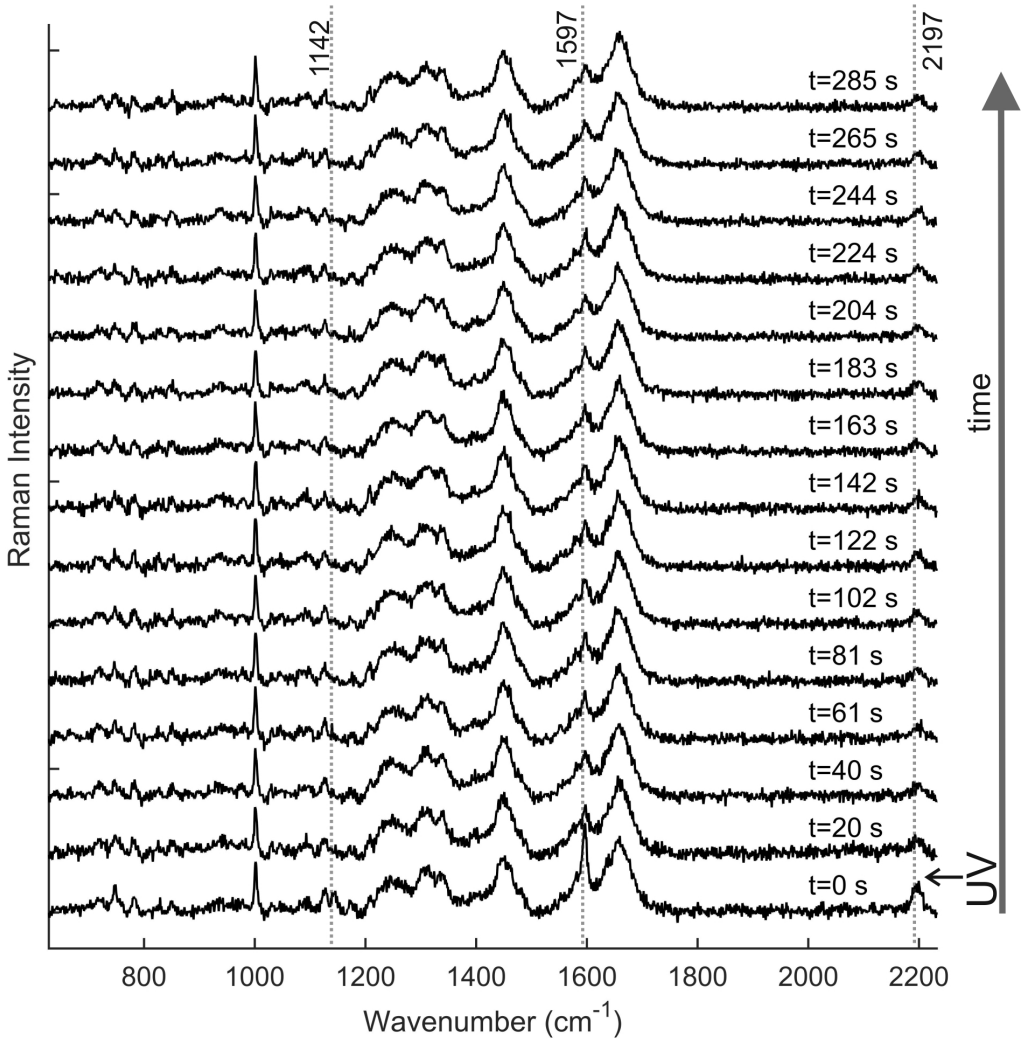


Figure 5.32: Example of a time-dependent single-cell response to the DC473 PDT[1] activation with UV light, where the first spectrum was taken before activation of the photosensitiser, and the remaining spectra were taken after activation. This example corresponds to a SW480 cell, and the data was pre-processed (cosmic ray removal, alignment to silicon peak, background subtraction and baseline subtraction).

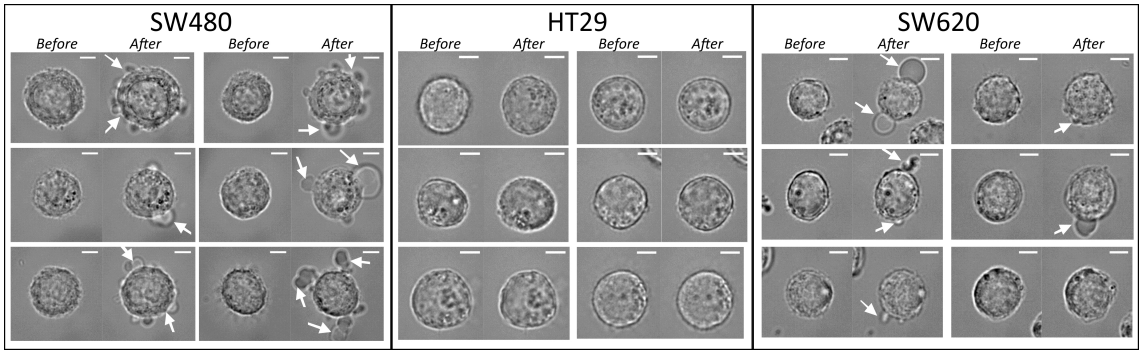


Figure 5.33: Examples of Bright Field images from following incubation with DC473 before and after PDT treatment with UV light and Raman acquisition. Some of the membrane blebs observed have been indicated with white arrows.

5.4.3 Analysis of the Average and Difference Spectra

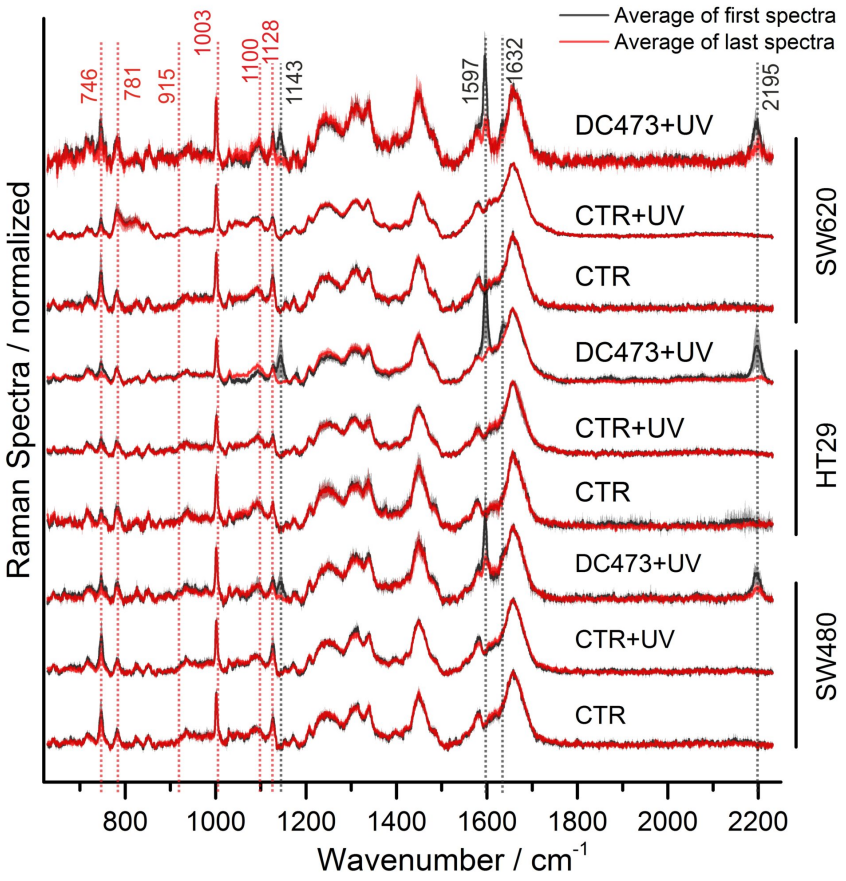


Figure 5.34: Average of the data before (black) and after (red) each PDT experiment for each of the samples and cell lines, where CRT=Control, CRT+UV=Control exposed to equivalent dose of UV light, DC473+UV=DC473 sample exposed to UV light. The main bands have been labelled. The area around the curve represents one standard deviation.

To illustrate the changes observed in the different datasets, the first and last spectrum for each cell in each experimental group (Control, Control + UV and DC473 + UV) in each cell line were averaged and compared. These spectra were overlaid for each cell line in figure 5.34. The difference in time between the *before* and *after* spectra corresponds to 285 s (4 min 45 s). The exposure to UV light – done in all samples but the control sample (CTR) – was done between the first two acquisition in a manual way, at t=8±5 s. When analysing the DC473+UV samples, the main DC473 bands at 1597 and 2195 cm^{-1} showed a strong drop between the *before* and *after* spectra, but not always down to undetectable levels (see bands at 1143, 1597, shoulder at 1632 and band at 2195 cm^{-1}). Particularly, SW480 and SW620 cells often showed detectable levels of photosensitiser after exposure to this dose of UV light, but HT29 cells showed a drastic drop of the compound's Raman signal. This results confirmed that the photosensitiser was undergoing a chemical reaction within these CRC[1] cells. It also suggested differential subcellular localization between SW480/SW620 cells and HT29 cells, that could explain the differences in response observed in morphology. The control samples remained mostly unchanged. However, when looking at cyt C[2] changes,

[1] Colorectal Cancer
[2] Cytochrome C

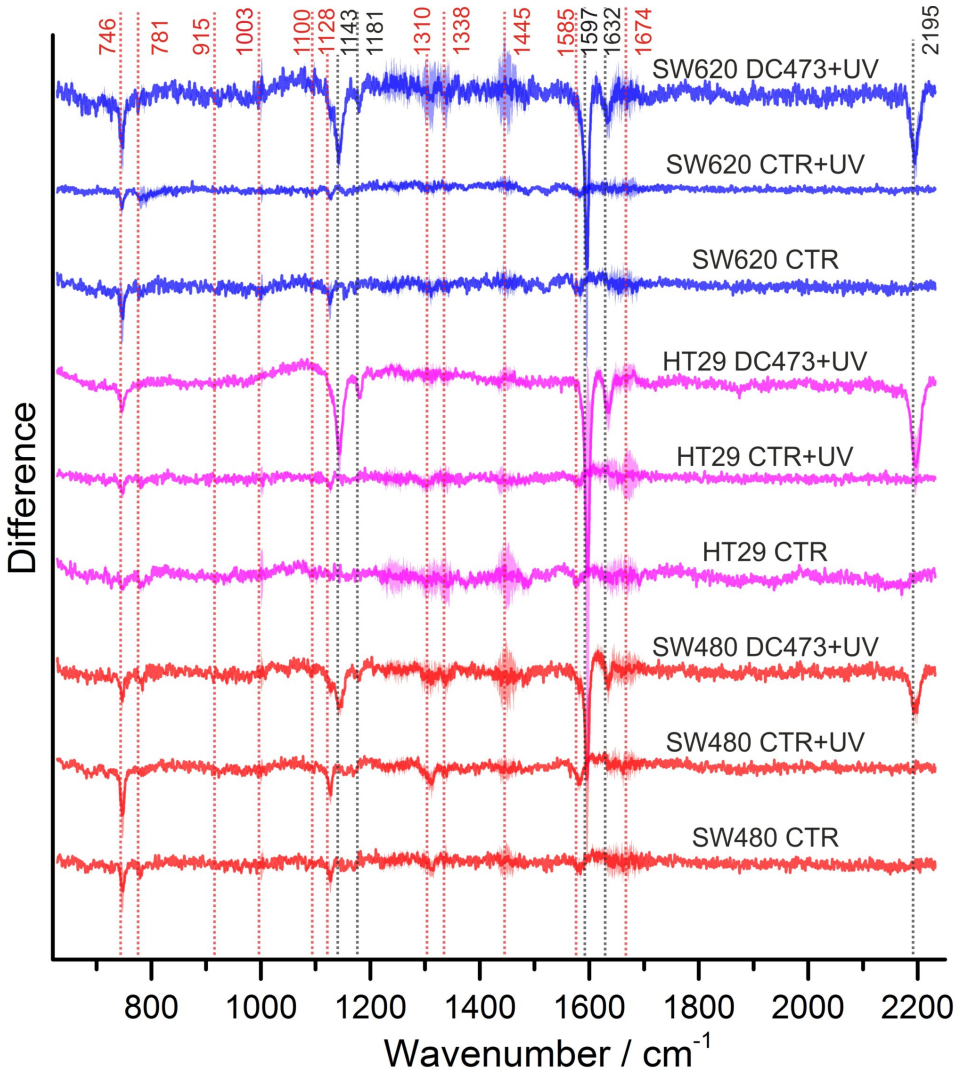


Figure 5.35: Average difference of the data (*last* - *first*) for the PDT experiments for each of the samples and cell lines, where CRT=Control, CRT+UV=Control exposed to equivalent dose of UV light, DC473+UV=DC473 sample exposed to UV light. The main bands have been labelled. The area around the curve represents the propagated error.

most samples showed a clear decrease in the band at 746 and 1128 cm^{-1}, indicating some phototoxicity not only from the PDT but also associated with the Raman acquisition laser and with the UV light dose.

When calculating the difference between the average spectra, the main effects of the PDT became evident, see figure 5.35. Results showed a drop in all DC473 peaks and in cell bands at 1443, 1181, 1597, 1632 and 2195 cm^{-1}. All datasets showed negative peak in the main cyt C bands (746, 1128, 1310 and 1585 cm^{-1}), indicating phototoxicity of the measuring laser light at the powers used (532 nm, 525 kJ/cm^2) and phototoxicity of the UV light used, with also DNA damage in some of the samples observed at 781 cm^{-1}. Additional changes observed showed the heterogeneity of the system, with variable changes in the area around 1338 cm^{-1} – with mixed contributions of proteins and nucleic acids –, in the lipid band at 1445 cm^{-1} and on the side of the Amide I at 1674 cm^{-1}.

5.4.4 Principal Component Analysis of the PDT Data

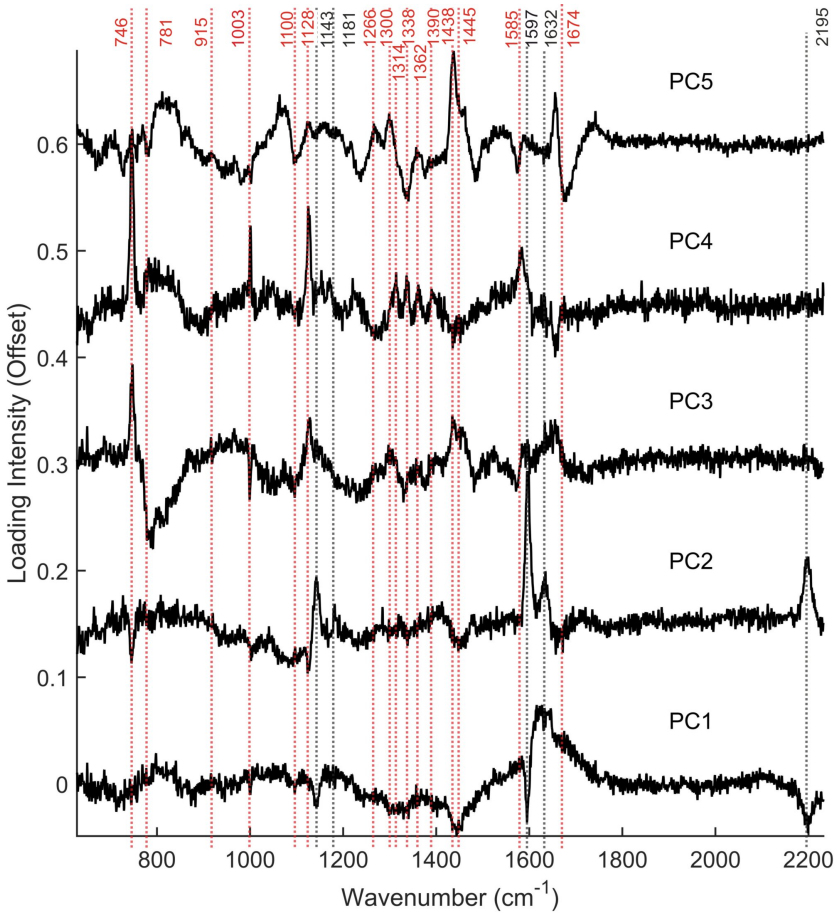


Figure 5.36: PCA loadings for the first 5 PCs for the PDT data. The main peaks have been labelled.

To better understand the data, PCA[1] was performed on all the DC473-PDT data, with variance explained for the first 5 coefficients of 55%. Figure 5.36 shows the loadings for the first 5 coefficients, where the main peaks have been labelled. Figure 5.37 shows the variation of the average scores for each of the conditions and cell lines with time – all traces were

[1] Principal Component Analysis

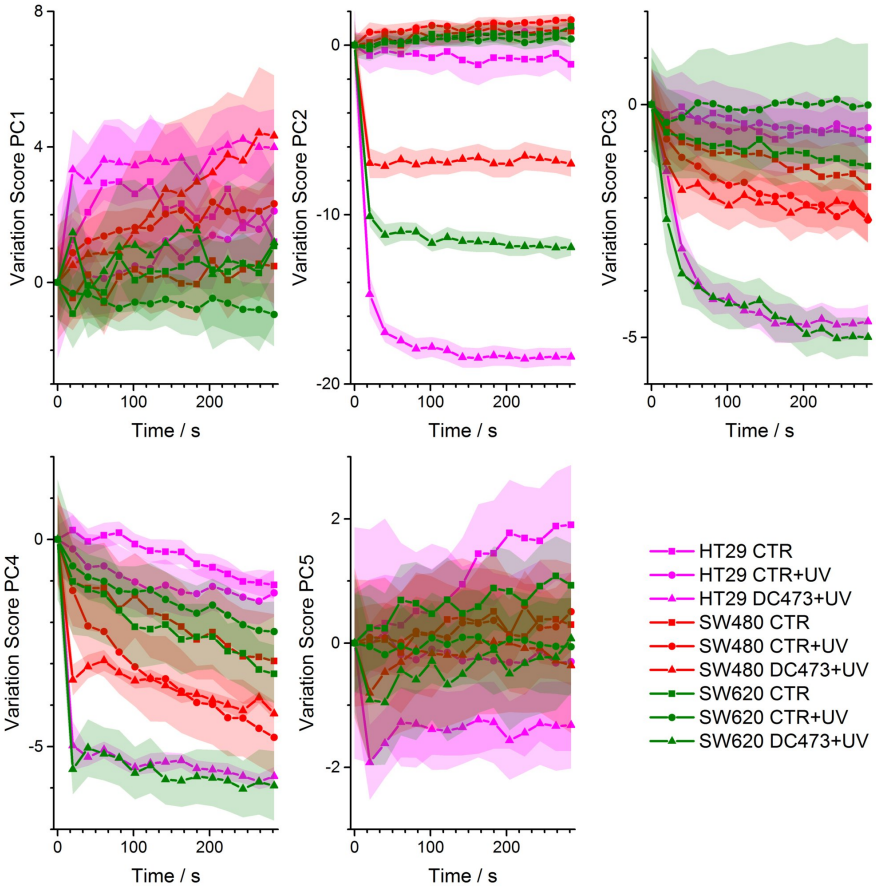


Figure 5.37: PCA average score-time traces for the first 5 PCs for the PDT data. All curves have been translated so the first point of each curve is at zero to better see the time trends. The area around the curve represents the standard error.

aligned to start at zero to better see the trends with time. PC1, similarly to PC5, accounted mostly for substrate variability, showing variability between cells within a cell line and between cell lines, not correlated with the time dependency.

PC2 contained the main DC473 contributions, with all the DC473 bands (1443, 1181, 1597, 1632 and 2195 cm^{-1}) present. It showed barely any time changes for the control samples and exponential decay time dependence for the DC473 samples for all cell lines. It also had negative contribution at 746 and 1128 cm^{-1}, indicating a possible negative correlation with cyt C. This seems contradictory with the previous DC473 accumulation results, where it showed a positive correlation with the 746 cm^{-1} cyt C, but it is likely due to the presence of multiple observations post UV treatment where the DC473 bands are weakly present and the cyt C bands drop. The biggest drop in this coefficient was on HT29 cells, followed by SW620 cells and then SW480 cells.

PC3 contributed mostly with positive cyt C bands (746, 1128 and 1585 cm^{-1}) and negative contributions in the region around 780-840 cm^{-1}, at the Phe band at 1003 cm^{-1} and a broad positive contribution at around 1455 (lipids) and 1656 cm^{-1} (Amide I). This coefficient could be indicating that a drop in cyt C bands was accompanied by a drop in lipids and protein. The scores of PC3 showed a slight and constant negative slope for the control samples, with a more drastic exponential-like reduction for the DC473+UV sample right after PDT. The HT29 and SW620 PDT samples showed very big drops in PC3 scores, with smaller changes for the SW480 PDT sample, that overlapped with the SW480 CTR+UV sample.

PC4 had a similar loading to PC3, but with a positive band around 780-840 cm^{-1}, a negative band around 1455 cm^{-1} and negative or neutral contribution around 1656 cm^{-1}. Additionally, it showed a series of repetitive bands at 1314, 1338, 1362 and 1390 cm^{-1}. The scores for PC4 showed a similar trend to PC3, with a low slope reduction for some of the control traces and a sharp reduction for the DC473+UV samples. Again, HT29 and SW620 cells showed a stronger reduction in this coefficient after PDT, with lower responses and overlap of SW480 DC473+UV and CTR+UV samples.

PCA was unable to identify any clear Raman marker for the cell blebbing morphology changes observed after PDT, and only showed a strong reduction on PC3 and 4 on all treated cells, in line with the amount of DC473 that the cells accumulated according to previous experiments.

A 2D plot of the single-cell score traces for PC2 and PC4 is shown in figure 5.38 as a summary of the PCA findings. Here, the first point for each cell was shown with a marker, and the remaining points were shown with a line trace. This figure shows the overall picture of the cell responses to the compound, with PC2 showing DC473 contents, with a clear distinction between control and treated cells, and PC4 showing mainly a drop in cyt C bands, referring to a decrease in cell viability. In the control groups, HT29 cells showed Brownian-motion like traces in this plot, whereas SW480 and SW620 cells showed more variability on PC4 score and a drop in the score over time, likely due to the phototoxicity of the 532 nm excitation laser and the UV light. However, this was a constant reduction in the PC4 score, compared to the drastic reduction observed on the PDT samples.

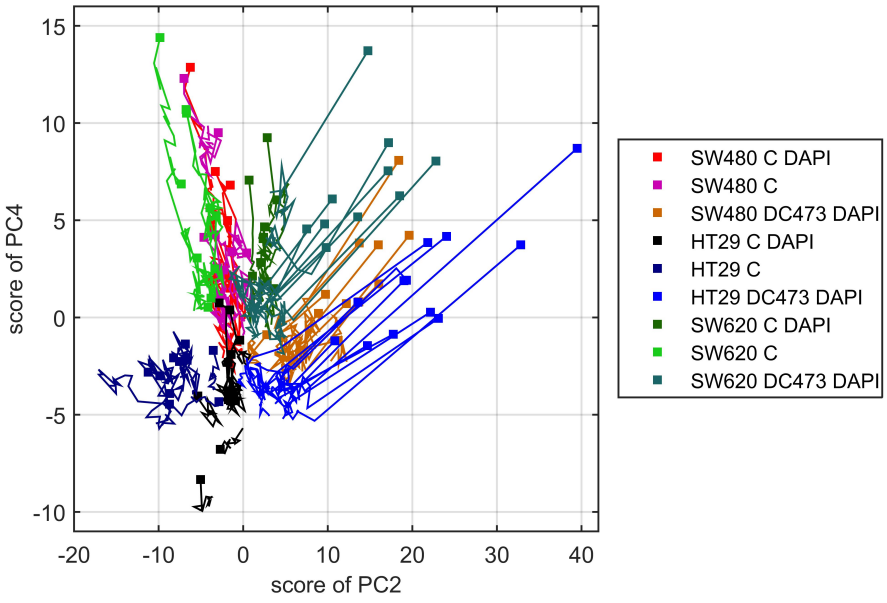


Figure 5.38: PCA traces for PC4 and PC2 for the PDT data, where the first time point is shown as a square, and remaining observations in that cell are shown as a trace.

5.4.5 PCA/LDA Analysis of the PDT Data

An alternative approach to identify PDT biomarkers is to train a PCA/LDA model to separate between control and treated cells. For this, cells were separated into two groups: control and DC473 treated cells before UV light on the first group and DC473 treated cells after UV light on the second group. When doing a 5-fold validation, the model showed a correct rate of 98.2±0.2%, with a sensitivity of 98.9±0.2% and specificity of 96.9±0.3%. Running 200 models, the average LDA loading after smoothing is shown in figure 5.39C, with a black shadow showing one standard deviation.

The first thing worth noting was in relation to the DC473 bands. Interestingly, the 1146 and 1642 cm^{-1} bands were positive, whereas the two strongly negative bands in the vicinity to the two stronger DC473 peaks were negative: around the 1597 cm^{-1} Phe band there was a positive contribution at 1591 cm^{-1} and a strong negative contribution at 1608 cm^{-1}, and around the 2197 cm^{-1} alkyne band there was a positive contribution at 2190 cm^{-1} and a strongly negative contribution at 2210 cm^{-1}. Given that the loading had negative scores for the post-PDT samples, this indicated that these two bands were shifting to higher wavenumbers after PDT. However, the other DC473 bands present did not show the same behaviour in the LD, simply showing a drop in contribution.

From the remaining cell bands present, the 747 cm^{-1} cyt C was not chosen as a good biomarker, probably due to a drop in this band also in the control groups likely due to phototoxicity of the laser. However, the 1228 cm^{-1} band – on the side of the Amide III – was seen to drop in the post PDT groups more than in the control groups. In the area between 1250 – 1550 cm^{-1}, multiple bands were present, with a duo of positive contributions at 1340 and 1368 cm^{-1} – with mainly mixed assignments – and a trio of positive contributions at 1419, 1442 and 1467 cm^{-1}, in the area of the broad lipid band at 1455, confirming the

previous observation that post-PDT samples showed a drop in lipids. Among the remaining peaks, the LDA loading showed a positive contribution at 791 followed by a sharp negative contribution at 807 cm^{-1}, but the meaning of this contribution is unknown.

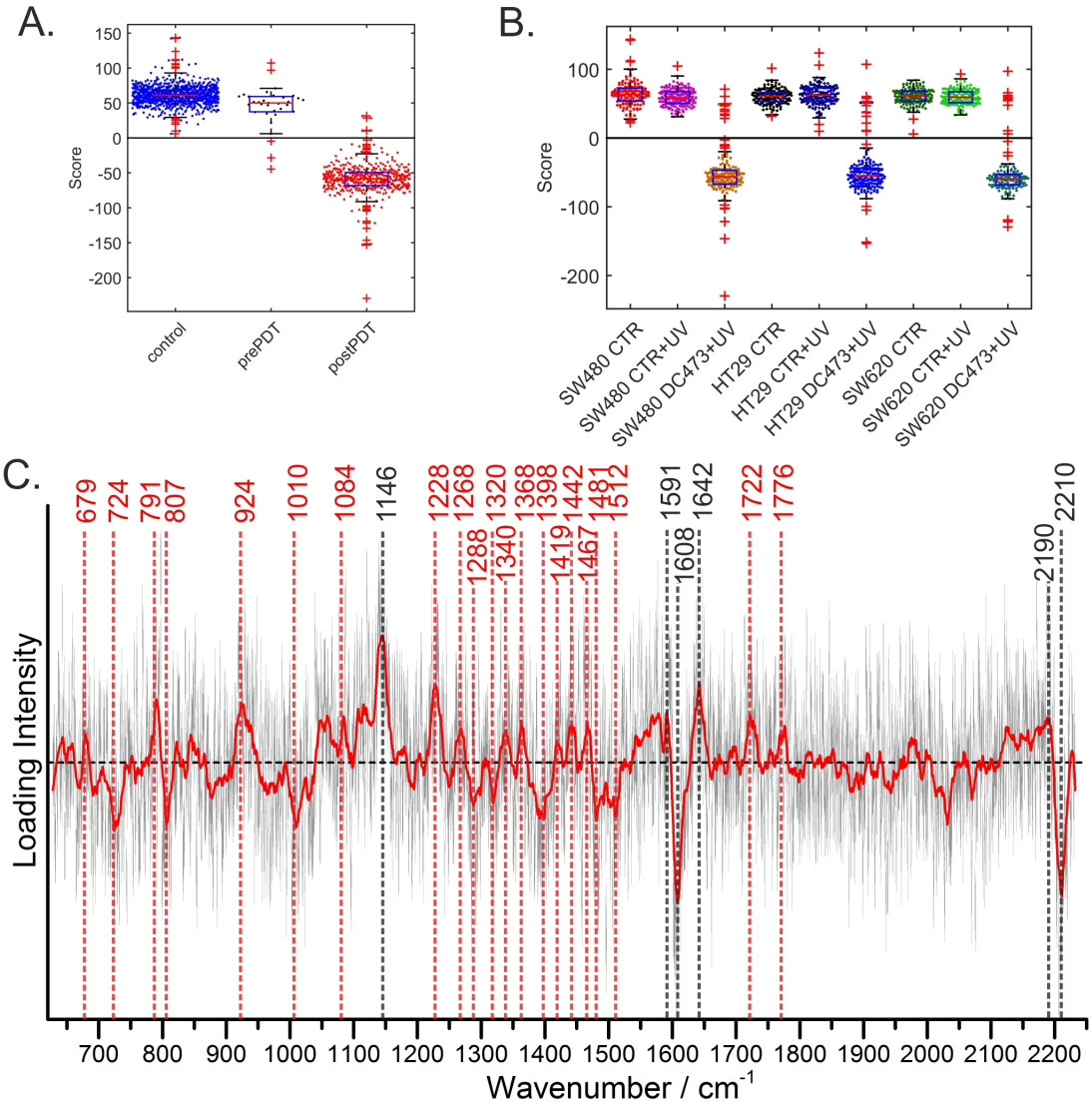


Figure 5.39: A. Boxplots and beeswarms of the LDA scores for the data when separating by control, pre-PDT and post-PDT. **B.** Boxplots and beeswarms of the LDA scores when separating by cell line and experiment. **C.** LDA loading average for 200 LDA models ran on the PDT data. The standard deviation is shown as a shadow and a smoothed loading is shown in overlayed in red to ease the analysis. A line at zero is shown in dashed black. The main peaks have been labelled, where black peaks were assigned as DC473 and red peaks were assigned as cell contributions.

5.4.6 Band Fitting of the PDT Data

Following the LDA results, the peaks at 1597, 2195, 747 and 1228 cm^{-1} were fitted or integrated to see their trend with PDT, analysed in this section.

DC473 bands fitting For the DC473 bands, the data was first thresholded by fitting amplitude to remove those fittings where the band was not present. This was done so that the amplitude minus the error was higher than 0.12 for the 1597 cm^{-1} band and 0.05 for

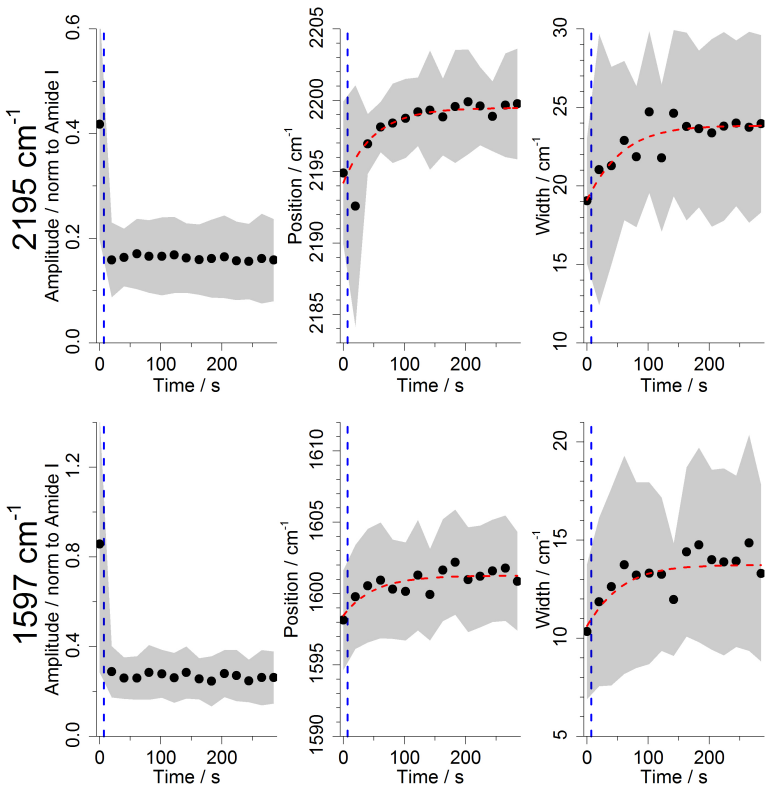


Figure 5.40: Average time dependence of the fitting parameters for the 2195 cm^{-1} band (top row) and 1597 cm^{-1} band (bottom row), for all cell observations after amplitude filtering. UV dose has been represented with a vertical blue dashed line. For both cases, the amplitude shows a strong drop after the cell was exposed to UV, with not all DC473 treated cells having any remainder photosensitiser signal. However, the remainder signal showed a slow shift to higher wavenumbers for both peaks, with also a slow broadening of the band. Red dashed lines show the fittings done on the data, and the area behind the curve corresponds to one standard deviation.

the 2195 cm^{-1} band. Figure 5.40 shows the fitting average results for all cells after the amplitude filtering on both DC473 bands, where the area around the curves corresponds to one standard deviation, and each data point was an average of a minimum of 20 fittings. Results showed a sharp drop in the amplitude of both bands caused by the UV irradiation, followed by a plateau. However, the position and widths of the bands showed a slower time-dependent trend post-UV exposure. The four curves were fitted to a sigmoid function as shown in equation 5.1 to better quantify the time dependence of the changes, where a is the asymptote of the sigmoid (in cm^{-1}), b is the ratio between the asymptotic shift and the initial value (adimensional) and τ (Hz) is the inverse of the lifetime. The values for the parameters of the fitted curves are shown on table 5.6 ($\pm$ standard errors).

$$y = \frac{a}{1 + b \cdot e^{-\tau \cdot t}} \tag{5.1}$$

Table 5.6: Time dependence sigmoid fittings of the DC473 average band GLS fitting parameters: band position and GSL width. The corresponding fittings are shown in figure 5.40

	Position / cm^{-1}			Width / cm^{-1}		
Band	a	b	τ	a	b	τ
2195 cm^{-1}	2199.5 ± 0.2	0.0024 ± 0.0003	0.020 ± 0.004	23.9 ± 0.4	0.25 ± 0.04	0.021 ± 0.007
1597 cm^{-1}	1601.3 ± 0.3	0.0018 ± 0.0004	0.02 ± 0.01	13.7 ± 0.4	0.29 ± 0.08	0.02 ± 0.01

When looking at the position of the bands, the 2195 cm^{-1} band showed a slow shift towards higher wavenumbers, saturating at 2199.5 cm^{-1}, for a total shift of $\approx$ 4.8 cm^{-1}. The 1597 cm^{-1} band had a smaller shift 1601.3 cm^{-1} (total shift of $\approx$ 3.6 cm^{-1}). In terms of the GLS width, the alkyne band showed an increase in width of $\approx$ 4.9 cm^{-1} whilst the sharper phenyl band showed a modest increase in width of around $\approx$ 3.0 cm^{-1}. Interestingly, all bands showed a similar τ constants of 0.02 s^{-1}, indicating that both the broadening and the position shift of the bands are likely due to the same biological effect on the remaining DC473. The inverse of τ, or the lifetime of these events, is of approximately 50 s.

The two most susceptible cell lines to PDT – based on morphology – both showed remaining DC473, whereas HT29 cell showed a complete absence of DC473 post-PDT in most observations. This could be indicating, that DC473 localized in two different compartments in the cell, one where it is susceptible to chemical reaction following irradiation but less active and another compartment where only a minority of DC473 molecules locate, where it is less susceptible to the reaction following irradiation but causes more cell damage. In this case, HT29 cells would have most of the compound located in the first compartment, whereas SW480 and SW620 cells would have a greater contribution in the second one. We hypothesize that the first compartment with less active DC473 are lipid droplets, whereas the second compartment would likely be in the cytosol or mitochondria. The remaining DC473 after irradiation would then be showing shifts in position and widening of bands due to changes in the environment after activation.

Cell related bands The fittings to the 747 cm^{-1} cyt C peak (figure 5.41) showed no correlation with the experimental settings used, with all samples showing a slow decrease in

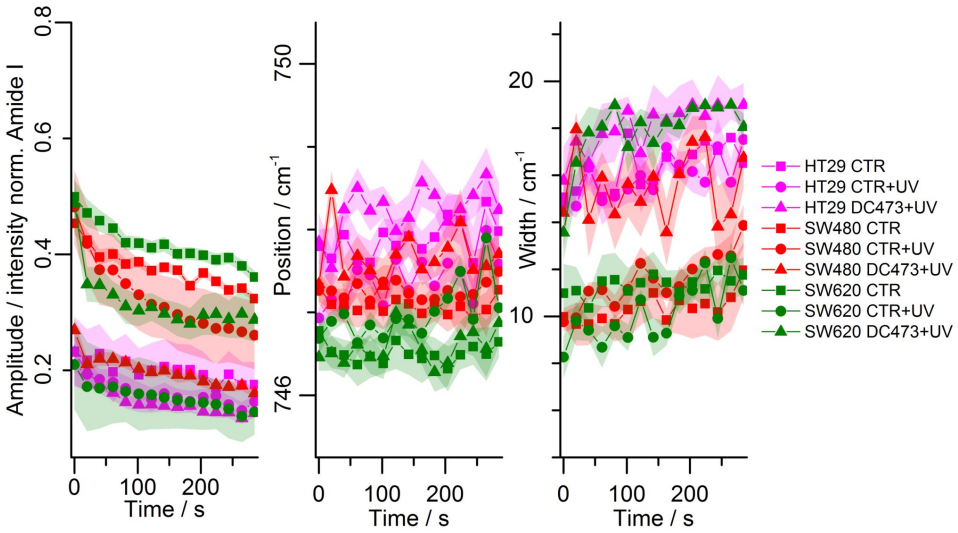


Figure 5.41: Fitting results of the 747 cm^{-1} cyt C band. The intensity showed a drop for all samples, with a certain variability between different experiments for the same cell line. The peak Position seemed to show that the peak is in general shifted to higher values for HT29 cells compared to SW620 or SW480 cells. Width showed in general wider peaks for the treated samples than for the untreated samples.

intensity over time – agreeing with the PCA/LDA results. Intensity also showed variability between the controls for the same cell line. The position seemed to show that the peak was in general shifted to higher values for HT29 cells compared to SW620 or SW480 cells. However, the data is in general overlapping and did not show a significant trend over time. Width showed in general wider peaks for the treated samples than for the untreated samples for SW480 and SW620 cells, with also wider bands for HT29 CTR+UV cells.

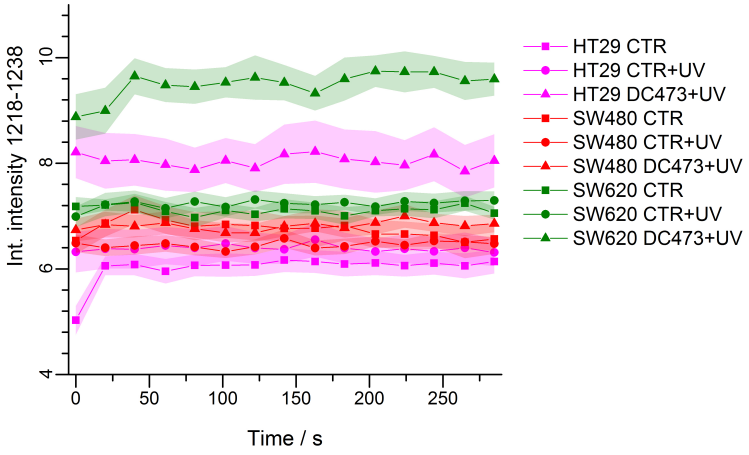


Figure 5.42: Dependence of the integrated intensity from 1218 to 1238 cm^{-1} with time for all the experiments and cell lines, where the area around the curve corresponds to the standard error.

Following the presence of a band at 1228 cm^{-1} on the LD loading, this region of the spectrum was investigated. As the amide III band overlaps with this region, the intensity was integrated from 1218 to 1238 cm^{-1} instead, with the integrated intensity traces shown in figure 5.42. This band showed slightly different contributions for different cell lines, with HT29<SW480<SW620. Results showed a significantly higher intensity for HT29 and SW620 cells after incubation with DC473 compared to their controls, whereas SW480

DC473 cells showed similar values to the SW480 controls. However, the band showed no significant time dependence.

5.4.6.1 Viability Assays of Cells incubated with DC473

Collaborators in LightOx provided preliminary PI[1]-FDA[2] live-dead assay results for the model system cell lines after incubation with DC473 and exposure to light. Cells were seeded at 8000 cells/well in 96 well plates and incubated for 2 days, adding the compound at day 3 and incubated for 24 h, when the light exposed group was exposed to 6.9 J/cm^2 of 405 nm laser light. Cells were then incubated with both PI and FDA according to manufacturer recommendations and fluorescence was read in a plate reader.

For the live-dead assay, both a cell enzyme activity and a membrane permeability tests are done. FDA is non-fluorescent but is taken up by live cells which convert it into the green fluorescent metabolite fluorescein via esterase. In contrast, the nuclei staining dye PI cannot pass through a viable cell membrane and is only uptaken by cells whose membrane is compromised, becoming fluorescent after intercalation with DNA. Results for the assays facilitated by LightOx are shown in figure 5.43, where A. 1, 2 and 3 show the live dead staining for each of the cell lines, and B. and C. show the overlap results for all cell lines for the fluorescein fluorescence and the PI respectively. The concentrations tested were of 0.5, 1, 2.5, 5 and 10 μM DC473 and all results are shown normalised to the 0.1% DMSO control.

All cells showed a relatively flat fluorescein fluorescence for the dark results, that dropped for 10μM for SW480 and HT29 cells, and at 5μM for SW620 cells after a 24 h incubation. The PI results showed a generally flat dark fluorescence, meaning that even though the cells have a compromised proliferation for high concentrations of DC473, they do not show loss of membrane integrity after 24h. This confirms that the accumulation of the compound at the concentration used in the present report (4 μM) is generally non-toxic for the cells, although can reduce proliferation, which could be related with the observed drop in lipids contributions in the Raman results. Interestingly, for the SW620 cells, some cells incubated with low concentrations of DC473 showed a higher fluorescein fluorescence than the control under dark conditions, meaning that low concentrations of the compound could increase proliferation. This observation would need repeating to confirm the results.

After exposing the cells to the UV light, all cells showed a drop in fluorescein fluorescence and an increase in PI fluorescence. For the fluorescein results, SW480 and SW620 cells showed this drop already at the lowest concentration used (0.5 μM), whereas HT29 cells showed a fluorescein fluorescence drop at higher concentrations of DC473, showing less susceptibility for PDT treatment with DC473. All cells showed an increase in PI for the higher concentrations of DC473. Interestingly, after exposing the control cells or the DMSO cells to the light, no differences could be observed for nor the fluorescein or the PI, meaning that in the time-scale of this measurement cells did not die solely due to the UV light dose. This confirms that the laser acquisition light is likely the cause of any toxicity shown by the

[1] Propidium Iodide
[2] Fluorescein diacetate

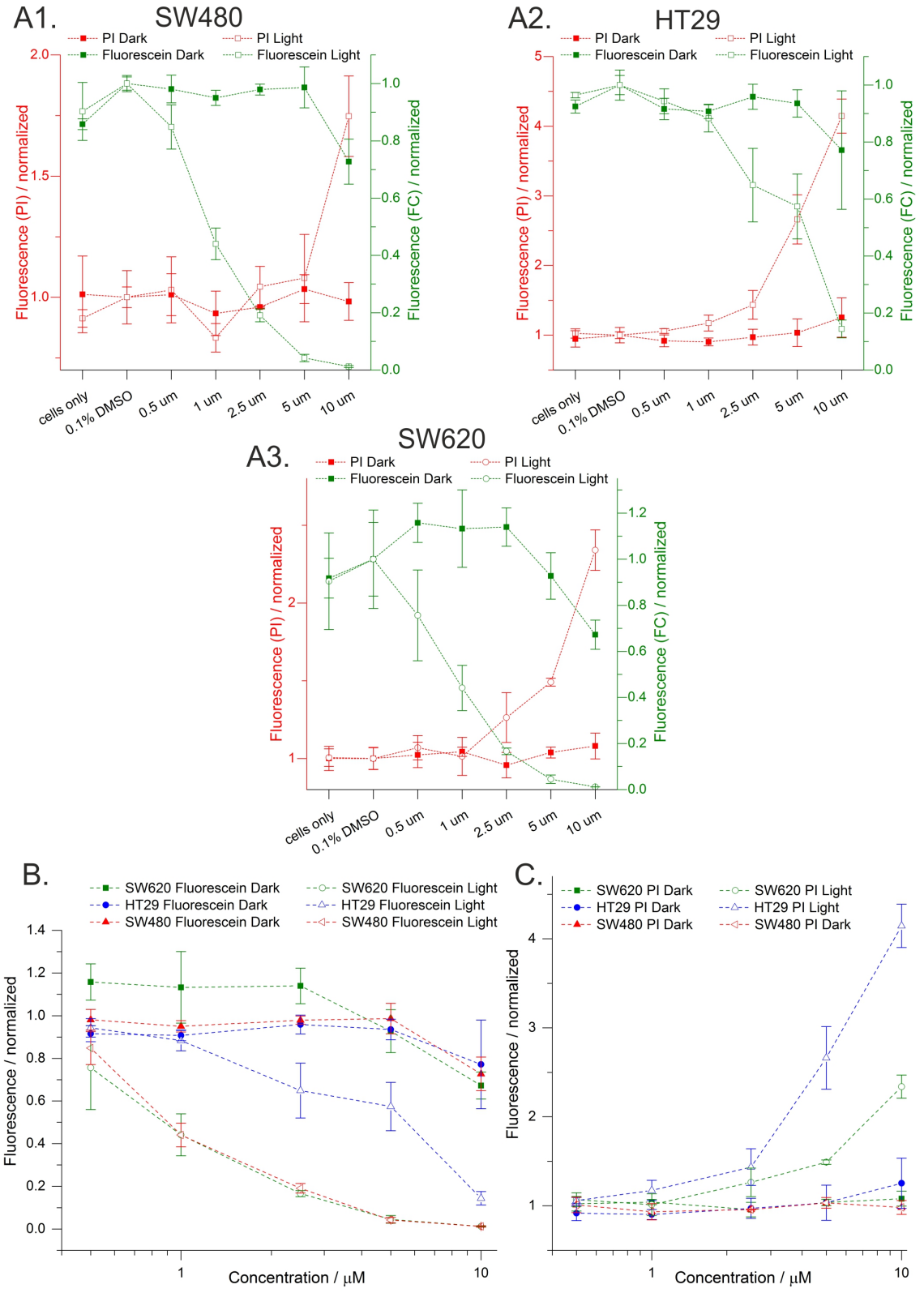


Figure 5.43: Results for the PI-FDA live-dead assay for cells treated with DC473 and light. **A.** 1, 2 and 3 are the results for the PI and FDA for each of the cell lines respectively. **B.** Fluorescein fluorescence with DC473 dose for all cell lines. **C.** PI fluorescence for all cell lines. All the data was normalised to the 0.1% DMSO well fluorescence and the error bars correspond to the standard deviation.

control groups in the PDT set of experiments. In the future, an MTT assay will be done to observe the mitochondrial activity of the cells after DC473 PDT treatment.

5.4.6.2 Conclusions and Future Experiments

The DC473 experiments showed that this molecule can be readily uptaken by different CRC cells following incubation. Accumulation mainly occurs in lipid droplets and in the cytosol, presumably in mitochondria, and was significantly higher for more advanced adenocarcinoma HT29 and SW620 cells compared to SW480 cells, likely linked to their higher lipid droplets content. Samples incubated with DC473 also showed significantly lower lipid contents, with lower integrated intensity in the CH stretching region. However, when activating the compound, SW480 and SW620 showed clear morphology changes in response to the drug, whereas HT29 cells showed no morphology changes. The compound's signal dropped upon activation with UV light, with a certain remaining compound with the UV light dose used for SW480 and SW620 cell, and minimal compound left in HT29 cells. Both PCA and LDA identified coefficients that showed time-dependent changes in multiple cell peaks over time after DC473 activation, including the 1228 cm^{-1} band on the side of the amide III and multiple bands between 1250-1550 cm^{-1}. The main two DC473 peaks at 1597 and 2195 cm^{-1} showed a strong drop in intensity after irradiation and additional shift to higher wavenumbers and increase in peak width over time. These effects possibly indicated relocalization of the compound, showing a similar time dependence (lifetime of 50 s). All these results indicate that the compound could be suitable for CRC PDT, however, its effects in lipid metabolism need confirming and some cells could be resistant to this concentrations and light doses, such as the data for HT29 seems to indicate. Future experiments should be done to fully understand the uptake mechanism and the UV light decomposition mechanism, and to confirm that the compound is accumulated in lipid droplets and not other similar vesicular organelles.

Chapter 6

Raman and Microfluidics

In the previous chapter, two PS[1]s were studied: PPIX[2], a traditional PS, and DC473, a novel PS. Experiments were done on live cells for detection and quantification of the amount of PS accumulated (relative to the Amide I band), and Raman was used for tracking the activation of the PS following its activation with UV[3] light. However, all studies reported in that chapter compared a control cell line subpopulation and a subpopulation incubated with the compound, obtaining information about the within sample heterogeneity, but being unable to link the initial cell phenotype with its accumulation of the compound. Working with cell traps and microfluidics would allow performing experiments where it is possible to take measurements before and after incubation in the same cells.

As was discussed in section 1.7 (page 18), the most extensively used material for microfluidic fabrication is the polymer PDMS[4], mainly due to its low cost. PDMS has multiple Raman peaks from 150 to 1409, with the strongest peaks occurring at 2903 and 2964 cm^{-1} due to the $C - H_3$ stretching.[20] This is a major challenge when working with RS[5] on a microfluidic system, as most of the PDMS bands are of the same order of magnitude or higher than the cell bands. To minimize the background contributions, an alternative is changing the materials for fabrication and obtaining glass devices instead. The fabrication of a glass device requires etching methods, where obtaining trapping structures with a high aspect ratio is very challenging. Glass devices also require etching, which requires special facilities not available in this Physics facilities. For this reason, the present report focused on the use of PDMS devices.

This chapter discusses different approaches to minimizing and working with this PDMS background, including coatings of PDMS (both metal and nanocomposites), alternative microfluidic fabrications, increasing the confocality and channel depth and PDMS contributions fitting and subtraction. It will additionally use a microfluidic system to track the accumulation of DC473 in live single-cells *on-chip*, and discuss data of an ongoing collaboration studying single *Escherichia coli* bacteria *on-chip*.

[1] Photosensitiser
[2] Protoporphyrin IX
[3] Ultraviolet
[4] Polydimethylsiloxane
[5] Raman spectroscopy

6.1 What are the Chemical Components of PDMS?

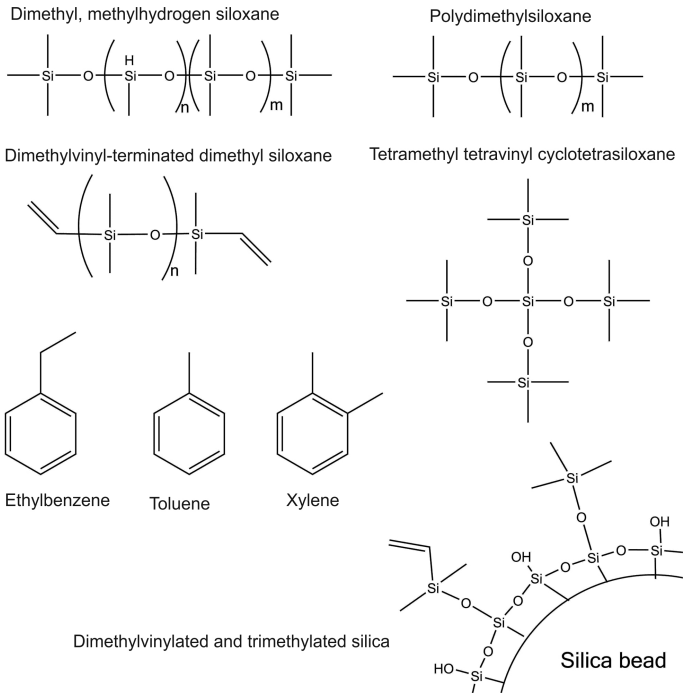


Figure 6.1: Chemical formula for the components of PDMS Sylgard 184 (Dow Corning).

PDMS[1] is a polymer sold as Sylgard 184 (Dow Corning) in two separate components to be mixed in a ratio 10:1: a Base resin and a Crosslinker or Curing Agent. The Base resin contains 0.5 wt% xylene, 0.2 wt% ethylbenzene, >60 wt% dimethylvinyl-terminated dimethyl siloxane (CAS Number 68083-19-2), 30 to 60 wt% dimethylvinylated and trimethylated silica (CAS Number 68988-89-6) and 1-5 wt% tetra(trimethylsiloxy) silane. The curing agent contains 0.19 wt% xylene, <0.1 wt% ethylbenzene, 55-75 wt% dimethyl, methylhydrogen siloxane, 15-35 wt% dimethylvinyl-terminated dimethyl siloxane, 10-30 wt% dimethylvinylated and trimethylated silica and 1 to 5 wt% tetramethyl tetravinyl cyclotetrasiloxane.[299, 300] The chemical structures for each of these components are shown in Fig 6.1. The general chemical formula for PDMS is also shown. However, it must be remembered that the polymer will contain a mixture of the different components, with branches due to reactions between methylhydrogen siloxane and dimethylvinylated and trimethylated silica, and covalently bonded silica beads. Both ethylbenzene, toluene and xylene are volatile, and would disappear in the baking steps for crosslinking of PDMS, and are not expected in the final product.

Figure 6.2 shows the Raman spectrum of crosslinked PDMS in a microfluidic device, when measuring through a bottom bonded coverslip exciting with the 532 nm laser. Table 6.1 gives the bands present in the spectrum and their assignments in the literature, mainly based on Cai *et al.*.[302] The absence of bands at ca. 1680 and 367 cm^{-1} indicate that the presence of uncrosslinked vinyl containing species was too low to detect it with RS[2].[304] In general, the PDMS Raman spectrum shows multiple contributions along the whole fingerprint region.

[1] Polydimethylsiloxane
[2] Raman spectroscopy

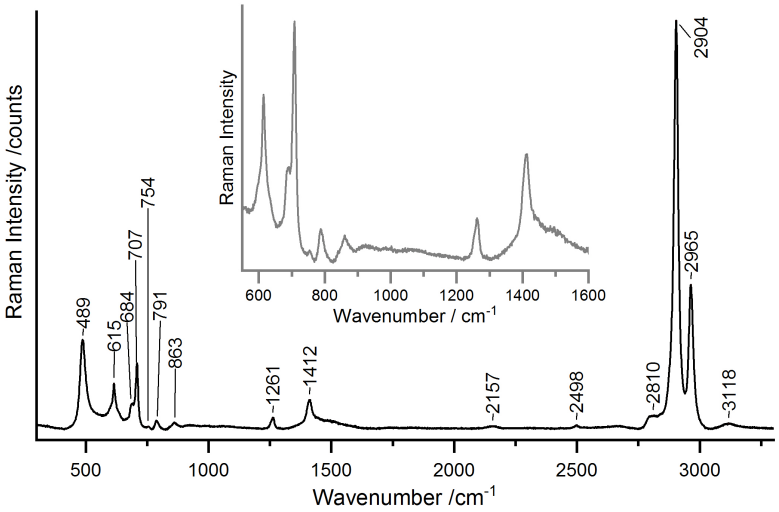


Figure 6.2: Raman PDMS spectrum excited with the 532 nm laser, where the positions of the main bands have been indicated, and the fingerprint region has been zoomed in the inset figure (grey axis). Tentative assignments of the bands based in the literature are listed in table 6.1

Table 6.1: PDMS bands positions, relative intensities to the maximum peak and tentative assignments based in the literature.

P / cm^{-1}	I / %	Literature / cm^{-1}	Reference	
186	6.8	190	[301]	C – Si – C wag
489	21.4	490.8/488	[302, 303]	Si – O – Si str
615	11.0	616.8	[302]	
684	5.9	689.1/687	[302, 303]	Si – CH_3 sym. rocking
707	14.8	711.8/708	[302, 303]	Si – C sym. str.
754	0.5	757.1	[302]	
791	1.6	790.8/787	[302, 303]	CH_3 asym. rocking and Si – C asym. str.
863	1.3	863/862	[302, 303]	CH_3 sym. rocking
1261	2.5	1266.1	[302]	CH_3 sym. bending
1412	7.0	1412.5/1411/1412	[301–303]	CH_3 asym. bending
2157	0.6	2152 (IR)	[302]	Si – H str.
2498	0.8	2500.7	[302]	glass coverslip
2810	3.1	2790.3	[302]	
2904	100.0	2890.5/2906/2907	[301–303]	CH_3 sym. str.
2965	35.2	2940.5/2965	[302, 303]	CH_3 asym. str.
3118	1.2			

6.2 Characterization of the Substrate Background on a PDMS Device

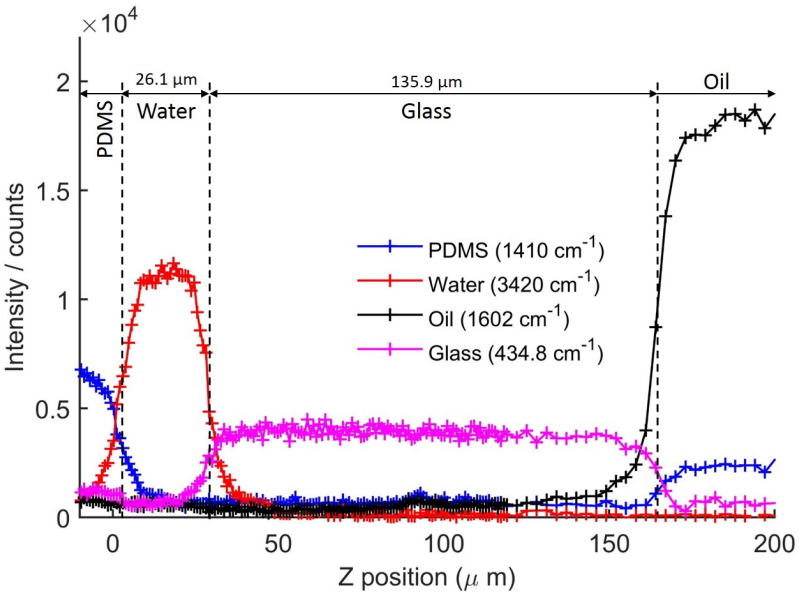


Figure 6.3: Raman depth scan on a microfluidic device using the 532 nm excitation laser and the 100× oil immersion objective, where the indicated peaks of each component were tracked as a function of depth (Z).

To show the contributions of each of the potential backgrounds to the signal within the channel, depth scanning of a 25 μm deep channel in a microfluidic device filled with water was done using the 532 nm laser, acquiring from 100 to 4500 cm^{-1} to also include water band at ca. 4000 cm^{-1}. The measurements were made with the 100× oil immersion objective using *standard* confocality and 0.3 s per spectrum with a Z step of 0.9 μm. Figure 6.3 shows the profile results, and figure 6.4 shows an example of spectra at different depths, with the main tracked bands indicated. The bands were not chosen as the maximum for each spectrum as they could overlap, but as contributions that would minimize overlapping.

Results showed the profile of each of the background contributions throughout the device, with a strong oil contribution while focused in the oil for the objective (Type F Immersion Oil, Leica), that transitions into a glass contribution while measuring on the coverslip, then a water contribution in the chamber and finishing with a PDMS[1] contribution past the microfluidic chamber. When measuring with the 100× oil immersion objective and *standard* confocality, the chamber always had a contribution of either the glass or the PDMS, that never reduced completely, which is more obvious when observing the contribution of the example spectra in the chamber on figure 6.4 in the CH-stretching region. The following experiments were done to minimize the contribution of this background.

6.2.1 Metal Evaporation on PDMS Devices for Background Reduction

When choosing the materials of the microfluidic device, it is known that perfect conductors have no Raman signal. Most bulk metals have such weak Raman signals that cannot be detected using conventional Raman spectroscopy.[305] However, thin metal films have been shown to produce Raman signatures, as the conductivity of nanometre thin films is

[1] Polydimethylsiloxane

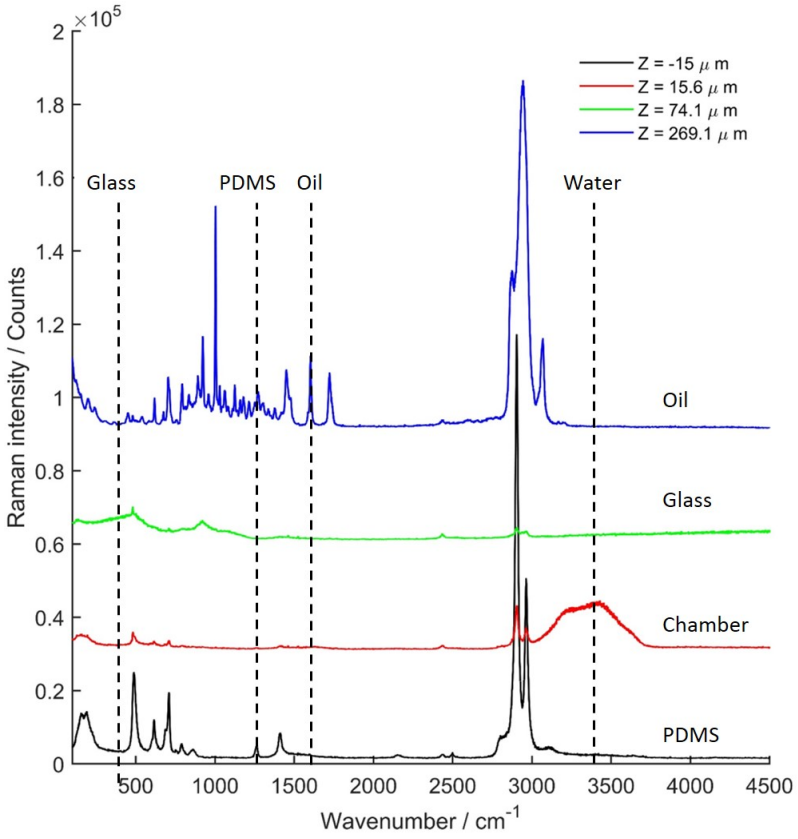


Figure 6.4: Example of spectra for the Z-scan in a 25 μm microfluidic device shown in figure 6.3 at different heights, representing an example for the immersion oil, the glass, the water and the PDMS. The peaks tracked in figure 6.3 are indicated with a dashed line.

low enough to allow photons to reach the metallic lattice bonds.[20] A reflective metal film coating the PDMS could minimize the contributions of the PDMS to the Raman spectrum.

In order to shield the PDMS and minimize the background, evaporating of metals was done onto PDMS devices following the methods detailed in subsection 3.8.3 (page 81). Aluminium was evaporated onto PDMS, tracking the evaporated mass to a final thickness of 50 nm, that was then bonded to a glass coverslip. However, the bonding of the PDMS to the coverslip is usually achieved by activation with oxygen plasma treatment, that could not be achieved with metal coated PDMS. Bonding was obtained by either manually cleaning the PDMS around the device, removing the metal layer, before O-plasma treatment, or by removing the metal layer using scotch tape. Figure 6.5A shows a picture of the microfluidic device with an evaporated 50 nm Al layer and figure 6.5B shows and schematic of the fabricated device. When observing the layer of Al on PDMS under the bright field microscope, the layer showed cracking, likely due to bending of the flexible PDMS layer after O-plasma treatment in order to bond it to the glass coverslip – see figure 6.5C.

When doing a Raman measurement on the Al-coated device using the 532 nm excitation laser, damage to the Al layer was observed, as shown in figure 6.5D and E. The signal at the surface of either the PDMS or the Al-coated PDMS was acquired using the 63× oil objective, where the intensity was minimized to reduce the damage on the Al layer (532 nm laser, 2.2 mW, 2.5 s). The signal comparison – acquired with the same settings in an area with and without Al in the device – is shown in figure 6.6. The Al layer shielded most of the PDMS background but caused a strong baseline on the spectrum.

Given the strong damage the visible laser had on the aluminium layer, gold coated PDMS devices were prepared instead. Gold is more expensive than aluminium but shows better

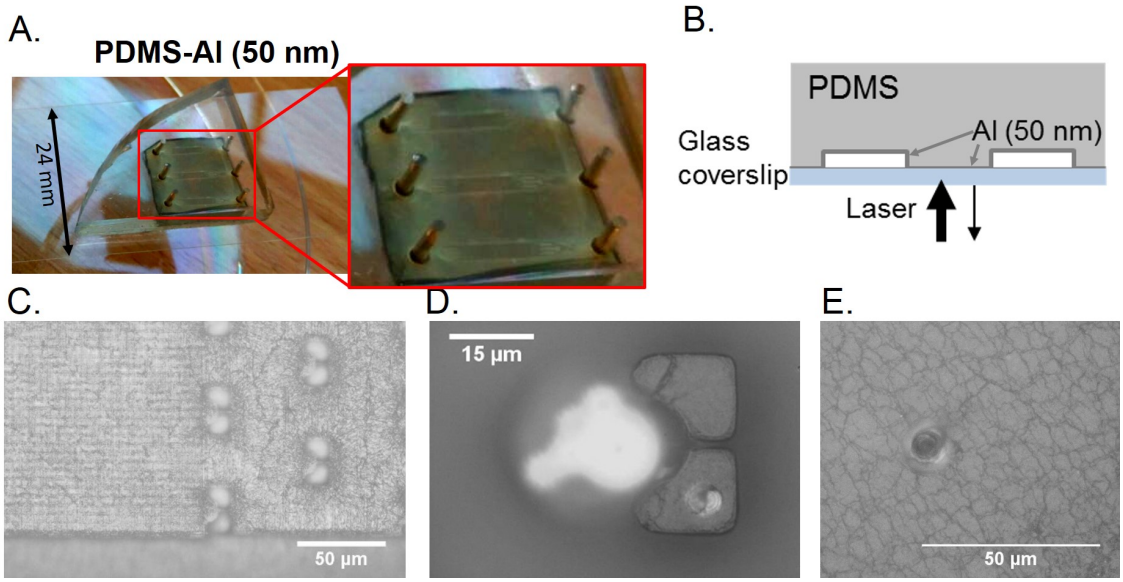


Figure 6.5: A. Picture of the microfluidic device, where the area around the traps is not transparent due to the layer of aluminium evaporated. Bonding was done by cleaning the PDMS around this area, obtaining a device with minimal leakage around the chambers but good surrounding overall bonding. **B.** Schematics of the final device, where the PDMS was covered by an Al film (50 nm) and then bonded to the coverslip, and the Raman signal was acquired through the glass coverslip in an inverted setup. **C.** Trapping area when focusing on the Al layer on the top of the channel, where a very cracked structure can be observed, likely due to bending the PDMS during the fabrication of the devices. **D.** Laser damage on the aluminium layer after exposing it to the visible excitation laser (532 nm, 22 mW, 1 s) using the 20× objective. **E.** Laser damage when using the 40× objective (532 nm, 11 mW, 5 s).

stability and biocompatibility compared to other metals easy to evaporate like silver. Gold coating on PDMS was previously reported by [306] and [307] to create electrodes in soft materials. To promote the adhesion of the Gold layer, an additional metal layer is often used. A layer of 5 nm of Chromium was evaporated between the PDMS and the Au, as shown in figure 6.8A. The Cr-Au-PDMS device showed no laser damage (22 mJ, 532 nm) as shown in figure 6.8B and the comparison of the background intensities obtained for the visible laser are shown in figure 6.7. After filling the microfluidic chamber with water, Raman spectrum acquisition with the visible laser caused the appearance of big air pockets. This could be explained by overheating due to interactions of the laser light with the gold layer,

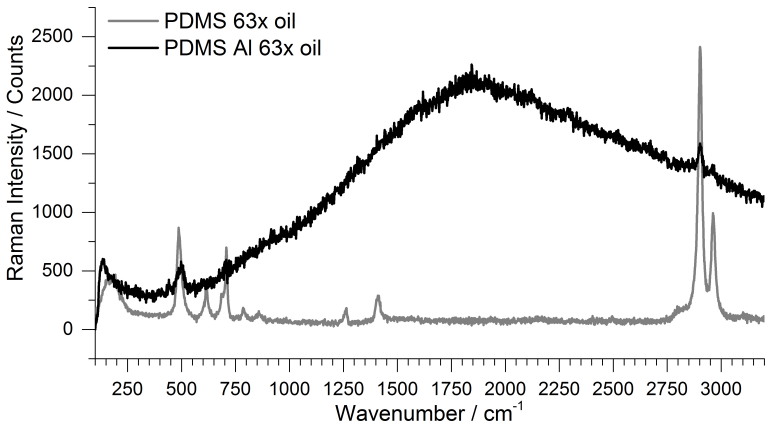


Figure 6.6: Spectrum of PDMS with and without an Al evaporated layer excited with the 532 nm laser. The Al coated PDMS showed a reduction in all the main PDMS bands, but a strong baseline contribution.

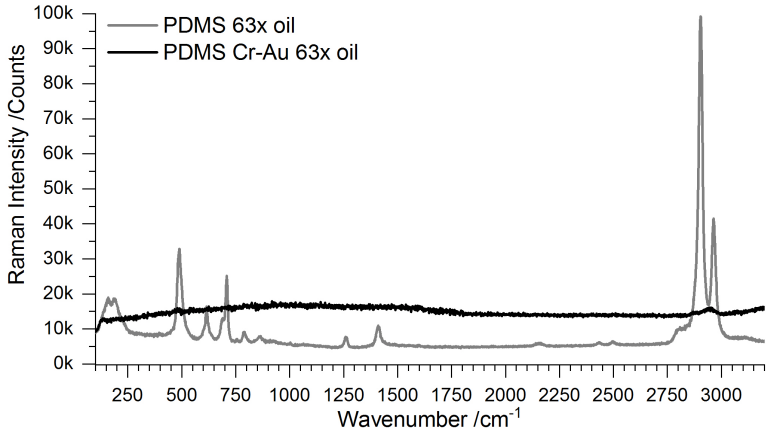


Figure 6.7: Raman spectrum of PDMS with and without a Cr/Au evaporated layer excited with the 532 nm laser. The Au coated PDMS showed an even better reduction in all the main PDMS bands, with very small contribution in the stronger bands wavenumbers.

causing water to evaporate (see figure 6.8C), or due to water electrolysis due to high currents produced by the laser beam in the thin metal layer.

The device showed functional traps, and HL60 cells suspended in DPBS[1] were trapped. When tested with cells, even very low laser intensities ($\approx 11\ mJ$) caused morphological changes in the cells, presumably due to overheating (see figure 6.8E1 and E2), with 0.22 mW found to be a non-damaging laser intensity (see figure 6.8E3 and E4). Signal was very noisy and had a strong baseline, causing problems of switching between the two windows of the step scan – see figure 6.9. The use of low intensities to avoid damaging the cells made the spectra very noisy.

In general, evaporation of a metal layer removed the background of the PDMS, but caused increased baseline and noise and increased cell thermal damage due to acquisition with the laser. As the intention is to measure live single-cells in a non-destructive manner, this is thus not a viable background reduction mechanism. However, for other purposes, this could be a useful technique for Raman *on chip* microfluidic devices production.

6.2.2 Gold Nanocomposites for Background Reduction

As an alternative to metal evaporation, gold nanocomposites were considered as a background reduction method. A nanocomposite consists of metal nanoparticles embedded in a dielectric matrix.[308] Gold nanocomposites act as very good light absorbers, with almost 100% absorbance from UV[2] to near-IR[3]. The main difference with metal evaporation is that metal increases the reflection considerably, which can increase the Raman signal intensity but can also reduce resolution due to side reflection contributions in the sample. A perfect absorber layer – ie., a completely black substrate –, would eliminate reflections but could show overheating of the black layer.

For this experiment, a brief collaboration was established with Prof. Elbahri's group at the University of Kiel, motivated by their publication in Advanced Materials.[309] They sent a sample to test the Raman signal and see if **i)** the nanocomposite had a strong signal in itself and **ii)** if the high absorbance reduced the background contributions from behind

[1] Dulbecco's Phosphate-Buffered Saline
[2] Ultraviolet
[3] Infrared

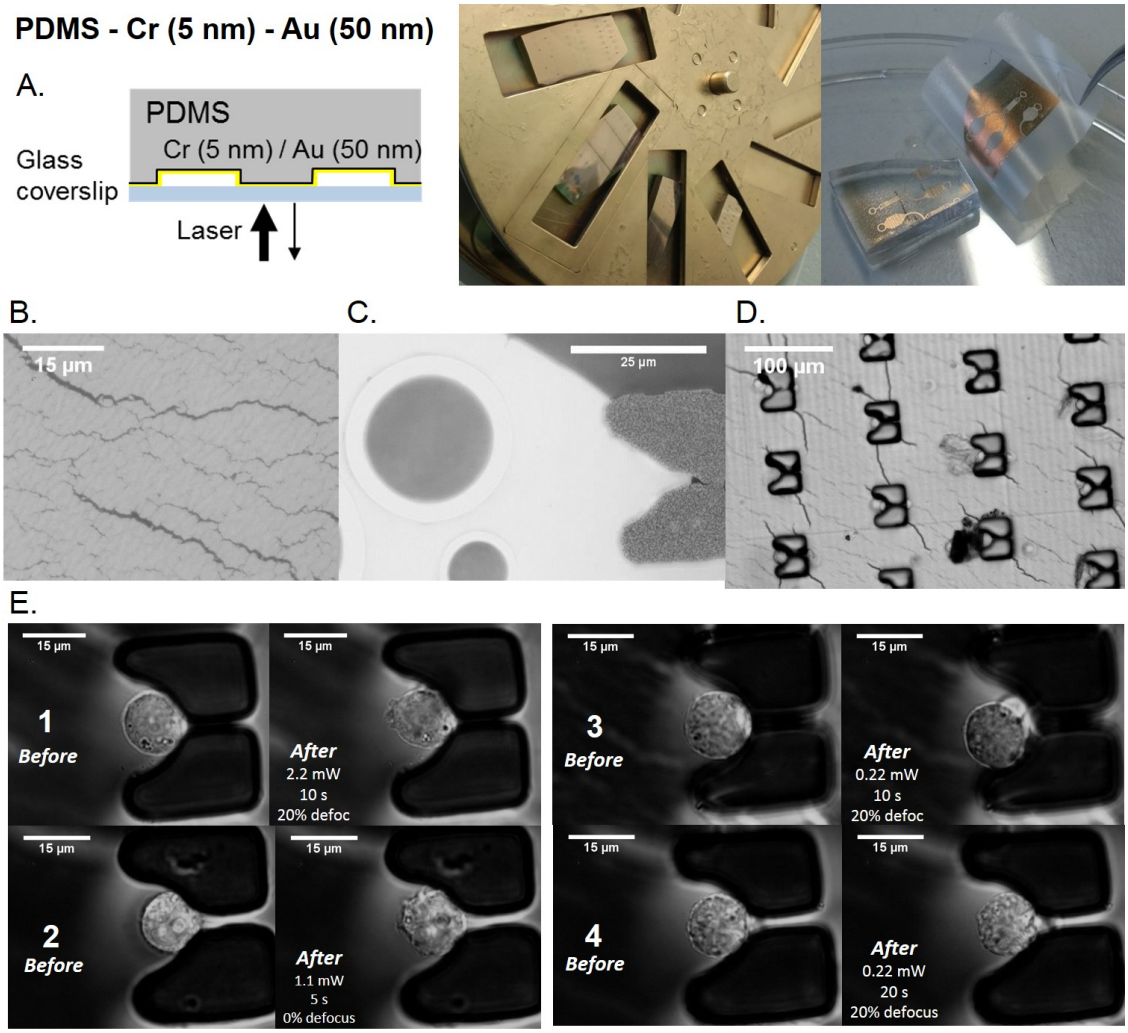


Figure 6.8: A. Schematics of the microfluidic device, where a layer of chromium (5 nm) and a layer of gold (50 nm) were evaporated onto the PDMS, that was then cleaned around the devices and bonded to a glass coverslip. Picture of the PDMS after the evaporation process. An alternative way of cleaning the areas that bond the glass by using tape. **B.** Close image of the layer of gold, showing some larger cracks than the ones found in the aluminium device. The layer had been exposed to the focused 22 mW laser spot and showed no signs of damage. **C.** Overheating in the trapping area caused water to evaporate, generating bubbles of water vapour. **D.** HL60 cells trapped in the device. **E.** Effects of the visible laser onto the trapped HL60 cells (before/after). Intensities higher than 2.2 mW caused the cells to bleb, likely due to overheating. Reducing the laser focus had strong effects on the cells even when lowering the intensity. When using 0.22 mW, the cells showed no visible damage.

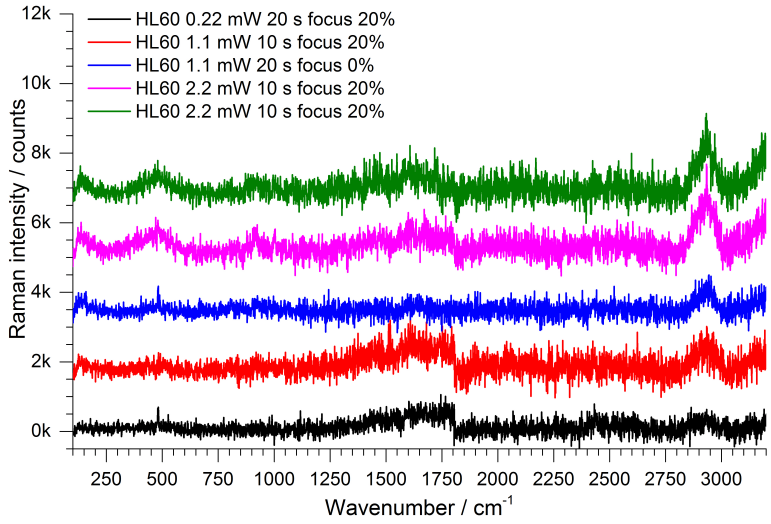


Figure 6.9: Trapped HL60 Raman signal from the cells in figure 6.8 when measuring with different settings after baseline subtraction. Due to the strong baseline on the raw data, the Renishaw software did bad stitching of the acquisition windows. The data is, in general, weak and very noisy.

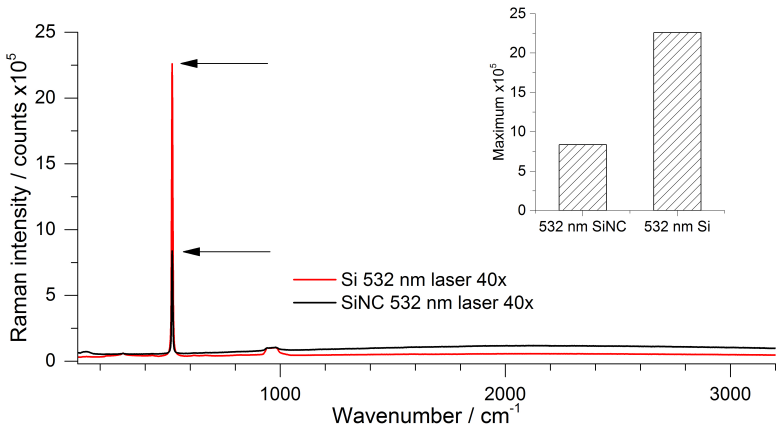


Figure 6.10: Silicon peak intensity reduction when measuring through a gold nanocomposite (ex. 532 nm). The measurements were done in the same sample on a region with and without nanocomposite. The data showed a reduction in the silicon peak intensity, with a final intensity of 37% the initial intensity.

the nanocomposite. For this, Dr Abdelaziz (from Elbari's lab) prepared samples of the perfect absorber nanocomposite in silicon for measurements of the attenuation of the silicon background peak. The sample was 20nm (SiO_2-Au 25%) on 70 nm SiO_2 spacer layer on a silicon substrate. Figure 6.10 shows the 532 nm laser results for the silicon sample. Results showed a small increase in the baseline that was overall flat, but only a small reduction in the silicon peak contribution, that was 37% as intense as the silicon sample peak. The sample was also measured with the 785 nm laser (data not shown), showing a similar background but with lower reduction, likely due to better penetration of the light in the sample and lower absorbance of the nanocomposite in the near-IR. The reduction on the background was not considered high enough for further experiments, although it is possible that a thicker layer of nanocomposite would result in a complete elimination of the background.

6.2.3 Glass SU8 Devices for Minimization of Background

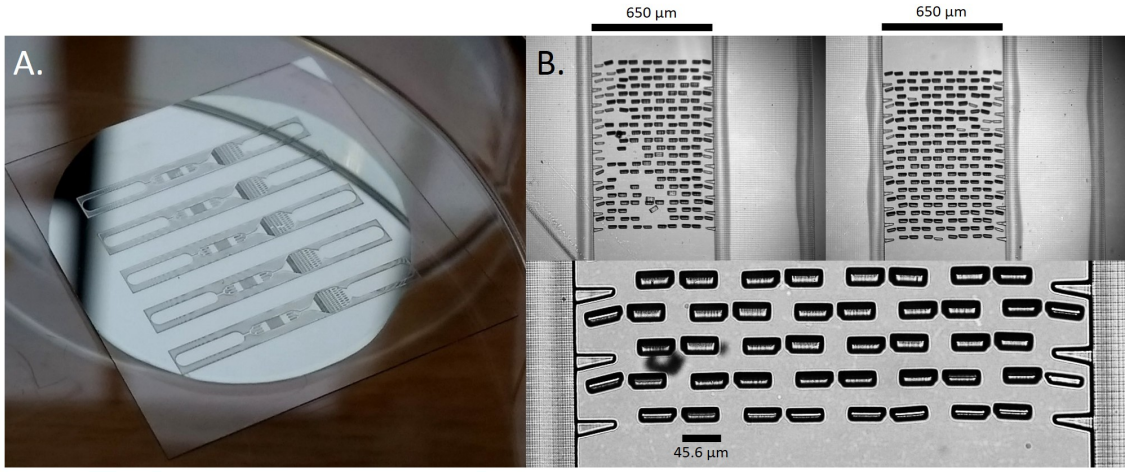


Figure 6.11: Glass-SU8 device fabrication results. **A.** Picture of the coverslip with SU8 after developing. **B.** Microscope images of the trapping region, showing most of the traps tilted due to bad adherence and low stability due to the high aspect ratio.

A further alternative to PDMS is the fabrication of glass-SU8 devices, where the glass coverslip and slide form the top and bottom of the device and the SU8 constitutes the inner structures of the device (walls). Fabrication was done following the SU8 fabrication parameters specified in Methods (see page 77) but onto a glass coverslip instead of a silicon wafer – usually showing lower SU8 adherence. When fabricating traps for PDMS moulding, the SU8 needs to be the negative of the final PDMS structure obtained, that is, the traps are fabricated as holes onto the bulk SU8 structure, so they result as pillars in the PDMS device. However, for a glass-SU8 device, the traps needed to be fabricated as the negative of the usual pattern, that is, as high aspect ratio pillars. The results for the fabrication are shown in figure 6.11. Figure 6.11A shows a picture of the developed devices, where the SU8 formed the walls around the device and the traps inside the device. Figure 6.11B shows the microscope images of the device, where developing caused most of the trapping pillars to fall. This was likely due to the high aspect ratio of the traps and the lower adherence of

SU8 on glass. The bonding to another coverslip was also unsuccessful, and thus this line of fabrication was abandoned.

6.2.4 Confocality Reduction of PDMS Background with Channel Height

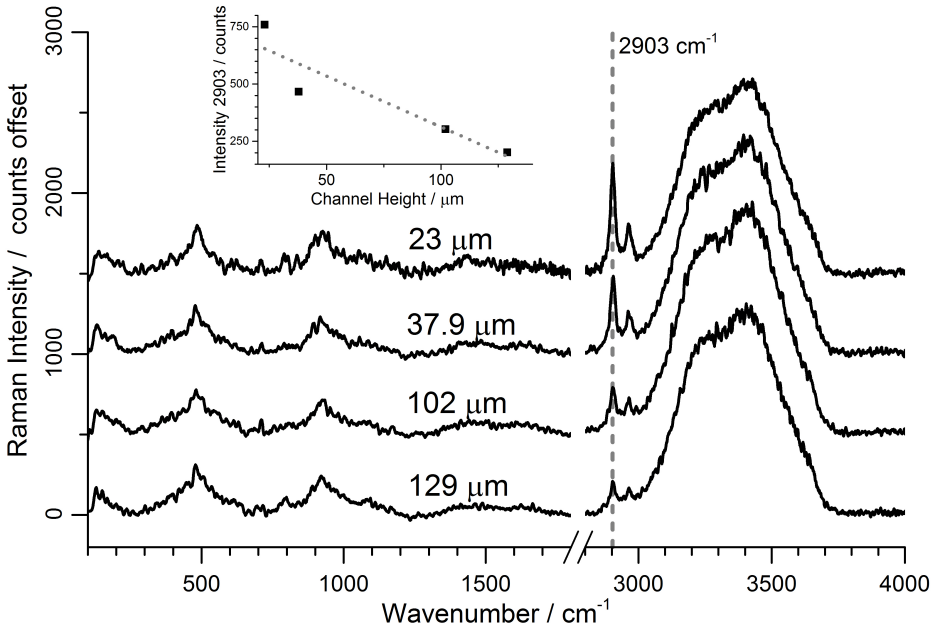


Figure 6.12: PDMS background with channel height when focussed on the inner coverslip part of the microfluidic device. The inset plot shows the 2003 cm^{-1} band maximum intensity dependence with the channel height, fitted with a linear regression.

PDMS devices are usually around 0.5-1 cm thick to ensure that the tubing used to interface the device to microfluidic pumps can be fitted in the inlets easily. However, this means that the amount of PDMS behind the sample is usually hundreds of times wider than the channel height. For this reason, even though the system is confocal and the signal from the background should not contribute strongly to the measurement, the strong PDMS background is always present in the spectrum, especially when measuring small volumes such as mammalian single cells or bacteria. To measure the contribution of PDMS background near the coverslip and how it relates with the sample thickness, devices of different thicknesses (23, 37.9, 102 and 129 μm) were prepared using SU8 2025 and SU8 2075 (Step profiling done by Liam Hunter in the School of Engineering). Measuring with the 100× oil immersion objective and *high* confocality mode next to the inner side of the coverslip, the spectrum obtained is shown in figure 6.12. The maximum intensity at 2903 cm^{-1} after removing the water band baseline was plotted in an inset plot with respect to the channel height and fitted to a linear regression. Data showed that there is a clear dependence of the PDMS background contribution with the height of the channel.

To test the PDMS background contribution relative to the measured single-cell spectrum, HL60 cells were measured in the deeper device (129 μm) at stop flow, near the coverslip inner side and using the 100× oil immersion objective – see figure 6.13. The measurement was taken with 1 s acquisition time and 5 accumulations but without defocusing the laser to increase the signal to noise ratio. Data showed that increasing the channel height significantly

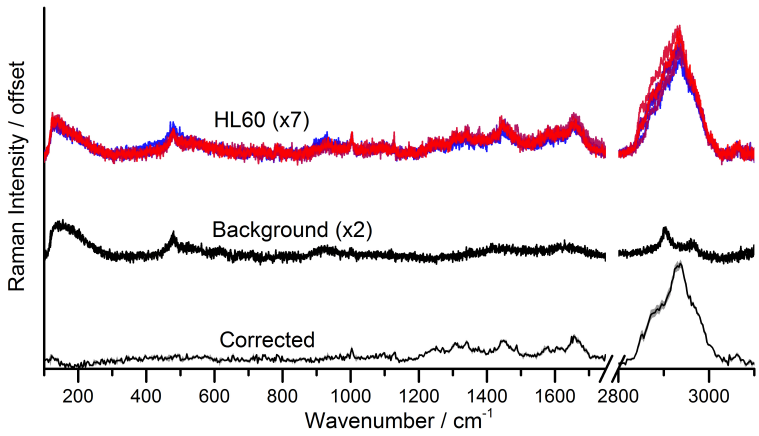


Figure 6.13: Single cell HL60 *on chip* data measured through the coverslip on the 129 μm channel. Data shows the signal on 7 different cells and 2 backgrounds, where the average of each was taken and used to subtract the background from the cell data, obtaining the corrected spectrum. The area around the spectrum corresponds to the propagated error using the standard error of each average.

improved the cell signal quality, and thus when the microfluidic design allows it, the channel should be chosen to be as deep as possible.

As a proof of concept, the signal of a single cell when flowing through the channel and crossing the acquisition laser was acquired. Signal was acquired for 0.5 s continuously on the CH stretching region using the 532 nm laser, at a constant flow of 0.5 μL/min. The acquisition region was limited to the CH stretching area to speed up acquisition while observing the strongest cell band. The laser spot was not defocused to increase the signal intensity. Figure 6.14 shows a single event acquired when an HL60 cell crosses the laser scanning point, changing the shape of the spectrum in the CH stretching region from the two PDMS bands to a broader cell band overlapped with PDMS signal. The data plotted in this figure was smoothed to reduce the noise. Even though the cell signal could be acquired, it was extremely noisy and unlikely of any diagnostic use. The experiment did, however, show the feasibility of measuring in flow using RS[1], without trapping. However, the flow rates required would have to be very low to allow sufficient time for data acquisition (0.5-1 s per cell). To ensure a sufficient number of events could be captured, the microfluidics would need redesigning. One option would be to use a flow focusing device followed by a concentrating mechanism. An alternative design would be to use a technique to align the cells such as dielectrophoresis or surface acoustic waves. The use of non-linear Raman would help to obtain a stronger signal using significantly shorter acquisition times, increasing the throughput. This is discussed further in chapter 7 (page 192).

6.2.5 Cell Raman Signal in a Microfluidic Channel

To assess the quality of the signal of a cell, a trapped HEK293 cell in a 25 μm deep PDMS device was mapped using RS (532 nm laser, high confocality). The 63× oil objective was used with a step of 0.7 μm and a mapping size of 82×57 px^2, with acquisition times of 0.7 s per pixel, taking a total time of 1 h for the whole map (4674 spectra). The map was analysed using the Wire software, with cosmic ray subtraction, baseline subtraction and vector

[1] Raman spectroscopy

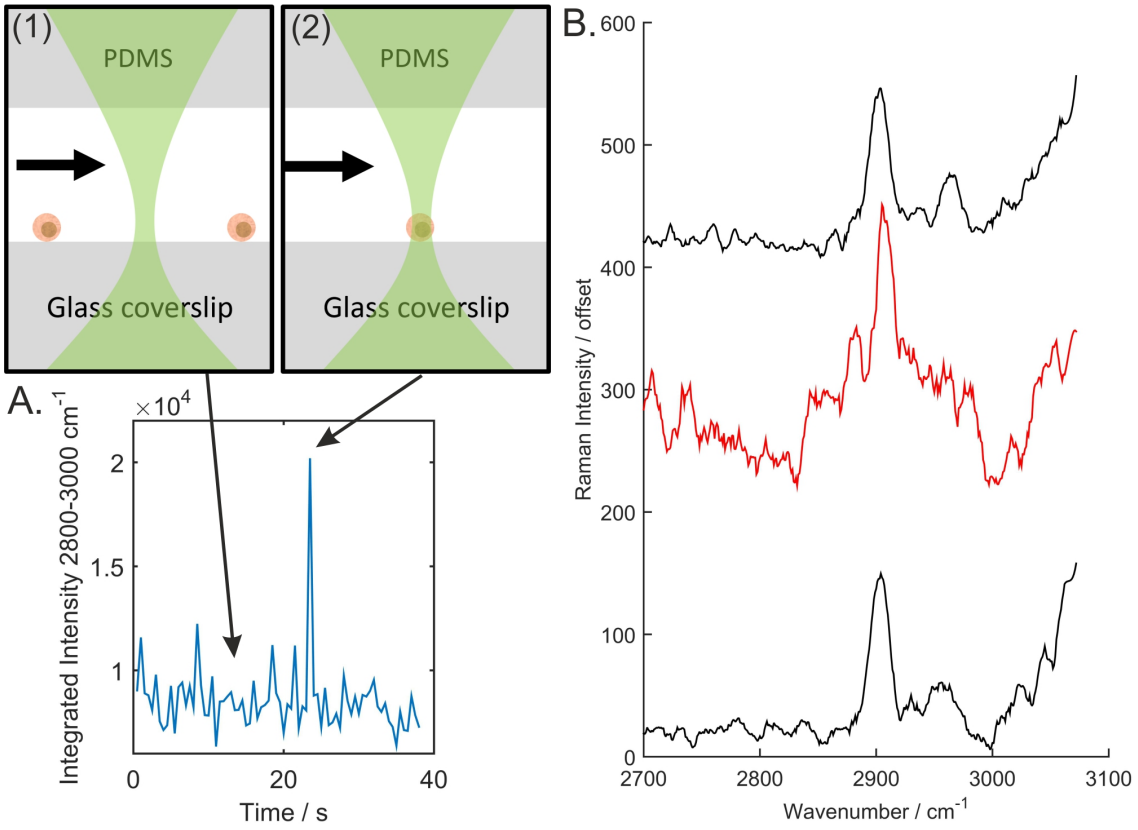


Figure 6.14: Single event of an HL60 cell crossing the laser scanning point when measuring at 0.5 μL/min with acquisition times of 0.5 s per spectrum on the CH stretching region, when exciting with the 532 nm laser. **A.** shows the time series detection of the event and **B.** shows the Raman spectra CH-stretching region of the event (red) compared with before and after (black.

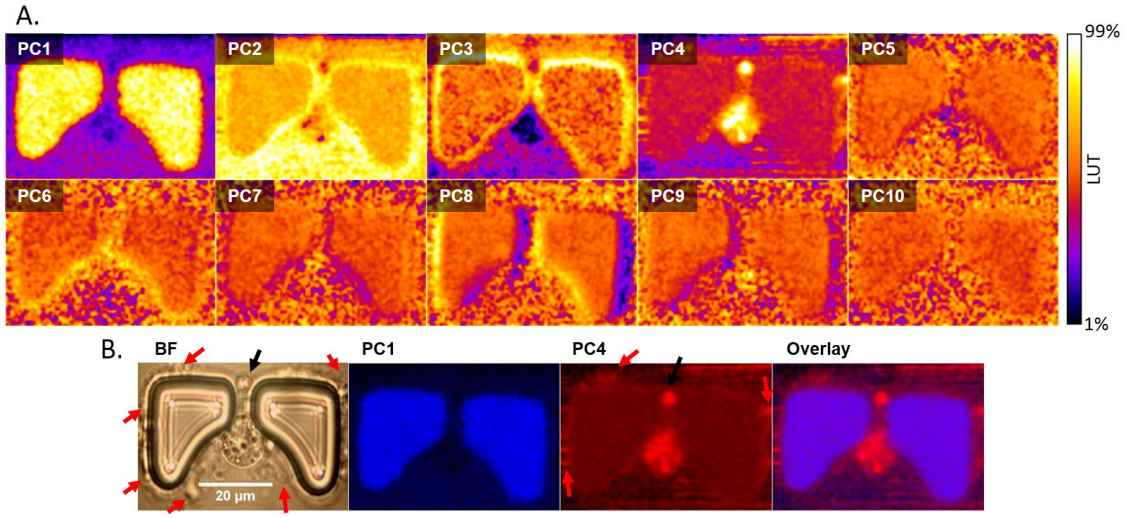


Figure 6.15: A. PCA scores false-colour images for the Raman mapping of a trapped HEK293 cell in a PDMS microfluidic device, where the intensity was linearly adjusted with the score so 1% of pixels were saturated in either side of the LUT. **B.** Bright field image of the mapped area and false colour images of the scores of the coefficients for the bulk of the trap (PC1, in blue) and the cell signal (PC4, in red). Some of the cell debris around the trap are indicated with red arrows, and the black arrow indicates the cell protuberance deformed through the trap.

normalization. It was then plotted using Matlab (see figure 6.15A) with a LUT[1] where 1% of the pixels were saturated at either side of the score distribution, and the map was Gaussian smoothed with a radius of half a pixel size. PC1 explained 43% of the variability, with PC2 only explaining 4% and the first 10 coefficients explaining 55% of the variability. From the first 10 coefficients, only one of them showed a distribution with strong contributions in the cell area (PC4, 2% of the variability), emphasizing how strong the PDMS contribution was compared to the cell signal. The HEK293 cell often deformed through the trap channel, showing a small protuberance on the other side of the trap. When imaging PC1 and PC4 in blue and red (figure 6.15B), is is shown that Raman detects the contributions of these deformed structures in addition to the main cell body.

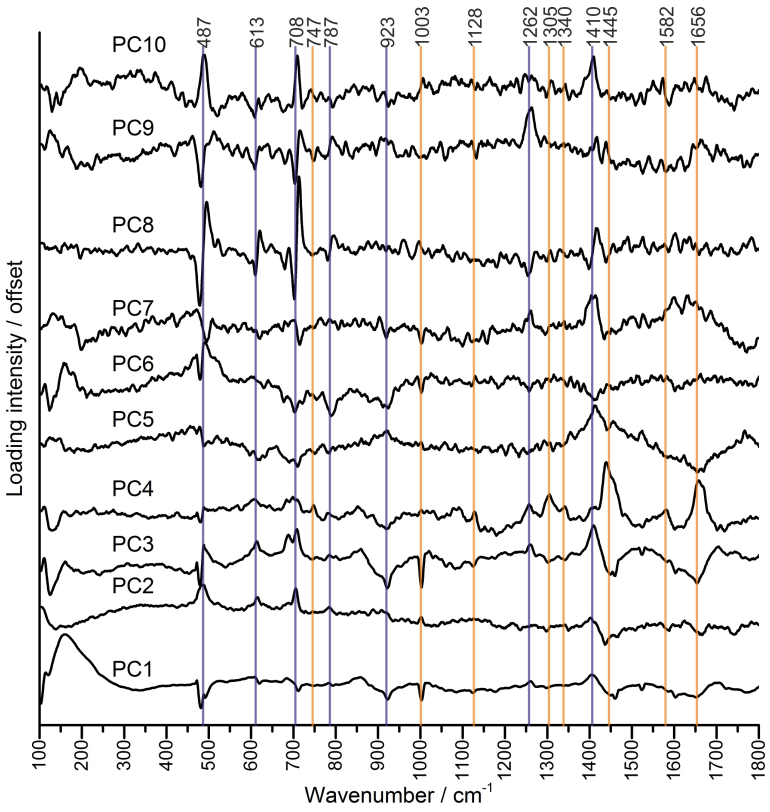


Figure 6.16: First 10 coefficients for PCA analysis of the map from figure 6.15, where the main bands for PDMS and cell have been labelled in blue and orange respectively.

Figure 6.16 shows the loadings for PCs 1 to 10. PC4 showed clearly cell contributions, with the broad bands at around 1455 cm^{-1} (lipids) and 1657 cm^{-1} (Amide I) and cyt C[2] contributions (747, 1128 and 1582 cm^{-1}). This is evidence that measuring cells on a PDMS microfluidic chip is possible. An interesting result of this map was the changes on the PDMS background along the surface of the trap. For instance, PC1, that showed high scores in the trap area, showed mainly glass contribution, with a broad contribution below 250 cm^{-1} and with negative PDMS contributions and cell contributions. This indicated that, when measuring in the trap area, a higher contribution of the glass is present in the spectra. PC2 however, showed higher scores in the areas around the trap, showing the constant PDMS background contributions. The PDMS contributions were variable, both in the intensity of the Raman signal and in the position of the bands. For instance, PC8 showed PDMS bands apparently shifted to higher wavenumbers, showing higher contribution in the left-hand side of the trap structure compared to the right-hand side of the trap, likely due to

[1] Look-up table
[2] Cytochrome C

increased scattering caused by the water-PDMS interface. In general, the data showed that the PDMS background contribution is not constant and both its intensity and shape vary along the microfluidic channel, and thus its subtraction is not straightforward, particularly for the sharpest bands.

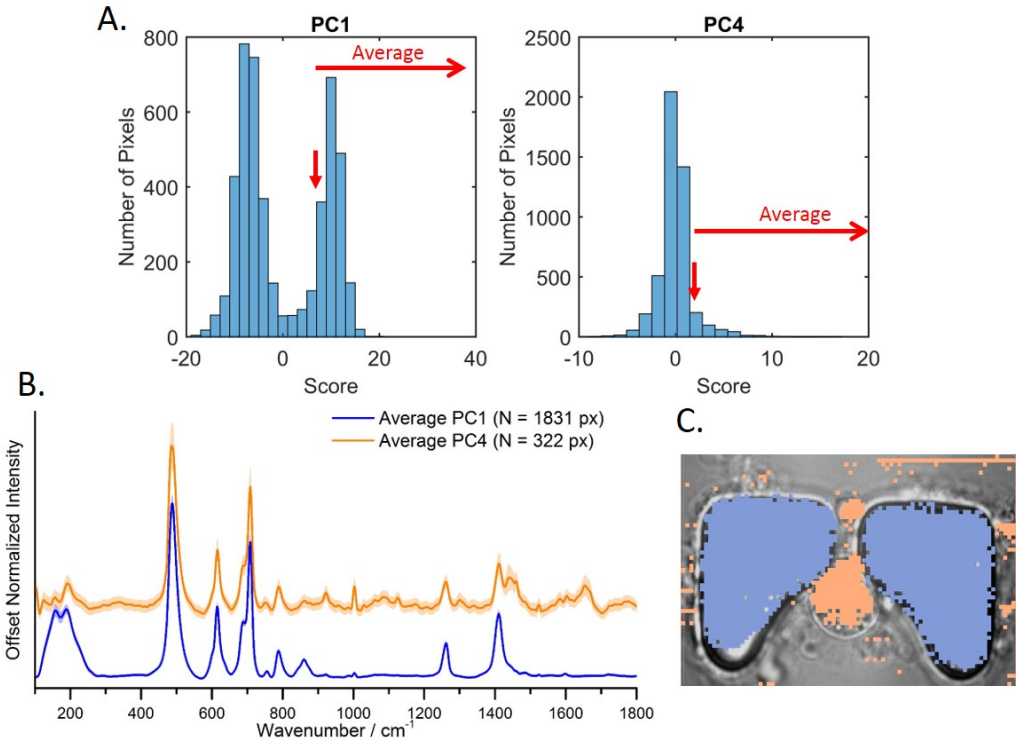


Figure 6.17: A. Histograms for the scores for PC1 (trap) and PC4 (cell signal) from the map from figure 6.15. The red arrows show the chosen threshold, where pixels with a score higher than that threshold were used for averaging. **B.** Average signal for the trap (blue) and for the cell (orange) for the pixel showed in blue or orange in C, where the fain broad line around the average represents one standard deviation. **C.** Bright field image with the overlay of the pixels included in each of the averages.

The mapping data from figure 6.15 was filtered according to the score of PC1 and PC4 to obtain the average signal of the trap and the cell. The chosen thresholds are indicated in figure 6.17A with a red arrow, corresponding to a value of 7.5 and of 4 for PC1 and PC4 respectively. The pixels whose score was higher than the threshold were used to calculate the average, shown in figure 6.17B, showing the contribution of the PDMS background to the cell spectrum. The selected pixels are shown as a mask on top of the bright field image in figure 6.17C for the trap (blue) and for the cell (orange). The cell spectra's strongest band in the fingerprint region, usually the Amide I, showed an intensity around 7 times weaker than the strongest PDMS band in this region. Even though some parts of the cell spectrum were still visible in the presence of the PDMS signal, most of the information below 900 cm^{-1} overlapped with strong PDMS bands.

Increasing the acquisition time but reducing the scanned area increased the quality of the obtained maps. Another example of a trapped cell is shown in figure 6.18, where live HCT116 cells were trapped in a microfluidic device in DPBS and the 100× oil immersion objective was used to map on the cell (ex. 532 nm), with an acquisition time of 1 s per pixel and high confocality. This map took 9 min 25 s acquisition time – compared to the previous map that

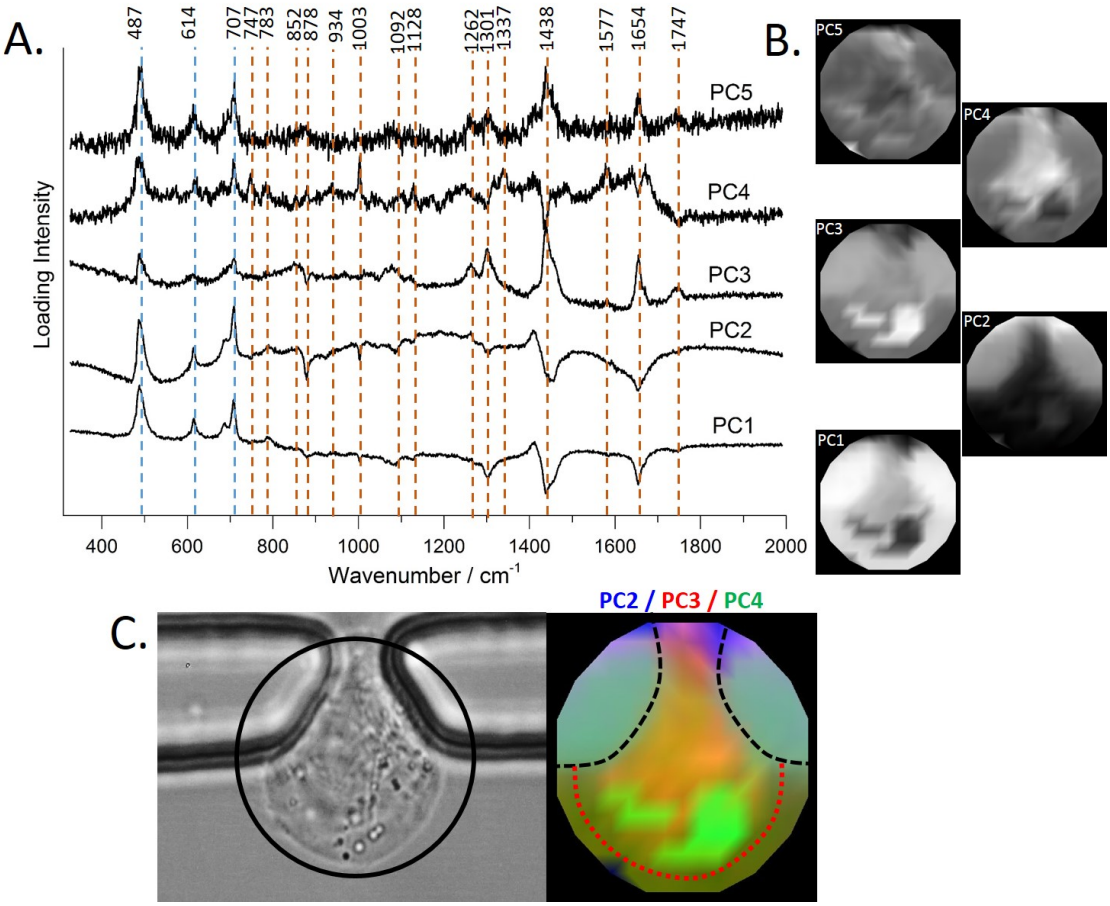


Figure 6.18: Map for a trapped HCT116 cell. **A.** Loadings for the first 5 PCs for the map, where the main peaks were labelled in blue (PDMS) and orange (cell). **B.** The triangulated maps for each of the coefficients is shown next to them, with a grey LUT expanded for all scores for each of the PCs. **C.** As PC2 (blue) was identified as the trap and PC3 (green) and PC4 (red) were identified as coefficients for the cell, a false-colour image was made with them, showing it next to the bright field image. The main outlines of the trap and the cell were drawn as a visual aid.

took 1 h, and even though it has lower resolution, the signal quality allowed resolving within cell heterogeneities. The radius of the map was of 15.12 μm and the step was 1 μm with a total of 177 spectra. Figure 6.18A shows the loadings for the map analysis. All coefficients showed the strong PDMS contributions up to 707 cm^{-1}, with changes in shape and baseline. PC3, PC4 and PC5 were identified as the coefficients with the most cell signal, where PC3 and PC4 showed higher spatial correlations in the score maps, whereas PC5 was noisier. PC4 presented the most typical features of a cell average spectrum, with the Phe band at 1003 cm^{-1}, the cyt C bands at 747 and 1128, the Amide III band at around 1240 cm^{-1} and a broader contribution where the Amide I band should be, with a band at 1337 cm^{-1}. However, the 1445 cm^{-1} lipids band was absent but strongly present for PC3 and PC5, likely detecting areas of lipid droplets. PC3 showed strong contributions at 1262, 1301, 1438, 1654 and 1747 cm^{-1}, in agreement with previous reports of lipid droplets signals.[310, 311]

In conclusion, the cell signal was strong enough to obtain high-quality spectra in a microfluidic channel and mapping with subcellular resolution. Mapping of a cell in a microfluidic trap was possible when using a highly confocal Raman system. However, the PDMS contribution impeded the use of the regions of the spectra below $\approx$ 710 cm^{-1} with additional weaker contributions at 1262 and 1410 cm^{-1}, and a robust background subtraction method would be desirable.

6.3 Least Square Fitting for on-chip Spectra Background Subtraction

Given that no suitable fabrication alternative was found to reduce the PDMS[1] background, a robust and reliable method for background subtraction was searched. On a simplified version of the work done by Ibrahim *et al.* for formalin contribution subtraction using least square fitting [312], a least square fitting with a PDMS background acquired at a similar height in the microfluidic channel was done. At least 5 backgrounds acquired at the same Z position as the cells of interest were acquired per experiment, and the average was used to create fitting function by fitting a spline to it. The spline was then fitted to the data using the following code in Matlab:

```
bb=spline(Wavenumber,Background); % Fitting spline to background
f=@(a,x) a*ppval(bb,x); % Creating function
% Fitting Options:
f0=fitoptions('Weights',BackgroundAverageZero.^4,'Exclude',Xintvalues>2000);
% Fitting background to data:
IntPeak=fit(Wavenumber.',RamanSpectraBaseline.',f,f0);
% Background subtraction:
RamanSpectraBackgSubs=RamanSpectraBaseline-(Background.')*IntPeak.a;
```

where the wavenumbers vector is `Wavenumber`, the background average spectrum after interpolation and baseline subtraction is `Background`, the fitting factor is `a` (saved inside the structure `IntPeak`), the uncorrected Raman spectrum after interpolation and gentle

[1] Polydimethylsiloxane

baseline subtraction is `BackgroundAverageZero` and the corrected Raman spectrum is `RamanSpectraBackgroundSubs`. It was found that the fitting was better when weighing the fitting with the background spectrum, making the regions with PDMS bands the ones that needed the best fitting, and thus removing weight from regions with strong cell features such as the Amide I. The weight chosen was to a power of 4, although weighting to a power of 6 also gave good fitting results. The CH stretching region, as it has such strong cell signal, was eliminated from the fitting by excluding wavenumbers >2000 cm^{-1}.

An example of an SW480 *on-chip* cell spectrum corrected using this method is shown in figure 6.19. PDMS bands were not visible after subtraction and the CH stretching region showed a similar shape to the one observed when measuring *off-chip*.

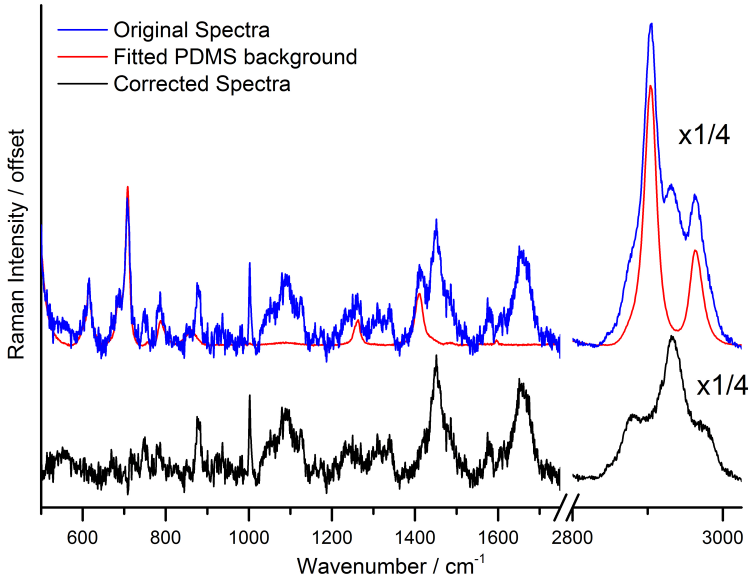


Figure 6.19: Example of PDMS background fitting and subtraction onto the spectrum of a SW480 live single cell measured trapped *on-chip*, exciting with the 532 nm laser.

6.4 Raman on-chip Results

6.4.1 On-chip Raman for Tracking Accumulation of DC473 on Live Single SW480 Cells

Following experiments with DC473 from Chapter 5, DC473 was seen to cause changes in cells following incubation, such as decrease of the lipids/Amide I signal. It was also shown that UV[1] irradiation of cells with DC473 showed a chemical reaction of most of the DC473. To complement the understanding of the DC473 effect in single cells, the ability to measure on the same single cell before and after incubation with DC473 would be highly desirable. As a proof of concept, the uptake of DC473 by SW480 cells was measured *on-chip*, where the cells were trapped, washed, measured, incubated with DC473 in media, washed and measured.

For this experiment, to ease cell tracking over time and assignment, Takeuchi-style traps were made, and cells were trapped with high efficiency, with fewer events of cells deforming

[1] Ultraviolet

through the trapping channel, but lower trapping density in general – see figure 6.20. More on the design and testing of these traps can be found in Section 3.8.5 (page 83). As Raman acquisition is slow and cell viability in buffer limited the total acquisition time to about an hour, only a few trapped cells were required, and hence this device was sufficient for such experiments.

The height of the device was 25 μm and the 12-15 μm cells were trapped in a PDMS trap of 12 μm diameter and 3 μm channel width. Microfluidic preparation is described in subsection 3.8.2 (page 81), and the trap design in subsection 3.8.4 (page 81) for Takeuchi-like cell traps. Cells were seeded in wells 1 day before the experiment at a concentration of 3.5×10^5 cells/well in 6 well plates and detached as usual (see subsection 3.6.1, page 74), to a final concentration $\approx4\times10^6$ cells/mL in cell dissociation buffer. The experiment was performed in a temperature and gas controlled chamber (37°C,20% O_2, 5% CO_2).

Cells were trapped using a flow rate of 10 μL/min until trap occupancy was close to 100%, and all cells were measured along the same line of traps. The cells were washed with DPBS[1] (10 min) and incubated in 4 μM DC473 in complete media for 20 min, followed by DPBS wash (10 min). Raman was acquired from the cells in DPBS before and after incubation (ex. 532 nm), during a period of around 26 min (*before*) and 25 min (*after*), under low flow (1 μL/min) to minimize any heating caused by the laser on the cells. When finishing the DC473 incubation, cells were washed with DPBS and confocal imaged, with a total washing time of about 46 min to ensure all the non-uptaken DC473 was removed. The beam was defocused by 50% to cover the whole cell and part of the trap, using the 100× immersion objective (22mW, 20s/cell, 532 nm laser). Spectra from cells was acquired from 100-3200 cm^{-1}. PDMS spectra taken from a trapping region was used for least-squares fitting and background subtraction. Spectra were normalized to the Amide I peak and baseline subtracted as usual.

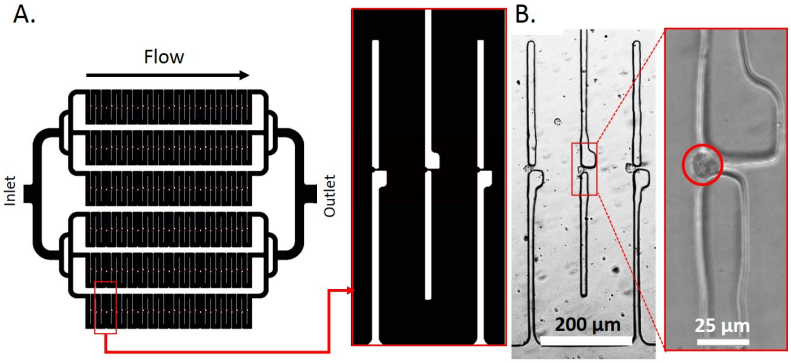


Figure 6.20: **A.** Design of the microfluidic device, with multiple traps in parallel. **B.** SW480 cells trapped in the device, with a zoom in a trapped cell. The red circle indicates the Raman acquisition region.

Some of the main issues found during the experiment were that sometimes cells would deform through the trap channel during buffer changes. Bright field pictures were taken from the cells before and after the incubation, omitting those cells that were washed away. Additionally, some cells showed a more than 50% intensity drop in the cell signal between the *before* and the *after* spectra and were also discarded. From 30 cells originally, 22 cells were kept for further analysis. The spectra post-incubation sometimes showed a fluorescent baseline due to the tail of the fluorescent emission of DC473. This was subtracted before analysis. The

[1] Dulbecco's Phosphate-Buffered Saline

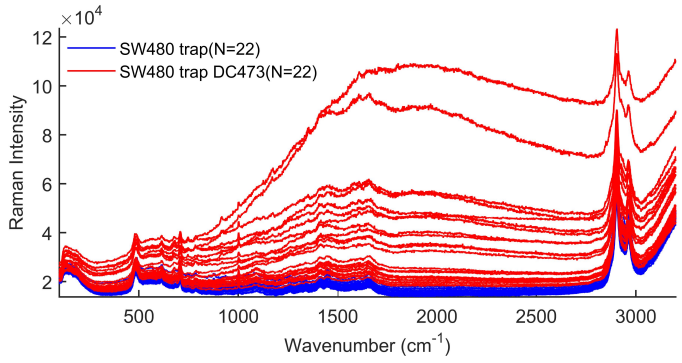


Figure 6.21: Raw spectra before and after incubation with DC473 *on-chip* exciting with the 532 nm laser, where spectra after incubation with DC473 showed a strong fluorescent baseline and all spectra showed strong PDMS background contribution.

raw data for this experiment is shown in figure 6.21. Confocal fluorescence images of the cells after incubation were taken to validate DC473 uptake with an additional technique. Images are shown in figure 6.22.

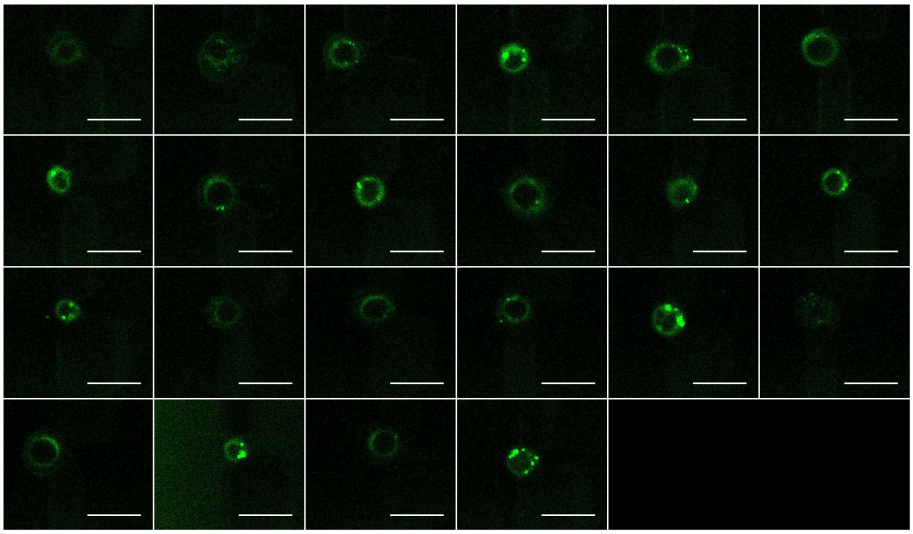


Figure 6.22: Confocal fluorescence images of the on-chip trapped cells after incubation with DC473, where the scale is of 25 μm. LUT[1] was adjusted equally in all images to increase contrast for printing purposes. Cell indexes for each row correspond to [1-6],[7-12],[13-18],[19-22].

The average of the spectra after background subtraction and baseline correction is shown in figure 6.23. As the data collected corresponds to the same cells before and after measurement the difference data for each cell could be calculated. Here, the difference spectrum was calculated as the average of the individual difference spectra for each cell – and not as the difference of the averages of each group as it was done in previous chapters. Main bands that increased post-incubation are indicated by the black lines. The two main DC473 bands at 1606 and 2173 cm^{-1} showed a clear increase post incubation. DC473 uptake after only 20 min incubation was moderate compared to the 4 h incubation from previous results from Chapter 5. Some of the DC473 bands showed a shift in position respect to previous results, with the Phe 1595 cm^{-1} bands appearing at 1606 cm^{-1}, and the alkyne 2195 cm^{-1} band weakly showing at around 2173 cm^{-1}. The 1141 cm^{-1} Phe vibrations here was not present, but a strong uptake band appeared at 1170 cm^{-1} instead. The 674 cm^{-1} DMSO[2] band showed a strong contribution at 677 cm^{-1}, indicating that even though the cells were washed with DPBS, DMSO still showed detectable levels in the cells. A strong band at 878 cm^{-1}

[2] Dimethyl sulfoxide

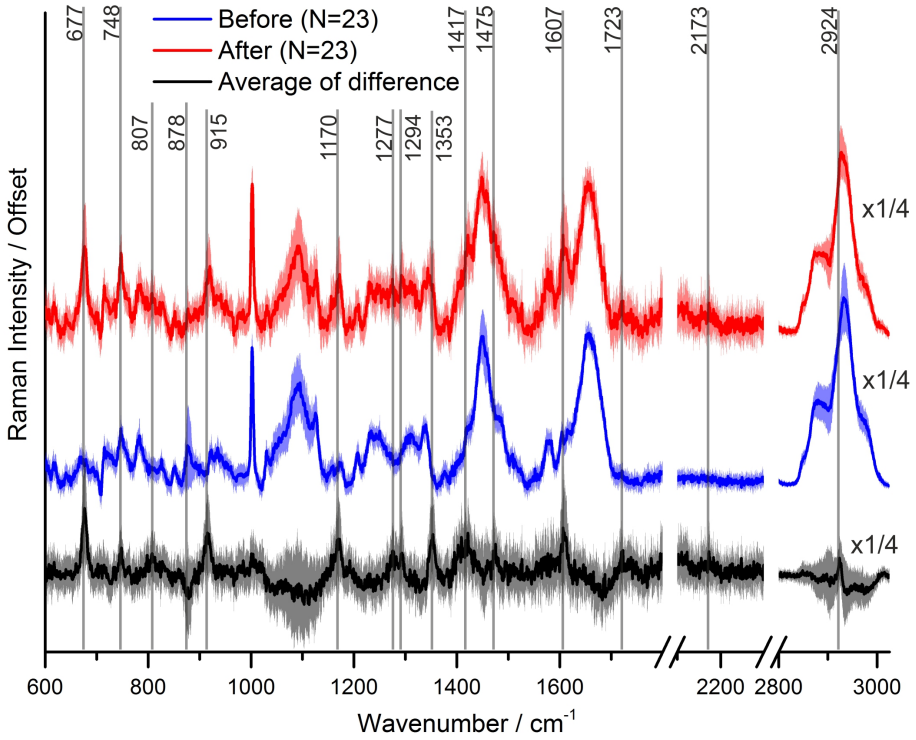


Figure 6.23: Average Raman spectra before and after incubation with DC473 *on-chip* and the average of the difference spectra for each cell (*after* minus *before*), excited with the 532 nm laser. The broad line around the average represents one standard deviation. Those bands that showed an increase in the difference spectrum were labelled as black. The CH stretching region is shown multiplied by a factor of 1/4 to increase the visibility of the fingerprint region.

showed a decrease following incubation. The boxplots and beeswarm plots of the integrated intensities of bands in the single cell difference spectra are shown in figure 6.25, with the significance level calculated using a paired t-test between the *before* and *after* single-cell data.

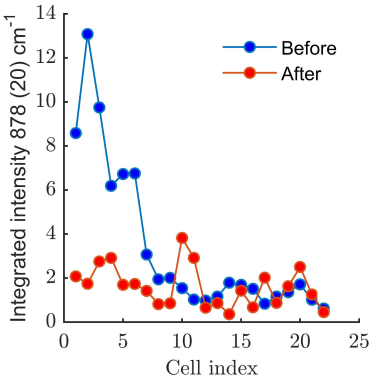


Figure 6.24: Plot of the integrated Raman intensity of the band at 878 cm^{-1} with the cell index, showing that only the first measurements showed high intensity. This band was tentatively assigned as remaining ethanol after washing, used to decrease the contact angle of PDMS. Raman excitation was done with the 532 nm laser.

Other bands that showed an increase following incubation were the 748 and 1126 cm^{-1} cyt C[1] bands, perhaps indicating an increase in viability following incubation with DC473 in media. A broad contribution was present at around 807 cm^{-1} in the difference spectra, in the region of vibrations of phosphodiester bonds and phosphate stretching, maybe indicating an increase in RNA or DNA.

The band at 878 cm^{-1} showed a stronger contribution in the spectra before incubation.

[1] Cytochrome C

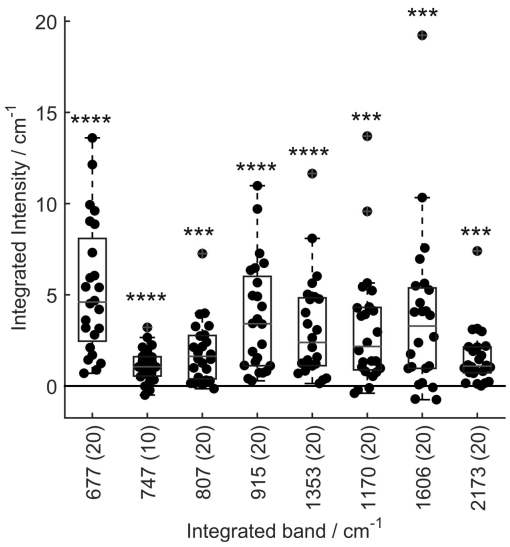


Figure 6.25: Boxplots and beeswarm plots for integrated intensities of selected bands from the Raman difference spectrum (ex. 532 nm) of each cell (after – before), with indicated significance (paired t-test). Each peak and their half width in wavenumbers are shown as *peak (halfwidth)* in the abscissa.

To identify the source of the 878 cm^{-1} band, its dependence with cell index was plotted in figure 6.24. Interestingly, only the first 6 cells measured before incubation with DC473 showed contributions of this band, with a clear trend over cell index – and thus over time. This could be indicating that it is due to a chemical that was being washed away over time with the DPBS flow after trapping. This band has been tentatively assigned to contributions of the remaining ethanol coating the PDMS[1] walls after the device washing steps,[313] used to reduce the contact angle of PDMS walls [314] and help remove air bubbles.

Another strong increase was for the band at 915 cm^{-1}, likely due to a glucose peak due to incubation with media, or due to ribose or proline/hydroxyproline contributions. The difference spectrum showed a sharp increase at 1353 cm^{-1}, as a result of a broadening of the 1338 cm^{-1} band, in a region of mixed contributions of DNA, proteins and CH_2/CH_3 vibrations. Another band that broadened in the DC473 sample was the lipid band at 1445 cm^{-1}, with increased intensity at 1417 and 1475 cm^{-1}. The CH-stretching region showed an anomalous shape caused by the subtraction of the PDMS in this region, that is more challenging than the fingerprint region due to the overlap of the sharp PDMS bands.

The PDMS bands at 485 cm^{-1} and in the CH stretching region showed worse background subtraction, with remaining contributions of PDMS, and thus for PCA, only the data from 500-2250 cm^{-1} was considered. PCA gave variance explained of the first 5 coefficients of around 43%, with variance explained of the first coefficient of 22%. The first 5 PCA scores and loadings are plotted in figure 6.26A and B respectively.

PC1 showed a very similar shape to the difference spectrum, with strong bands at 677, 915, 1170, 1277, 1353, 1413 and 1607 cm^{-1}, and weaker bands at 748, 1475, 1723 cm^{-1}. In the alkyne region, two weak broad bands could be seen at 2122 and 2173 cm^{-1}, however, these could be caused by artefacts in the fluorescence baseline subtraction. This coefficient showed low scores for the cells before incubation with DC473, and two populations for the cells after incubation with DC473, with around half the cells showing low uptake, with partial overlap with the scores of the cells before incubation, and the other half showing high uptake.

[1] Polydimethylsiloxane

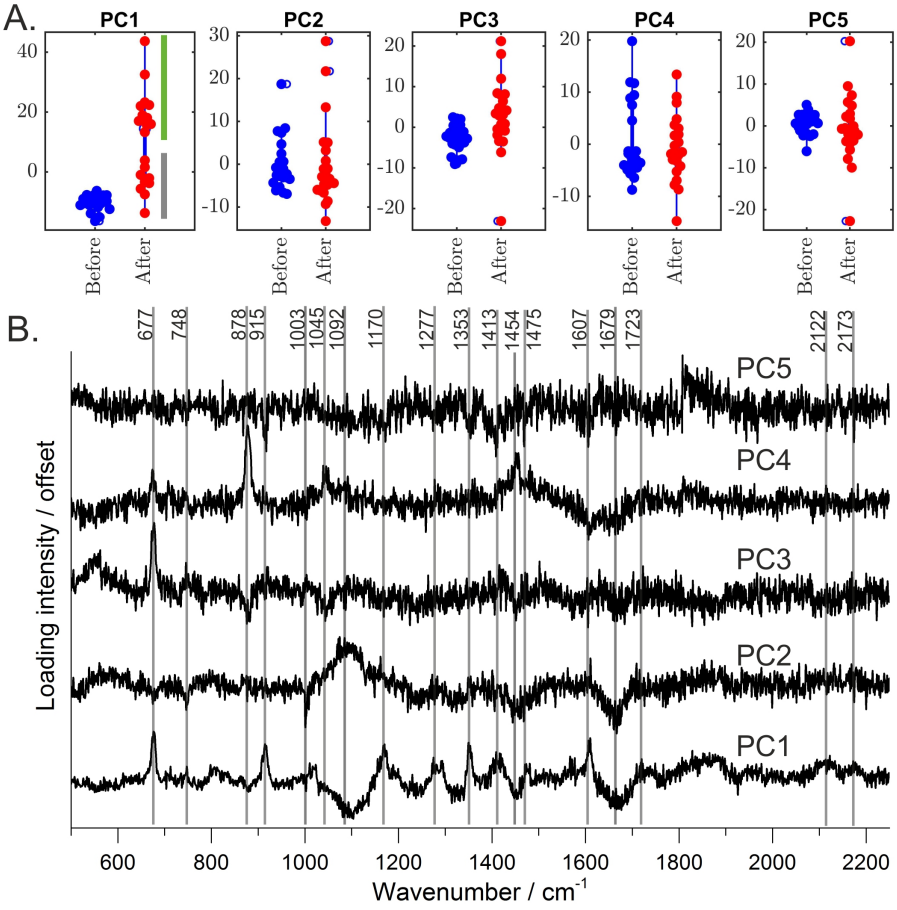


Figure 6.26: First 5 PCA beeswarms and boxplots of the scores (**A.**) and first 5 PCA loadings (**B.**) for the Raman data shown in figure 6.23 (ex. 532 nm laser), where the main peak positions have been labelled. In the PC1 scores in A., the low uptakers and high uptakers have been labelled in grey and green respectively with a band next to the beeswarm plot.

PC2 showed mainly broad positive and negative bands, probably due to artefacts in baseline subtraction caused by the sample fluorescence, with a sharp negative Phe peak at 1003 cm^{-1}, and broad negative lipids (1455 cm^{-1}) and Amide I bands (1656 cm^{-1}), showing in general within sample heterogeneity with overlap between the scores of the *before* and *after* samples.

PC3 showed a strong band at 677 cm^{-1} (DMSO) with in general higher scores for the *after* measurements and a tight population around score zero in the *before* measurements. Interestingly, the presence of DMSO in the cells did not show two populations, indicating that some cells accumulated more drug even though they all had a similar presence of DMSO.

PC4 showed a sharp band at 878 cm^{-1}, tentatively assigned to ethanol – seen at 886 cm^{-1} by Burikov *et al.* and assigned to C – C vibrations –, with contributions in the first few cells from the *before* spectra, with additional bands at 1092 cm^{-1} and at 1454 cm^{-1} – assigned to C – O ν (1056) and CH_3 and CH_2 bending (1456) of ethanol respectively.[313] PC5 barely showed any distinguishable features, with very noisy contributions and more within cell variability in the *after* samples than in the *before*, both centred around score zero.

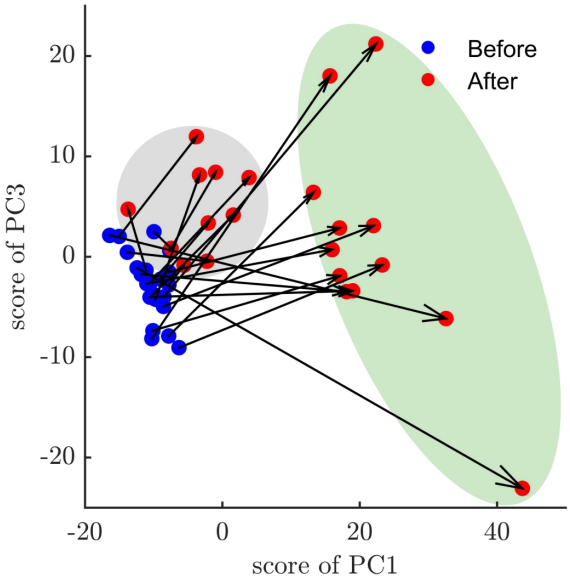


Figure 6.27: PC1 vs PC3 scores scatter plot for the Raman *on chip* SW480 results before and after incubation with DC473 (ex. 532 nm), where each cell *before* and *after* scores were joined with an arrow. The low uptakers and high uptakers groups have been labelled in grey and green respectively.

PC1 and PC3 were thus chosen as DC473 related. A scatter plot with PC1 and 3 is shown in figure 6.27, where each cell *before* and *after* scores are joined with an arrow. The scatter plot shows clearly two distinct behaviours of the cells, with a group of high uptakers (green area) and a group with undetectable uptake or very moderate uptake (grey area). The reason why some cells showed significantly higher uptake than others was unclear, although it could be due to differences in the viability or cell cycle stage of the cells.

A very interesting question is if the amount of DC473 uptaken per cell can be predicted from the before dataset, that is, if there is a Raman biomarker for drug uptake. For this, the datasets were divided in two depending on the scores on PC1, with high uptakers (score > 8, 12 cells) and low uptakers (score < 8, 10 cells). A PCA/LDA model (2-fold validation) was trained with the high/low uptakers data, and the scores and coefficients for the classification are plotted in figure 6.28. The scores on figure 6.28A showed that low and high uptakers were almost perfectly separated by the LD with the *before* observations. The loading from

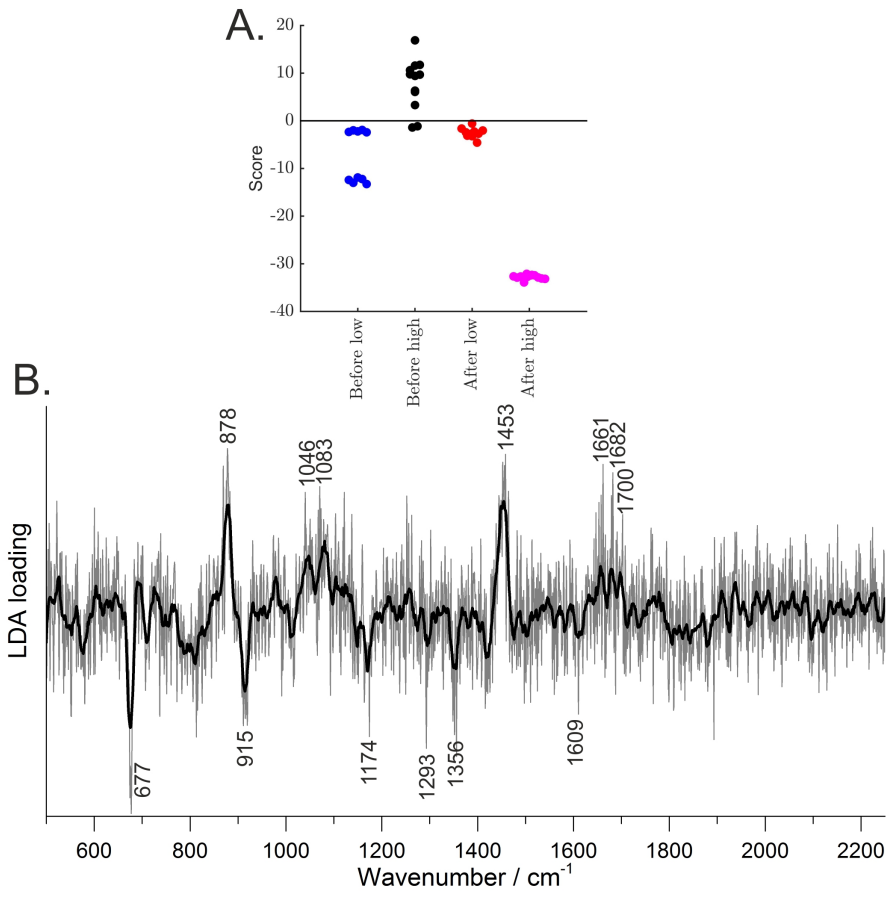


Figure 6.28: LDA analysis for the prediction of low and high SW480 uptakers for the DC473 *on chip* Raman data (ex. 532 nm). **A.** Scores for the LD of best separation between low and high uptakers based on the spectra from *before*. **B.** Loading for the LD of best separation, where the smoothed loading has been plotted in black overlapping as a visual aid and the main bands have been labelled.

figure 6.28B showed in general similar negative bands as the positive bands on PC1, but with strong positive bands at 878 and 1453 cm^{-1} and a weaker band at 1046 cm^{-1} tentatively assigned to ethanol [313]*star*. This could be indicating that the remaining ethanol in the walls of the trapping area of these cells is increasing their membrane permeability, increasing their DC473 uptake. Other bands were present at 1083 and on the side of the Amide I at 1661, 1682 and 1700 cm^{-1}, that could be indicating changes in proteins.

The experimental design, in this case, allowed the measurement of changes on the same single cell before and after incubation, but has a limitation, as the design did not allow incubation with media of a control set of cells in parallel. The changes seen could have been caused by the media incubation and the shear stress caused by the flow on the cells, and a control experiment where the cells are incubated with media without DC473 is necessary. Additionally, the number of cells analysed is low (N=22), and bigger datasets will be required to confirm these results.

The main issues found in this experiment were the changes between buffers, that sometimes caused air bubbles to flow in the device or caused backflow and cell untrapping. However, for easier and more effective switch between solutions and to ensure that sterility is kept during the whole experiment, a matrix fluid changer such as the MUX system from Elveflow could be used in future experiments. In addition to this, cell signal acquisition in media instead of buffer should be considered to minimize the stress on the cells, subtracting the Raman signal of the media. Future experiments should aim to activate the photodynamic activity of DC473 *on-chip* following incubation.

6.4.2 Other on chip Raman Projects

During the course of the PhD, Raman *on chip* was also applied to different systems, such as microbubble trapping, measuring and analysis *on chip* (Dr. Churchman's thesis, section 6.7)[315], single *Escherichia coli* bacteria analysis or Raman for oil permeation tracking over time onto a PDMS device (to be published by Hunter *et al.*). This subsection presents a taste over the *E. coli* project, whereas the other projects are discussed in depth elsewhere.

6.4.2.1 Raman on Single Escherichia coli on-chip

Within bacterial subpopulations, when treating them with antibiotics, at least three different possible phenotypes can occur: susceptible bacteria, killed by the antibiotic, persister bacteria, that continue proliferating shortly after the antibiotic is removed, and VBNC[1] bacteria, that can remain non-growing for prolonged periods of time after the antibiotic is removed. These bacteria can elude traditional microbiology assays, and thus require characterization via single-cell resolution assays. These single cells are often associated with dormancy and are associated with the relapse of diseases such as tuberculosis, candidiasis or cholera among others. They present in *E. coli* cultures with a fraction between 10-0.1%, and have shown mild activity in the *tolC* promoter, associated with multidrug efflux pump AcrAB-TolC, tracked

[1] viable but non-culturable (bacteria)

with a *tolC-GFP*[1] construct strain. Previous results showed differential gene transcription of persister and VBNC cells even before antibiotic treatment, with lower fluorescence in the tolC GFP reporter strains for both persister and VBNC cells.[316]

This is yet another example of cells with the same genetic background showing distinct phenotypes. As a collaboration with Dr Stephano Pagliara's lab at the University of Exeter, single *E. coli* bacteria were measured *on-chip*, with the final aim of observing if VBNC bacteria showed a distinct biochemical signature compared to persister and susceptible bacteria. If the measuring settings were non-toxic for the bacteria, this would allow measuring before and after antibiotic treatment, observing if there is a distinct biochemical phenotype of persister and VBNC cells before drug treatment.

As a first experimental approach, single-cells were measured for both a WT[2] *E. coli* strain and a GFP-*tolC* reporter *E. coli* strain in a 10 μm depth PDMS channel. Samples were prepared in Exeter and posted to Leeds, and refrigerated before the measurements. Samples were washed with fresh minimal LB media to remove most of the bacteria before measuring. Single bacteria measurements were acquired using 100% intensity of the visible 532 nm excitation laser (22 mW) with the 100× oil objective and *high* confocality. Each measurement was exposed for 20 s and acquired using the Andor Detector. The fingerprint region was acquired (100 – 1800 cm^{-1}) with a laser spot size of $\approx$ 1 μm. For both samples, multiple backgrounds were taken from the same Z-position on a region with no visible bacteria and were used together to correct all the data from both samples. For each GFP-tolC bacteria, confocal imaging was done before the acquisition (ex.488 nm /em.493-650 nm), and non-fluorescent cells were discarded. Cell fluorescence was then bleached before each Raman measurement.

Figure 6.29A shows the raw data for single bacteria for each of the two strains. Even though the signal was acquired for 20 s, the biological features of the samples are dominated by the glass and PDMS background. GFP samples were bleached before each measurement, but some of the samples still showed a certain fluorescent baseline. Figure 6.29B shows the pre-processed data, where cosmic rays were removed, the background was fitted and subtracted, data were aligned to the silicon peak and normalized to the Amide I. The region below 690 cm^{-1} was removed due to variable contributions of glass and PDMS background. Spectra were very noisy, but showed the main features of single cell spectra, including the Phe peak at around 1003 cm^{-1}, Amide III band around 1250 cm^{-1}, lipids band at 1451 cm^{-1} and Amide I band at 1656 cm^{-1}, showing that single *E. coli* bacteria could be measured *on chip*. This can be seen better in the average spectra shown in figure 6.30, showing the average of each of the strains with the main bands labelled. The spectral shape agrees with previous reports of bulk Raman spectroscopy of *E. coli off chip*.[317]

Results showed the feasibility of measuring single bacteria on PDMS microfluidic devices. Further experiments will aim to optimize the microfluidics for single cell RS[3] and to increase the sample sizes to 50-100 bacteria and ensuring reproducibility, with the final aim to test the bacterial resistance to antibiotics. Figure 6.31 shows the design constructed by our

[1] Green Fluorescent Protein
[2] Wild type
[3] Raman spectroscopy

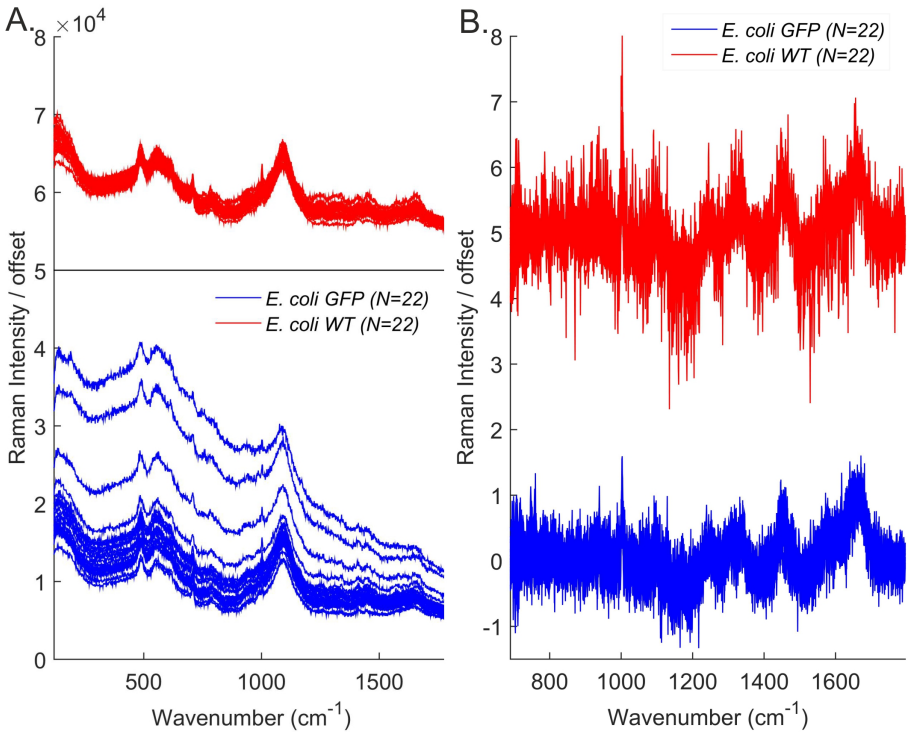


Figure 6.29: **A.** Raw Raman data for single *E. coli* measured at the bottom of a 10 μm channel (ex. 532 nm) for the tolC-GFP sample (GFP) and the WT sample. The GFP sample was bleached before the Raman acquisition, but some of the measurements showed a considerable fluorescent baseline. **B.** Pre-processed data for the dataset from A., where the data below 690 cm^{-1} was truncated. Data was offset for clarity.

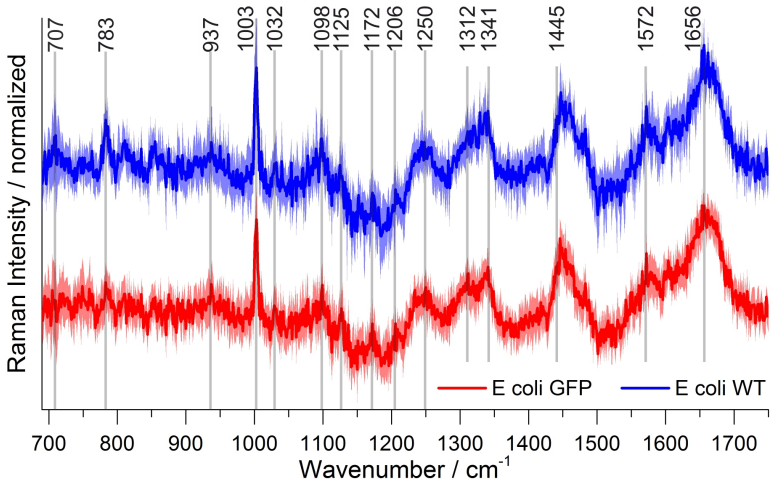


Figure 6.30: Average of *on chip* Raman spectra (ex. 532 nm) of single *E. coli*, where the broad line around the curves represents one standard deviation, and the spectra were offset for clarity. Main bands were labelled.

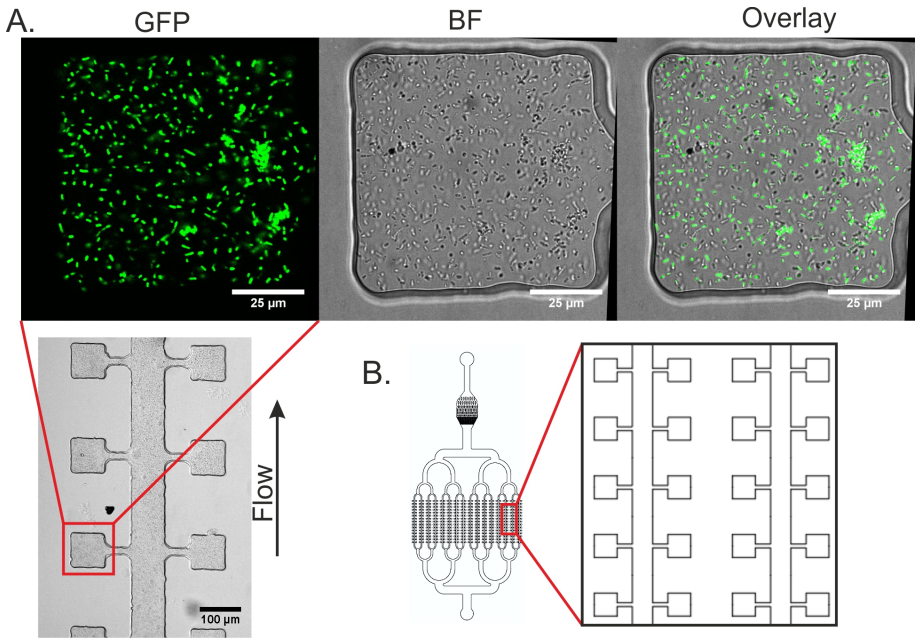


Figure 6.31: Design for a device for measuring bacteria *on chip* using Raman spectroscopy. **A.** Confocal images of the GFP-*E. coli* on a chambers microfluidic device, where the depth is of 10 μm. **B.** Design of a microfluidic device based on a device sent by our collaborators in Exeter.

collaborators in Exeter and tested and imaged in Leeds. Bacteria can be seeded in the lateral chambers with a diffusion of media from the central channels without disturbing the bacteria. Treatment of the surface can ensure that bacteria remain adhered. Incubation with antibiotics and washing of the device can be done from the central channel. Additionally, a gradient generator could be added to the design, allowing treating the bacteria with a distribution of antibiotic concentrations.

6.4.2.2 Conclusions and Future Challenges

This chapter explored the compatibility of RS with PDMS microfluidic devices. Coating of the polymer with metals showed a strong reduction of the background, with complete reduction for the gold samples. However, the strong optical absorbance of the gold layer caused overheating when working with trapped cells and thus was not compatible with live single cell RS. Gold nanocomposites showed a promising shielding of the background of 63%. However, did not completely eliminate the background in the tested settings. Experiments with thicker layers of gold nanocomposites could completely eliminate the background without causing strong reflections. Increasing the system confocality and performing a PDMS fitting and subtraction proved to be effective for cell spectrum acquisition *on chip*. Proof of concept experiments with cells trapped *on chip* showed successful mapping. A proof of concept experiment with SW480 cells and DC473 uptake showed that microfluidic and live cell RS could obtain information from the changes induced by a drug in every single cell. It also showed that the DMSO band at 677 cm^{-1} – absent from the 4 h incubation spectra (chapter 5) – was present at the early stages during drug incubation. Additionally, obtaining single cell data from the same cell from before and after incubation allowed the search of predictors in the *before* spectra for the DC473 uptake, showing a correlation with the presence of bands

tentatively assigned to ethanol and increased uptake. The ethanol presence would have never been detected using a different *on chip* technique, showing how Raman *on chip* can give essential information to understand heterogeneous responses to drugs of cells subpopulations at the single cell level. However, further optimization of the microfluidic platform will be necessary to be able to perform this experiment in a reproducible manner and with increased throughput. To finish, early stages of single bacteria measurements *on chip* are shown, proving that it is feasible to obtain the single bacteria signal in a PDMS device, with further experiments to be done in this direction.

Part IV

Conclusions and Future Work

Chapter 7

Conclusions and Future Work

7.1 Chapter by Chapter Overview

Chapter 4 (page 91), as the first Results chapter, was this thesis's first approach to live single cell RS[1], with the analysis of hundreds of cells from different cell lines using the 532 nm visible laser: from leukaemia (Hl60), to colorectal carcinoma (HCT116) and colorectal adenocarcinoma cells from different stages, SW480 (Duke's B), HT29 (Duke's C primary) and SW620 (Duke's C metastasis). Correlation analysis proved to be a valuable tool for band assignment on single-cell spectra. It was also a first approach to the chemometric analysis of large Raman datasets, where we tested classification using PCA[2], PCA/LDA[3], DT[4]s and SVM[5]s, both considering good classification and ease to understand the output in relation to the underlying biology. For the metastatic model system, we found that when normalising to the Amide I peak, SW620 cells showed higher saccharides, phosphates, nucleic acid content, α-helix, β-sheet and $\alpha + \beta$ secondary structure, with increased ratio of α/β secondary structure and increased ratio of $CH_2 : CH_3$ stretching bands. The SW480 cells displayed a higher proportion of disordered structure and increased overall CH stretching intensity. PCA discrimination indicated that the cyt C[6] peaks accounted for most of the within sample variability whilst the protein, nucleic acids and lipid-associated peaks gave the largest variability between cell lines. LDA/PCA and SVM results yielded >98% correct rate in classification between the SW620/SW480 cell lines compared to DTs and C5.0 DTs, that gave a lower performance but had an output easy to understand, with single peak biomarkers. When considering PCA/LDA, HCT116 and SW620 cells showed the best classification, whereas HT29 cells showed the worst, with frequent misclassification of HT29 cells and SW480 cells. Overall, the PCA/LDA performance for the separation of different cancer types was 92.4±0.4%, with even higher classification rates for HCT116 cells and SW620 cells, showing the potential of Raman spectroscopy to separate between types of live cancerous cells.

[1] Raman spectroscopy
[2] Principal Component Analysis
[3] Linear Discriminant Analysis
[4] Discrimination Trees
[5] Support Vector Machines
[6] Cytochrome C

When comparing multiple CRC[1] cell lines we found that the primary CRC cell lines (SW480, HT29 and HCT116) showed increased $\alpha + \beta/\beta$-sheet ratio in the Amide III band compared to the HL60 and SW620 cells. The carcinoma cell line HCT116 showed lower cyt C, CH_2 symmetric stretching and free phosphates, and higher lactate contributions compared to the adenocarcinoma cell lines. The analysis of the average and PLSR[2] analysis with the colorectal adenocarcinoma stage showed that lactate (1725 cm^{-1}), phosphates (810-813 cm^{-1}) and double bonds (1556 cm^{-1}) increased with adenocarcinoma stage, whereas the contributions above 2900 cm^{-1}, DNA (787 cm^{-1}) and the Amide I contribution above 1675 cm^{-1} decreased with adenocarcinoma stage. The increase in phosphates correlated with the higher overall phosphorylated status of more advanced CRC stages, previously seen by Yousef *et al.* in SW480 and SW620 cells.[96] Our results confirmed an increase of the 1556 cm^{-1} double bonds band with metastasis previously shown by Tsikritsis *et al.* in fixed cells [97] even when adding HT29 cells to the model. PLSR also highlighted the changes in lipid composition between the different cell lines, with changes in the 1455 and the 2900 cm^{-1} lipids bands, correlating with the increase in lipid droplets of HT29 and SW620 when compared to SW480 cells. An increased amount of lipid droplets has previously been correlated with tumorigenic potential in CRC.[318]

Chapter 5 (page 111) aimed to detect accumulation of photodynamic compounds in live single cells from the adenocarcinoma model system: SW480, HT29 and SW620 cells. It initially explored the detection of the accumulation of PPIX[3] as the ending metabolite of 5-ALA[4]. PPIX showed red fluorescence when accumulated in the cells, and caused toxicity to UV[5] light. Even though PPIX showed RR[6] when excited with the visible laser, detection of PPIX was found to be challenging due to its similarity to cyt C, present in viable cells. Results from PCA analysis showed that, in subconfluent cultures, the main PC[7]s with PPIX contribution were PC5 and PC6, showing accumulation trend of HT29 > SW480 > SW620, with very significant difference between them ($p<0.0001$), in agreement with previous literature reports from traditional biochemical techniques, showing higher accumulation of PPIX in HT29 cells than SW480 cells.[153, 259]

As a novel photosensitiser, DC473 was studied as part of a collaboration with the University of Durham and LightOX, who provided the purified molecule. This molecule had a very strong Raman signal, including a sharp phenyl band at 1597 cm^{-1} and a broad alkyne band in the silent region, at 2197 cm^{-1}. We used confocal RS and confocal fluorescence to map its location within an SW480 cell using the near IR[8] laser. Raman results showed accumulation in lipid droplets, where fluorescence was strong, in the cytosol, with lower fluorescence intensity, and in an area with overlap with the nucleus, potentially the nucleolus, where fluorescence was absent, indicating that DC473 was in a polar environment for this subcellular location. Accumulation of DC473 caused a very significant drop in the CH stretching bands

[1] Colorectal Cancer
[2] Partial Least Squares Regression
[3] Protoporphyrin IX
[4] 5-aminolevulinic acid
[5] Ultraviolet
[6] Resonant Raman
[7] Principal coefficient
[8] Infrared

($p<0.0001$), that was hypothesised to be due to inference with lipid metabolism when accumulating in lipid droplets, and an increase in cyt C related bands, likely due to interference with mitochondria. Additionally, we obtained the DC473/Amide I ratio for each of the cell lines by doing single-cell RS, where both fitting the alkyne band and the phenyl band gave significantly higher accumulation for HT29 and SW620 cells compared to SW480 cells. These cells also showed wider alkyne peaks than SW480.

Additionally, RS was used to track apoptosis on these cells after UV light activation of the compound (PDT[1]). Even though HT29 and SW620 cells accumulated more DC473, SW480 and SW620 cells were found to show changes of morphology when exposed to UV light unlike HT29 cells, an observation confirmed with the PI/FDA live-dead assay. The excitation laser light and the UV light used in the experiment showed moderate effect in the Raman signal of the cells, with a drop of cyt C 747 cm^{-1} but no cell morphology changes. Activation of the compound showed a drop in its Raman intensity, but additionally the remaining signal showed changes in some bands positions, with a shift of both the phenyl band at 1597 towards 1601 cm^{-1} and the alkyne band at 2195 towards 2200 cm^{-1}, both following a sigmoid-like trend with an exponential τ value of around 0.02 s^{-1}, corresponding to a lifetime of 50 s. Similarly, the FWHM of the peaks also followed an exponential broadening, with a width increase of $\approx$ 4.9 and $\approx$ 3 cm^{-1} for the alkyne and the phenyl bands respectively, with a similar life-time. In terms of cell bands changes, the 1128 cm^{-1} cyt C band showed a decrease in intensity post-PDT, with additional drop of bands around 1340 and 1368 cm^{-1} (mixed assignments) and in the lipids fingerprint region (1419, 1442 and 1467 cm^{-1}). DC473 was also activated with UV light in DMSO[2] solution, showing a drop in the main DC473 peaks and indicating the chemical reaction of the molecule. To understand the mechanism of action of DC473 in the cell, reproducing it *off cell* is essential, and thus more will be done in this direction in the future, tracking the activation of DC473 in aqueous solution with RS.

Chapter 6 (page 161) explored the applicability of RS to PDMS[3] microfluidic systems. The PDMS background contribution inside the chamber was minimal when confocality is kept high and was reduced by increasing the height of the channel. Aluminium evaporation onto PDMS was shown to cause a considerable reduction of the PDMS background, with a broad baseline contribution, but showed laser damage of the aluminium layer when measuring with the 532 nm laser. Gold/Chromium evaporation created a photostable metal layer onto the PDMS with almost complete reduction of the PDMS background. However, caused overheating and thermal damage of the trapped cells even at low laser intensities. This method for background reduction could be used for other microfluidic applications where the viability of the cells is not a concern. Gold nanocomposites were also tested for their potential to reduce background being strong light absorbers. On silicon samples, they showed a 63% reduction in the background, that could be increased by increasing the nanocomposite layer thickness. These materials should be tested as Raman substrates more extensively, as they could decrease background without increasing reflection of the laser, that can cause decreased resolution when using bulk metal substrates.

[1] Photodynamic Therapy
[2] Dimethyl sulfoxide
[3] Polydimethylsiloxane

When increasing the confocality of the system and the height of the channel, low PDMS background contribution could be achieved, obtaining cell spectra with minimal PDMS contribution. In this context, a cell flowing in a high PDMS channel could be detected when constantly acquiring Raman spectra every 0.5 s, meaning the optimization of this system to obtain cells arranged in a line moving at a low enough speed would pave the way for the RACC[1] system. Using a more sensitive technique, such as non-linear Raman, would allow quicker acquisition of these spectra, as will be discussed in the next steps section.

Chapter 6 also showed that Raman mapping on cells trapped in a PDMS structure is possible, obtaining clear signals of the cell and the PDMS traps, with within-cell resolution. Using least squares fitting of the PDMS background allowed reproducible background subtraction for spectroscopy results, to allow the tracking of DC473 uptake *on chip*. The uptake results were made on a small number of SW480 cells (N=22), and were just a first attempt at live single-cell drug incubation experiments *on chip*. However, they showed interesting results regarding studies of DC473 uptake. Results showed that in early incubation stages with DC473 (20 min) cells showed the 677 cm^{-1} DMSO band, that is lost after further incubation with the compound. Obtaining the *before* and *after* spectra for the same cell allowed the observation of the uptake in the *after* spectra and looking for predictors in the *before* spectra. Results also showed the presence of remaining ethanol in the microfluidic device after the washing stages, that correlated with a higher uptake of DC473. Although these studies are at a preliminary stage, these assays do not require high volumes of cells or long incubation assays and allow testing non-proliferative cells. Other *on-chip* results showed the ability to record single *Escherichia coli* spectra in a PDMS device (10 μm deep) obtaining good quality spectra that reproduced the literature *E. coli off chip* reported bands locations.

7.2 A Wider Context: The MRC Project

This PhD was part of a joint project between the University of Leeds and Leeds Teaching Hospitals NHS Trust in order to develop a microfluidic platform to integrate different phenotyping techniques, first to be developed separately. In this context, another PhD student, Fern Armistead, worked on developing Deformation Cytometry as a mechano phenotyping assay, using similar cell lines. Some of her results on HL60 cells were published in *Biophysical Journal* (2019) [3], where she proved cell deformability by joining ultrafast imaging and microfluidic deformation in an extensional flow. Her first paper presented results for HL60 cells as a model circulatory cell as well as for the CRC[2] cell line, SW480. Her second paper – currently in production – studies the deformability of the three CRC model system, seeing mechanical changes associated with CRC progression. In this sense, our results can give an overview of the mechanical and chemical phenotype of CRC. The future aim is to make these techniques compatible and be able to obtain the mechanical and chemical phenotype of each single-cell in a joint microfluidic platform.

The deformation cytometry results are summarised in Table 7.1. In terms of the diameter, SW480 and HT29 cells have similar sizes, with HT29 being slightly smaller, whereas SW620

[1] Raman activated cell cytometer/cytometry
[2] Colorectal Cancer

cells are significantly smaller. In terms of the nuclear diameter, HL60 had the smallest nucleus, with a nuclear ratio of 55 ± 2 %. Of the adenocarcinoma cells, the cell line with highest nuclear ratio was SW480, with a nuclear ratio of 72 ± 1 %, with HT29 and SW620 showing similar nuclear ratios of 63 ± 1 % and 67 ± 6 % respectively. Interestingly, Raman results showed that the 782 and 810 cm^{-1} bands indicated DNA contents (normalised to Amide I) following a trend HL60>SW620>HT29>SW480, inverse to the nuclear ratio. This could be indicating that RS[1] is being sensitive to how compressed the nucleus is, obtaining higher DNA to Amide I content in cells with a more compressed nucleus (i.e. HL60), and lower ratio on cells whose nucleus is more expanded (i.e. SW480).

Table 7.1: Summary table of the mechanical phenotyping results from Fern Armistead's research on the adenocarcinoma model system and HL60, when using 5 $\mu L/min$ in an extensional flow, corresponding to a shear dominated regime (33 cP viscosity).

Cell line	HL60	SW480	HT29	SW620
Diameter (μm)	12.3 ± 0.2	15.2 ± 0.3	14.5 ± 0.2	11.5 ± 0.1
Nuclear diameter (μm)	8.9 ± 0.1	11.2 ± 0.1	11.0 ± 0.1	9.2 ± 0.1
Nuclear ratio (%)	55 ± 2	72 ± 1	63 ± 1	67 ± 6
Maximum strain	0.18 ± 0.01	0.08 ± 0.01	0.12 ± 0.01	0.11 ± 0.01
Final strain	0.030 ± 0.004	0.010 ± 0.003	0.056 ± 0.006	0.049 ± 0.006
Relaxation time (ms)	3.5 ± 0.2	1.36 ± 0.06	0.89 ± 0.05	1.04 ± 0.05

When looking at cell deformation, different parameters were available, such as the maximum strain, the final strain and the relaxation time of the cells. In terms of strain, HT29 and SW620 cells showed higher maximum strains and final strains than SW480 cells, suggesting that, in terms of their mechanical phenotype, HT29 cells (primary, Duke's C) showed a similar phenotype to SW620 cells (metastasis, Duke's C). This would be indicating that HT29 cells had already acquired a mechanical metastatic phenotype before leaving the primary site. However, when looking at the RS results, HT29 and SW480 cells showed the most similar Raman signatures, with more than 7 % of cell overlap respectively, whereas SW620 cells showed a very distinct signature, with a correct rate of 96.2 ± 0.7 %. This would be indicative of HT29 cells being in an intermediate state between SW480 cells and SW620 cells, with the mechanical phenotype of metastatic cells, but the chemical phenotype of primary cells.

7.3 Next Steps

Regarding the work on this thesis, multiple open questions are left. Chapter 4 studied the classification of different stages of colorectal cancer cell lines, but more cell lines should be studied to confirm this trend. Additionally, healthy cell lines should be studied to compare with the diseased stages. Furthermore, live cells isolated from patient's biopsies should be studied to translate the results from cell lines to patients. In terms of the analysis, expertise on single cell analysis was developed during the course of the PhD, and tools for single band fitting used on Chapters 5 and 6 could be used on the data from Chapter 4 for Amide I and III bands.

[1] Raman spectroscopy

For the work on Chapter 5, multiple open questions are left. The effectiveness of the compound DC473 as a PS[1] needs to be confirmed in animal models, opening the possibility of Raman mapping on biopsies to confirm accumulation of DC473 in different tissues. LightOx will be soon doing the first tests of DC473 in a murine system. When working on tissue, autofluorescence is often strong, and thus RS[2] will be useful for unequivocal detection of the compound. Raman will be used to detect the accumulation of the compound in various tissues and its presence in induced CRC[3] tumours after PDT[4], observing the cell changes induced by PDT in vivo and the residual compound after PDT.

Understanding of the mechanism of uptake of DC473 on cells is very limited, and little is known about its metabolism on live cells and its accumulation. Further experiments are needed to confirm that it accumulated preferentially in lipid droplets. Raman has high resolution and thus can be used to track the compound during uptake through endosomes, lysosomes or exosomes. Additionally, the compound could not be completely bleached neither in solution or within cells, indicating it is possible that it is aggregating. Little is known about the effects on the fluorescence spectra of aggregation, that needs further study. Furthermore, other compounds from the DC473 series could also be studied using RS, specially non-fluorescent compounds that are otherwise difficult to detect on live cells.

For the work on Chapter 6, RS on PDMS[5] devices is shown to be very challenging. More powerful methods for background subtraction such as EMSC[6] could help reduce the PDMS contribution in a more reproducible manner [312], allowing measuring in shallower channels. Additionally, the results obtained for background reduction using nanocomposites were promising, and further experiments should be done using a different layers of nanocomposites to optimise the background reduction as a new substrate for RS.

Regarding the Raman on bacteria work from Chapter 6, the ability of RS to detect changes in the biochemical composition of a cell, being sensitive to changes in viability and proliferative states, make it a valuable asset to study changes at the single cell level. This is especially beneficial in systems where bacteria are non-culturable, like in the case of VBNC[7]. In this context, further experiments should be done to test tolerance/persistence/susceptibility to antibiotics *on chip* at the single-cell level.

However, to realize fast Raman *on-a-chip*, non-linear RS should be done. Recently, Goda's group developed the dual-comb CARS[8] with the fastest to date RACC[9], with a proof of concept study of the astaxanthin production of *Haematococcus lacustris* at a rate of 2000 events per second.[205, 206] Their system is one step away from a RACS[10] system, holding spectacular potential for liquid biopsy single-cell applications. A benchtop RACC system in every hospital would mean fast cancer screening and patient single-cell drug response testing,

[1] Photosensitiser
[2] Raman spectroscopy
[3] Colorectal Cancer
[4] Photodynamic Therapy
[5] Polydimethylsiloxane
[6] Extended Multiplicative Scatter Correction
[7] viable but non-culturable (bacteria)
[8] Coherent anti-Stokes Raman spectroscopy
[9] Raman activated cell cytometer/cytometry
[10] Raman activated cell sorter

a step further towards personalised medicine. Future work will be done in collaboration with Goda's lab to develop and optimize Goda's current RACC system for human samples. This holds the potential to push the system to the highest possible throughput towards the analysis of liquid biopsies and the detection of very rare events such as CTC[1], assessing the applicability into the current health system.

[1] Circulating Tumour Cells

References

[1] J. Gala de Pablo, F. J. Armistead, S. A. Peyman, D. Bonthron, M. Lones, S. Smith, and S. D. Evans, "Biochemical fingerprint of colorectal cancer cell lines using label-free live single-cell Raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 49, pp. 1323–1332, aug 2018. i, 7

[2] J. Gala de Pablo, D. R. Chisholm, A. Steffen, A. K. Nelson, C. Mahler, T. B. Marder, S. A. Peyman, J. M. Girkin, C. A. Ambler, A. Whiting, and S. D. Evans, "Tandem fluorescence and Raman (fluoRaman) characterisation of a novel photosensitiser in colorectal cancer cell line SW480," *The Analyst*, vol. 143, no. 24, pp. 6113–6120, 2018. i, 12, 127

[3] F. J. Armistead, J. Gala De Pablo, H. Gadêlha, S. A. Peyman, and S. D. Evans, "Cells Under Stress: An Inertial-Shear Microfluidic Determination of Cell Behavior," *Biophysical Journal*, vol. 116, pp. 1127–1135, mar 2019. i, ii, 2, 195

[4] M. Van Dyke, *An Album of Fluid Motion*. Parabolic Press, 12th editi ed., 1982. ii, 47

[5] D. J. Collins, B. Morahan, J. Garcia-Bustos, C. Doerig, M. Plebanski, and A. Neild, "Two-dimensional single-cell patterning with one cell per well driven by surface acoustic waves," *Nature Communications*, vol. 6, pp. 1–11, 2015. 2

[6] C. Matthäus, T. Chernenko, C. Stiebing, L. Quintero, M. Miljkovi, L. Milane, A. Kale, M. Amiji, S. Lorkowski, and V. Torchilin, *Confocal Raman Microscopy*, vol. 66 of *Springer Series in Surface Sciences*. Cham: Springer International Publishing, 2018. 2, 13

[7] J. V. Sweedler and E. a. Arriaga, "Single cell analysis," *Analytical and Bioanalytical Chemistry*, vol. 387, no. 1, pp. 1–2, 2006. 2

[8] M. Li, J. Xu, M. Romero-Gonzalez, S. A. Banwart, and W. E. Huang, "Single cell Raman spectroscopy for cell sorting and imaging," *Current Opinion in Biotechnology*, vol. 23, no. 1, pp. 56–63, 2012. 2, 3

[9] H. Kobayashi, C. Lei, Y. Wu, A. Mao, Y. Jiang, B. Guo, Y. Ozeki, and K. Goda, "Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning," *Scientific Reports*, vol. 7, p. 12454, dec 2017. 2

[10] I. W. Schie, J. Rüger, A. S. Mondol, A. Ramoji, U. Neugebauer, C. Krafft, and J. Popp, "High-Throughput Screening Raman Spectroscopy Platform for Label-Free Cellomics," *Analytical Chemistry*, vol. 90, pp. 2023–2030, feb 2018. 3, 12

[11] H. Andersson and A. van den Berg, "Microfluidic devices for cellomics: a review," *Sensors and Actuators B: Chemical*, vol. 92, pp. 315–325, jul 2003. 3

[12] L. Charon and L. Lilge, "Single-Cell Analysis in Microfluidic Devices," in *Encyclopedia of Microfluidics and Nanofluidics*, pp. 3017–3027, New York, NY: Springer New York, 2015. 3

[13] A. Smekal, "Zur Quantentheorie der Dispersion," *Die Naturwissenschaften*, vol. 11, pp. 873–875, oct 1923. 3

[14] C. V. Raman and K. S. Krishnan, "A New Type of Secondary Radiation," *Nature*, vol. 121, pp. 501–502, mar 1928. 3, 24

[15] C. Krafft, "Raman and coherent anti-Stokes Raman scattering microspectroscopy for biomedical applications," *Journal of Biomedical Optics*, vol. 17, p. 040801, apr 2012. 3, 5, 25, 26, 27

[16] C. Matthäus, B. Bird, M. Miljković, T. Chernenko, M. Romeo, and M. Diem, "Chapter 10: Infrared and Raman Microscopy in Cell Biology," in *Biophysical Tools for Biologists: In Vivo Techniques*, pp. 275–308, Methods Cell Biol, 2008. 3, 7, 13

[17] A. F. Palonpon, M. Sodeoka, and K. Fujita, "Molecular imaging of live cells by Raman microscopy," *Current Opinion in Chemical Biology*, vol. 17, pp. 708–715, aug 2013. 4, 6, 7

[18] K. Kong, C. Kendall, N. Stone, and I. Notingher, "Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection," *Advanced Drug Delivery Reviews*, vol. 89, pp. 121–134, jul 2015. 12

[19] R. Smith, K. L. Wright, and L. Ashton, "Raman spectroscopy: an evolving technique for live cell studies," *The Analyst*, vol. 141, no. 12, pp. 3590–3600, 2016. 3, 12

[20] A. F. Chrimes, K. Khoshmanesh, P. R. Stoddart, A. Mitchell, and K. Kalantar-zadeh, "Microfluidics and Raman microscopy: current applications and future challenges," *Chemical Society Reviews*, vol. 42, no. 13, p. 5880, 2013. 3, 18, 19, 161, 165

[21] Y. Zheng, J. Nguyen, Y. Wei, and Y. Sun, "Recent advances in microfluidic techniques for single-cell biophysical characterization," *Lab on a Chip*, vol. 13, no. 13, p. 2464, 2013. 3

[22] J. R. Heath, A. Ribas, and P. S. Mischel, "Single-cell analysis tools for drug discovery and development," *Nature Reviews Drug Discovery*, vol. 15, pp. 204–216, dec 2015. 3

[23] Q. Zhang, P. Zhang, H. Gou, C. Mou, W. E. Huang, M. Yang, J. Xu, and B. Ma, "Towards high-throughput microfluidic Raman-activated cell sorting.," *The Analyst*, pp. 23–25, 2015. 3, 7, 18

[24] C. Mallidis, V. Sanchez, J. Wistuba, F. Wuebbeling, M. Burger, C. Fallnich, and S. Schlatt, "Raman microspectroscopy: shining a new light on reproductive medicine," *Human Reproduction Update*, vol. 20, no. 3, pp. 403–414, 2014. 4, 24, 26, 32

[25] C. Krafft and J. Popp, "The many facets of Raman spectroscopy for biomedical analysis," *Analytical and Bioanalytical Chemistry*, vol. 407, no. 3, pp. 699–717, 2015. 6, 7, 14, 19, 32, 52

[26] S. Webster, *Raman microscopy and optical spectroscopy of conjugated polymers at high pressure*. Phd thesis, University of Leeds, 1994. 24, 25, 26, 28, 29

[27] K. J. Baldwin, *Photophysics of conjugated polymers and oligomers*. Phd thesis, University of Leeds, 1997. 4, 7, 24, 28, 29, 32

[28] S. E. Cope, *Raman Spectroscopy in Photodynamic Therapy and Cancer Diagnosis*. PhD thesis, University of Leeds, 1991. 4, 14, 24, 25, 26, 32, 45, 46

[29] M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljković, C. Krafft, and J. Popp, "Molecular pathology via IR and Raman spectral imaging," *Journal of Biophotonics*, vol. 6, no. 11-12, pp. 855–886, 2013. 4, 12

[30] C. M. Krishna, G. Sockalingum, J. Kurien, L. Rao, L. Venteo, M. Pluot, M. Manfait, and V. Kartha, "Micro-Raman Spectroscopy for Optical Pathology of Oral Squamous Cell Carcinoma," *Applied Spectroscopy*, vol. 58, pp. 1128–1135, sep 2004. 4, 8

[31] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Jablonska, and R. Kordek, "The label-free Raman imaging of human breast cancer," *Journal of Molecular Liquids*, vol. 164, pp. 123–131, nov 2011. 4, 7, 8

[32] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman Spectroscopy of Biological Tissues," *Applied Spectroscopy Reviews*, vol. 42, pp. 493–541, sep 2007. 4, 6, 94, 95, 98, 136

[33] I. Rehman, Z. Movasaghi, and S. Rehman, *Vibrational Spectroscopy for Tissue Analysis*, vol. 20124820 of *Series in Medical Physics and Biomedical Engineering*. CRC Press, sep 2012. 4, 7, 27, 29

[34] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy," *EPJ Techniques and Instrumentation*, vol. 2, no. 1, p. 8, 2015. 4, 38

[35] M. Diem, M. Miljković, B. Bird, T. Chernenko, J. Schubert, E. Marcsisin, A. Mazur, E. Kingston, E. Zuser, K. Papamarkakis, and N. Laver, "Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises," *Spectroscopy: An International Journal*, vol. 27, no. 5-6, pp. 463–496, 2012. 4, 8, 35

[36] B. Alberts, *Molecular Biology of the Cell*. New York: Garland Science, 4th ed., 2002. 5

[37] E. C. Faria and P. Gardner, *Single-Cell Analysis*, vol. 853 of *Methods in Molecular Biology*. Totowa, NJ: Humana Press, 2012. 4

[38] D. W. Mayo, F. A. Miller, and R. W. Hannah, *Course Notes on the Interpretation of Infrared and Raman Spectra*. Hoboken, NJ, USA: John Wiley & Sons, Inc., may 2004. 5, 26, 29

[39] Z. Wen, "Raman spectroscopy of protein pharmaceuticals," *Journal of Pharmaceutical Sciences*, vol. 96, pp. 2861–2878, nov 2007. 5

[40] S. F. El-Mashtoly, H. K. Yosef, D. Petersen, L. Mavarani, A. Maghnouj, S. Hahn, C. Kötting, and K. Gerwert, "Label-Free Raman Spectroscopic Imaging Monitors the Integral Physiologically Relevant Drug Responses in Cancer Cells," *Analytical Chemistry*, vol. 87, pp. 7297–7304, jul 2015. 5, 6, 13

[41] D. Kurouski, R. P. Van Duyne, and I. K. Lednev, "Exploring the structure and formation mechanism of amyloid fibrils by Raman spectroscopy: a review," *The Analyst*, vol. 140, no. 15, pp. 4967–4980, 2015. 6

[42] N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey, and V. E. Anderson, "Raman spectroscopic characterization of secondary structure in natively unfolded proteins: alpha-synuclein.," *Journal of the American Chemical Society*, vol. 126, no. 8, pp. 2399–2408, 2004. 6

[43] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, and M. Baranska, "Raman spectroscopy of proteins: A review," *Journal of Raman Spectroscopy*, vol. 44, no. 8, pp. 1061–1076, 2013. 6, 7, 94, 95

[44] C. David, "Raman Spectroscopy for proteins," *Horiba Scientific Webminar*, pp. 1–53, 2012. 6

[45] A. C. S. Talari, Z. Movasaghi, S. Rehman, and I. ur Rehman, "Raman Spectroscopy of Biological Tissues," *Applied Spectroscopy Reviews*, vol. 50, pp. 46–111, jan 2015. 6

[46] A. Downes and A. Elfick, "Raman Spectroscopy and Related Techniques in Biomedicine," *Sensors*, vol. 10, pp. 1871–1889, mar 2010. 6, 7, 35, 41, 52

[47] A. S. Haka, Z. Volynskaya, J. a. Gardecki, J. Nazemi, R. Shenk, N. Wang, R. R. Dasari, M. Fitzmaurice, and M. S. Feld, "Diagnosing breast cancer using Raman spectroscopy: prospective analysis," *Journal of Biomedical Optics*, vol. 14, no. 5, p. 054023, 2009. 6, 7, 52

[48] L. E. Kamemoto, A. K. Misra, S. K. Sharma, M. T. Goodman, H. Luk, A. C. Dykes, and T. Acosta, "Near-Infrared Micro-Raman Spectroscopy for in Vitro Detection of Cervical Cancer," *Applied Spectroscopy*, vol. 64, pp. 255–261, mar 2010. 7

[49] P. Carey, "Biochemical applications of Raman and resonance Raman spectroscopies," *Journal of Molecular Structure*, vol. 112, p. 337, feb 1984. 7, 25, 26, 27, 32

[50] M. Okada, N. I. Smith, A. F. Palonpon, H. Endo, S. Kawata, M. Sodeoka, and K. Fujita, "Label-free Raman observation of cytochrome c dynamics during apoptosis," *Proceedings of the National Academy of Sciences*, vol. 109, no. 1, pp. 28–32, 2012. 7, 32, 94, 95, 98, 99

[51] A. Gorski, A. Starukhin, S. Stavrov, S. Gawinkowski, and J. Waluk, "Resonance Raman spectroscopy study of protonated porphyrin," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 173, pp. 350–355, 2017. 7

[52] A. Pallaoro, M. R. Hoonejani, G. B. Braun, C. D. Meinhart, and M. Moskovits, "Rapid identification by surface-enhanced Raman spectroscopy of cancer cells at low concentrations flowing in a microfluidic channel," *ACS Nano*, vol. 9, no. 4, pp. 4328–4336, 2015. 7, 37

[53] B. Yan, B. Li, Z. Wen, X. Luo, L. Xue, and L. Li, "Label-free blood serum detection by using surface-enhanced Raman spectroscopy and support vector machine for the preoperative diagnosis of parotid gland tumors," *BMC Cancer*, vol. 15, p. 650, dec 2015. 7

[54] D. Cialla, A. März, R. Böhme, F. Theil, K. Weber, M. Schmitt, and J. Popp, "Surface-enhanced Raman spectroscopy (SERS): progress and trends," *Analytical and Bioanalytical Chemistry*, vol. 403, no. 1, pp. 27–54, 2012. 7

[55] Y. Zhang, H. Hong, D. V. Myklejord, and W. Cai, "Molecular Imaging with SERS-Active Nanoparticles," *Small*, vol. 7, pp. 3261–3269, dec 2011. 7

[56] K. Klein, A. M. Gigler, T. Aschenbrenner, R. Monetti, W. Bunk, F. Jamitzky, G. Morfill, R. W. Stark, and J. Schlegel, "Label-Free Live-Cell Imaging with Confocal Raman Microscopy," *Biophysical Journal*, vol. 102, no. 2, pp. 360–368, 2012. 7, 37, 41

[57] R. J. Swain, G. Jell, and M. M. Stevens, "Non-invasive analysis of cell cycle dynamics in single living cells with Raman micro-spectroscopy," *Journal of Cellular Biochemistry*, vol. 104, no. 4, pp. 1427–1438, 2008. 7, 35, 37, 41

[58] Z. J. Smith, C. Lee, T. Rojalin, R. P. Carney, S. Hazari, A. Knudson, K. Lam, H. Saari, E. L. Ibañez, T. Viitala, T. Laaksonen, M. Yliperttula, and S. Wachsmann-Hogiu, "Single exosome study reveals subpopulations distributed among cell lines with variability related to membrane content," *Journal of Extracellular Vesicles*, vol. 4, p. 28533, jan 2015. 7, 12, 13, 14

[59] A. C. S. Talari, C. A. Evans, I. Holen, R. E. Coleman, and I. U. Rehman, "Raman spectroscopic analysis differentiates between breast cancer cell lines," *Journal of Raman Spectroscopy*, vol. 46, pp. 421–427, may 2015. 7, 37

[60] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Jablonska-Gajewicz, and R. Kordek, "Raman 'optical biopsy' of human breast cancer," *Progress in Biophysics and Molecular Biology*, vol. 108, pp. 74–81, jan 2012. 98

[61] B. Brozek-Pluska, J. Musial, R. Kordek, E. Bailo, T. Dieing, and H. Abramczyk, "Raman spectroscopy and imaging: applications in human breast cancer diagnosis," *The Analyst*, vol. 137, no. 16, p. 3773, 2012. 7, 8, 17

[62] Y. Oshima, H. Shinzawa, T. Takenaka, C. Furihata, and H. Sato, "Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy," *Journal of Biomedical Optics*, vol. 15, no. 1, p. 017009, 2010. 7, 37

[63] U. Neugebauer, T. Bocklitz, J. H. Clement, C. Krafft, and J. Popp, "Towards detection and identification of circulating tumour cells using Raman spectroscopy," *The Analyst*, vol. 135, no. 12, p. 3178, 2010. 7

[64] M. Larraona-Puy, A. Ghita, A. Zoladek, W. Perkins, S. Varma, I. H. Leach, A. a. Koloydenko, H. Williams, and I. Notingher, "Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma," *Journal of Biomedical Optics*, vol. 14, no. 5, p. 054031, 2009.

[65] E. Canetta, A. Riches, E. Borger, S. Herrington, K. Dholakia, and A. K. Adya, "Discrimination of bladder cancer cells from normal urothelial cells with high specificity and sensitivity: Combined application of atomic force microscopy and modulated Raman spectroscopy," *Acta Biomaterialia*, vol. 10, pp. 2043–2055, may 2014. 7

[66] "Leeds GIFT 2 Research Tissue Project." 8

[67] F. Lyng, E. Gazi, and P. Gardner, *Biomedical Applications of Synchrotron Infrared Microspectroscopy*, vol. 11 of *RSC Analytical Spectroscopy Series*. Cambridge: Royal Society of Chemistry, 2010. 8

[68] E. Ó Faoláin, M. B. Hunter, J. M. Byrne, P. Kelehan, M. McNamara, H. J. Byrne, and F. M. Lyng, "A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy," *Vibrational Spectroscopy*, vol. 38, pp. 121–127, jul 2005.

[69] S. M. Ali, F. Bonnier, A. Tfayli, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Clive Lee, F. M. Lyng, and H. J. Byrne, "Raman spectroscopic analysis of human skin tissue sections ex-vivo: evaluation of the effects of tissue processing and dewaxing.," *Journal of biomedical optics*, vol. 18, no. 6, p. 061202, 2013. 8

[70] S. Anand, R. Cicchi, F. Martelli, F. Giordano, A. M. Buccoliero, R. Guerrini, and F. S. Pavone, "Effects of formalin fixation on tissue optical properties of in-vitro brain samples," in *Optical Interactions with Tissue and Cells XXVI* (E. D. Jansen, ed.), vol. 9321, (San Francisco, California), p. 93210Z, mar 2015. 8

[71] F. Lyng, E. Gazi, and P. Gardner, "Preparation of Tissues and Cells for Infrared and Raman Spectroscopy and Imaging," *Books/Book Chapters*, no. July, 2011. 8

[72] M. Hedegaard, C. Matthäus, S. Hassing, C. Krafft, M. Diem, and J. Popp, "Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging," *Theoretical Chemistry Accounts*, vol. 130, pp. 1249–1260, dec 2011. 8, 35

[73] M. M. Mariani, P. Lampen, J. Popp, B. R. Wood, and V. Deckert, "Impact of fixation on in vitro cell culture lines monitored with Raman spectroscopy.," *The Analyst*, vol. 134, no. 6, pp. 1154–1161, 2009. 8

[74] A. D. Meade, C. Clarke, F. Draux, G. D. Sockalingum, M. Manfait, F. M. Lyng, and H. J. Byrne, "Studies of chemical fixation effects in human cell lines using Raman microspectroscopy," *Analytical and Bioanalytical Chemistry*, vol. 396, no. 5, pp. 1781–1791, 2010. 8, 41

[75] S. Jeong, Y.-i. Kim, H. Kang, G. Kim, M. G. Cha, H. Chang, K. O. Jung, Y.-H. Kim, B.-H. Jun, D. W. Hwang, Y.-S. Lee, H. Youn, Y.-S. Lee, K. W. Kang, D. S. Lee, and D. H. Jeong, "Fluorescence-Raman Dual Modal Endoscopic System for Multiplexed Molecular Diagnostics," *Scientific Reports*, vol. 5, p. 9455, aug 2015. 9, 128

[76] A. Lombardini, V. Mytskaniuk, S. Sivankutty, E. R. Andresen, X. Chen, J. Wenger, M. Fabert, N. Joly, F. Louradour, A. Kudlinski, and H. Rigneault, "High-resolution multimodal flexible coherent Raman endoscope," *Light: Science & Applications*, vol. 7, p. 10, dec 2018. 9

[77] P. Zirak, G. Matz, B. Messerschmidt, T. Meyer, M. Schmitt, J. Popp, O. Uckermann, R. Galli, M. Kirsch, M. J. Winterhalder, and A. Zumbusch, "Invited Article: A rigid coherent anti-Stokes Raman scattering endoscope with high resolution and a large field of view," *APL Photonics*, vol. 3, p. 092409, sep 2018. 9

[78] K. Hirose, S. Fukushima, T. Furukawa, H. Niioka, and M. Hashimoto, "Invited Article: Label-free nerve imaging with a coherent anti-Stokes Raman scattering rigid endoscope using two optical fibers for laser delivery," *APL Photonics*, vol. 3, p. 092407, sep 2018.

[79] C. L. Zavaleta, E. Garai, J. T. C. Liu, S. Sensarn, M. J. Mandella, D. Van de Sompel, S. Friedland, J. Van Dam, C. H. Contag, and S. S. Gambhir, "A Raman-based endoscopic strategy for multiplexed molecular imaging," *Proceedings of the National Academy of Sciences*, vol. 110, pp. E2288–E2297, jun 2013. 9

[80] V. V. Pully, A. T. M. Lenferink, and C. Otto, "Time-lapse Raman imaging of single live lymphocytes," *Journal of Raman Spectroscopy*, vol. 42, no. 2, pp. 167–173, 2011. 9, 98

[81] T. J. Harvey, E. C. Faria, A. Henderson, E. Gazi, A. D. Ward, N. W. Clarke, M. D. Brown, R. D. Snook, and P. Gardner, "Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers," *Journal of Biomedical Optics*, vol. 13, no. 6, p. 064004, 2008. 9

[82] S. Casabella, P. Scully, N. Goddard, and P. Gardner, "Automated analysis of single cells using Laser Tweezers Raman Spectroscopy," *The Analyst*, 2016. 9, 20

[83] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, pp. 394–424, nov 2018. 9

[84] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," *Gut*, vol. 66, no. 4, pp. 683–691, 2017. 9

[85] C. C. Applegate and M. A. Lane, "Role of retinoids in the prevention and treatment of colorectal cancer," *World Journal of Gastrointestinal Oncology*, vol. 7, no. 10, p. 184, 2015. 9, 10

[86] R. Abdel-Samad, P. Aouad, H. Gali-Muhtasib, Z. Sweidan, R. Hmadi, H. Kadara, E. L. D'Andrea, A. Fucci, C. Pisano, and N. Darwiche, "Mechanism of action of the atypical retinoid ST1926 in colorectal cancer: DNA damage and DNA polymerase α.," *American journal of cancer research*, vol. 8, no. 1, pp. 39–55, 2018. 9

[87] M. Riihimäki, A. Hemminki, J. Sundquist, and K. Hemminki, "Patterns of metastasis in colon and rectal cancer.," *Scientific reports*, vol. 6, p. 29765, 2016. 9

[88] J. J. Granados-Romero, A. I. Valderrama-Treviño, E. H. Contreras-Flores, B. Barrera-Mera, M. Herrera Enríquez, K. Uriarte-Ruíz, J. C. Ceballos-Villalba, A. G. Estrada-Mata, C. Alvarado Rodríguez, and G. Arauz-Peña, "Colorectal cancer: a review," *International Journal of Research in Medical Sciences*, vol. 5, p. 4667, oct 2017. 9, 10, 11

[89] W. M. Grady and S. D. Markowitz, "The Molecular Pathogenesis of Colorectal Cancer and Its Potential Application to Colorectal Cancer Screening," *Digestive Diseases and Sciences*, vol. 60, pp. 762–772, mar 2015. 10

[90] D. Ahmed, P. W. Eide, I. A. Eilertsen, S. A. Danielsen, M. Eknæs, M. Hektoen, G. E. Lind, and R. A. Lothe, "Epigenetic and genetic features of 24 colon cancer cell lines.," *Oncogenesis*, vol. 2, no. 0424, p. e71, 2013. 10, 11

[91] A. Leslie, F. A. Carey, N. R. Pratt, and R. J. C. Steele, "The colorectal adenoma-carcinoma sequence," *British Journal of Surgery*, vol. 89, pp. 845–860, jul 2002. 10

[92] S. Singla, D. Kaushal, H. Sagoo, and N. Calton, "Comparative analysis of colorectal carcinoma staging using operative, histopathology and computed tomography findings," *International Journal of Applied and Basic Medical Research*, vol. 7, no. 1, p. 10, 2017. 11

[93] E. Buck, A. Eyzaguirre, S. Barr, S. Thompson, R. Sennello, D. Young, K. K. Iwata, N. W. Gibson, P. Cagnoni, and J. D. Haley, "Loss of homotypic cell adhesion by epithelial-mesenchymal transition or mutation limits sensitivity to epidermal growth factor receptor inhibition," *Mol Cancer Ther*, vol. 6, no. 2, pp. 532–541, 2007. 11, 106

[94] R. E. Hewitt, A. McMarlin, D. Kleiner, R. Wersto, P. Martin, M. Tsoskas, G. W. H. Stamp, and W. G. Stetler-Stevenson, "Validation of a model of colon cancer progression," *The Journal of Pathology*, vol. 192, pp. 446–454, dec 2000.

[95] A. Leibovitz, J. C. Stinson, W. B. McCombs, C. E. McCoy, K. C. Mazur, and N. D. Mabry, "Classification of human colorectal adenocarcinoma cell lines," *Cancer research*, vol. 36, no. 12, p. 4562, 1976. 11

[96] I. Yousef, J. Bréard, N. SidAhmed-Adrar, A. Maâmer-Azzabi, C. Marchal, P. Dumas, and F. Le Naour, "Infrared spectral signatures of CDCP1-induced effects in colon carcinoma cells," *The Analyst*, vol. 136, no. 24, p. 5162, 2011. 11, 96, 108, 193

[97] D. Tsikritsis, S. Richmond, P. Stewart, A. Elfick, and A. Downes, "Label-free identification and characterization of living human primary and secondary tumour cells.," *The Analyst*, vol. 140, no. 15, pp. 5162–8, 2015. 11, 96, 98, 99, 108, 134, 193

[98] J. Fogh, ed., *Human Tumor Cells in Vitro*. Boston, MA: Springer US, 1975. 11

[99] A. Beljebbar, H. Morjani, G. D. Sockalingum, and M. Manfait, "<title>Rapid identification of the multidrug resistance in the human leukemic cells by near-infrared Fourier transform Raman microspectroscopy</title>," in *Proc. SPIE 3257, Infrared Spectroscopy: New Tool in Medicine* (H. H. Mantsch and M. Jackson, eds.), vol. 3257, pp. 62–65, apr 1998. 12

[100] C. M. Krishna, G. D. Sockalingum, G. Kegelaer, S. Rubin, V. B. Kartha, and M. Manfait, "Micro-Raman spectroscopy of mixed cancer cell populations," *Vibrational Spectroscopy*, vol. 38, pp. 95–100, jul 2005. 12

[101] H.-H. Lin, Y.-C. Li, C.-H. Chang, C. Liu, A. L. Yu, and C.-H. Chen, "Single Nuclei Raman Spectroscopy for Drug Evaluation," *Analytical Chemistry*, vol. 84, pp. 113–120, jan 2012. 12

[102] K. Briviba, R. Bornemann, and U. Lemmer, "Visualization of astaxanthin localization in HT29 human colon adenocarcinoma cells by combined confocal resonance Raman and fluorescence microspectroscopy," *Molecular Nutrition & Food Research*, vol. 50, pp. 991–995, nov 2006. 12

[103] H. Zhang, J. Zheng, A. Liu, H. Xiao, and L. He, "Label-free Imaging and Characterization of Cancer Cell Responses to Polymethoxyflavones Using Raman Microscopy," *Journal of Agricultural and Food Chemistry*, vol. 64, pp. 9708–9713, dec 2016. 12

[104] S. Akyuz, A. Ozel, K. Balci, T. Akyuz, A. Coker, E. Arisan, N. Palavan-Unsal, and A. Ozalpan, "Raman micro-spectroscopic investigation of the interaction of cultured HCT116 colon cancer cells with alpha-difluoromethylornithine (DFMO), an irreversible inhibitor of ornithine decarboxylase," *Journal of Molecular Structure*, vol. 993, pp. 319–323, may 2011. 12

[105] S. Akyuz, A. Ozel, K. Balci, T. Akyuz, A. Coker, E. Arisan, N. Palavan-Unsal, and A. Ozalpan, "Raman micro-spectroscopic analysis of cultured HCT116 colon cancer cells in the presence of roscovitine," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 78, pp. 1540–1547, may 2011. 12

[106] X. Su, S. Fang, D. Zhang, Q. Zhang, X. Lu, J. Tian, J. Fan, and LiyunZhong, "Raman spectrum reveals Mesenchymal stem cells inhibiting HL60 cells growth," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 177, pp. 15–19, apr 2017. 12

[107] C. Cai, R. Chen, J. Lin, Y. Li, and S. Feng, "Micro-Raman spectroscopy of single leukemic cells," *Chinese Optics Letters*, vol. 6, no. 12, pp. 938–940, 2008. 12

[108] M. K. Hammoud, H. K. Yosef, T. Lechtonen, K. Aljakouch, M. Schuler, W. Alsaidi, I. Daho, A. Maghnouj, S. Hahn, S. F. El-Mashtoly, and K. Gerwert, "Raman micro-spectroscopy monitors acquired resistance to targeted cancer therapy at the cellular level," *Scientific Reports*, vol. 8, no. 1, p. 15278, 2018. 12, 13

[109] C. Krafft, M. Schmitt, I. W. Schie, D. Cialla-May, C. Matthäus, T. Bocklitz, and J. Popp, "Label-Free Molecular Imaging of Biological Cells and Tissues by Linear and Nonlinear Raman Spectroscopic Approaches," *Angewandte Chemie International Edition*, vol. 56, pp. 4392–4430, apr 2017. 12

[110] I. W. Schie and T. Huser, "Methods and applications of Raman microspectroscopy to single-cell analysis.," *Applied spectroscopy*, vol. 67, pp. 813–828, aug 2013. 12

[111] N. K. Das, Y. Dai, P. Liu, C. Hu, L. Tong, X. Chen, and Z. J. Smith, "Raman Plus X: Biomedical Applications of Multimodal Raman Spectroscopy," *Sensors*, vol. 17, p. 1592, jul 2017. 12, 127, 128

[112] I. Notingher and L. L. Hench, "Raman microspectroscopy: a noninvasive tool for studies of individual living cells in vitro," *Expert Review of Medical Devices*, vol. 3, pp. 215–234, mar 2006. 12

[113] H. J. Byrne, M. Baranska, G. J. Puppels, N. Stone, B. Wood, K. M. Gough, P. Lasch, P. Heraud, J. Sulé-Suso, and G. D. Sockalingum, "Spectropathology for the next generation: Quo vadis?," *The Analyst*, vol. 140, no. 7, pp. 2066–2073, 2015. 12

[114] J. Phillip Kennedy, P. Jeffrey Conn, and C. W. Lindsley, "A novel class of H3 antagonists derived from the natural product guided synthesis of unnatural analogs of the marine bromopyrrole alkaloid dispyrin," *Bioorganic & Medicinal Chemistry Letters*, vol. 19, pp. 3204–3208, jun 2009. 12

[115] B. Kann, H. L. Offerhaus, M. Windbergs, and C. Otto, "Raman microscopy for cellular investigations - From single cell imaging to drug carrier uptake visualization," *Advanced Drug Delivery Reviews*, vol. 89, pp. 71–90, 2015. 12

[116] F. Hu, L. Wei, C. Zheng, Y. Shen, and W. Min, "Live-cell vibrational imaging of choline metabolites by stimulated Raman scattering coupled with isotope-based metabolic labeling," *The Analyst*, vol. 139, no. 10, pp. 2312–2317, 2014. 12

[117] C. Matthäus, A. Kale, T. Chernenko, V. Torchilin, and M. Diem, "New ways of imaging uptake and intracellular fate of liposomal drug carrier systems inside individual cells, based on raman microscopy," *Molecular Pharmaceutics*, vol. 5, no. 2, pp. 287–293, 2008. 12

[118] S. Hong, L. Lin, M. Xiao, and X. Chen, "Live-cell bioorthogonal Raman imaging," *Current Opinion in Chemical Biology*, vol. 24, pp. 91–96, 2015. 12

[119] W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton, and A. N. Hulme, "Stimulated Raman scattering microscopy: An emerging tool for drug discovery," *Chemical Society Reviews*, vol. 45, no. 8, pp. 2075–2089, 2016.

[120] H. Yamakoshi, A. F. Palonpon, K. Dodo, J. Ando, S. Kawata, K. Fujita, and M. Sodeoka, "Simultaneous imaging of protonated and deprotonated carbonylcyanide p-trifluoromethoxyphenylhydrazone in live cells by Raman microscopy," *Chem. Commun.*, vol. 50, no. 11, pp. 1341–1343, 2014. 12, 13

[121] S. Hong, T. Chen, Y. Zhu, A. Li, Y. Huang, and X. Chen, "Live-Cell Stimulated Raman Scattering Imaging of Alkyne-Tagged Biomolecules," *Angewandte Chemie International Edition*, vol. 53, pp. 5827–5831, jun 2014. 128

[122] F. Hu, C. Zeng, R. Long, Y. Miao, L. Wei, Q. Xu, and W. Min, "Supermultiplexed optical imaging and barcoding with engineered polyynes," *Nature Methods*, vol. 15, no. 3, pp. 194–200, 2018. 128, 129

[123] A. F. Palonpon, J. Ando, H. Yamakoshi, K. Dodo, M. Sodeoka, S. Kawata, and K. Fujita, "Raman and SERS microscopy for molecular imaging of live cells," *Nature Protocols*, vol. 8, pp. 677–692, apr 2013. 128

[124] H. Yamakoshi, K. Dodo, M. Okada, J. Ando, A. Palonpon, K. Fujita, S. Kawata, and M. Sodeoka, "Imaging of EdU, an alkyne-tagged cell proliferation probe, by Raman microscopy," *Journal of the American Chemical Society*, vol. 133, no. 16, pp. 6102–6105, 2011. 128

[125] H. Yamakoshi, K. Dodo, A. Palonpon, J. Ando, K. Fujita, S. Kawata, and M. Sodeoka, "Alkyne-tag Raman imaging for visualization of mobile small molecules in live cells," *Journal of the American Chemical Society*, vol. 134, no. 51, pp. 20681–20689, 2012. 12, 128, 132

[126] K. Aljakouch, T. Lechtonen, H. K. Yosef, M. K. Hammoud, W. Alsaidi, C. Kötting, C. Mügge, R. Kourist, S. F. El-Mashtoly, and K. Gerwert, "Raman Microspectroscopic Evidence for the Metabolism of a Tyrosine Kinase Inhibitor, Neratinib, in Cancer Cells," *Angewandte Chemie - International Edition*, vol. 57, no. 24, pp. 7250–7254, 2018. 13

[127] S. F. El-Mashtoly, D. Petersen, H. K. Yosef, A. Mosig, A. Reinacher-Schick, C. Kötting, and K. Gerwert, "Label-free imaging of drug distribution and metabolism in colon cancer cells by Raman microscopy," *The Analyst*, vol. 139, no. 5, p. 1155, 2014. 13

[128] H. K. Yosef, L. Mavarani, A. Maghnouj, S. Hahn, S. F. El-Mashtoly, and K. Gerwert, "In vitro prediction of the efficacy of molecularly targeted cancer therapy by Raman spectral imaging Raman4Clinics," *Analytical and Bioanalytical Chemistry*, vol. 407, no. 27, pp. 8321–8331, 2015. 13

[129] H. K. Yosef, T. Frick, M. K. Hammoud, A. Maghnouj, S. Hahn, K. Gerwert, and S. F. El-Mashtoly, "Exploring the efficacy and cellular uptake of sorafenib in colon cancer cells by Raman micro-spectroscopy," *The Analyst*, 2018. 13

[130] D. Fu, J. Zhou, W. S. Zhu, P. W. Manley, Y. K. Wang, T. Hood, A. Wylie, and X. S. Xie, "Imaging the intracellular distribution of tyrosine kinase inhibitors in living cells with quantitative hyperspectral stimulated Raman scattering," *Nature Chemistry*, vol. 6, pp. 614–622, jul 2014. 13

[131] H. Nawaz, F. Bonnier, P. Knief, O. Howe, F. M. Lyng, A. D. Meade, and H. J. Byrne, "Evaluation of the potential of Raman microspectroscopy for prediction of chemotherapeutic response to cisplatin in lung adenocarcinoma," *The Analyst*, vol. 135, no. 12, p. 3070, 2010. 13

[132] H. Nawaz, F. Bonnier, A. D. Meade, F. M. Lyng, and H. J. Byrne, "Comparison of subcellular responses for the evaluation and prediction of the chemotherapeutic response to cisplatin in lung adenocarcinoma using Raman spectroscopy," *The Analyst*, vol. 136, no. 12, p. 2450, 2011. 13

[133] Z. Farhane, F. Bonnier, and H. J. Byrne, "Monitoring doxorubicin cellular uptake and trafficking using in vitro Raman microspectroscopy: short and long time exposure effects on lung cancer cell lines," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 5, pp. 1333–1346, 2017. 13

[134] Z. Farhane, F. Bonnier, O. Howe, A. Casey, and H. J. Byrne, "Doxorubicin kinetics and effects on lung cancer cell lines using in vitro Raman micro-spectroscopy: binding signatures, drug resistance and DNA repair," *Journal of Biophotonics*, vol. 11, no. 1, pp. 1–14, 2018. 13

[135] R. Mouras, G. Rischitor, A. Downes, D. Salter, and A. Elfick, "Nonlinear optical microscopy for drug delivery monitoring and cancer tissue imaging," *Journal of Raman Spectroscopy*, vol. 41, pp. 848–852, aug 2010. 13

[136] A. A. van Apeldoorn, H.-J. van Manen, J. M. Bezemer, J. D. de Bruijn, C. A. van Blitterswijk, and C. Otto, "Raman Imaging of PLGA Microsphere Degradation Inside Macrophages," *Journal of the American Chemical Society*, vol. 126, pp. 13226–13227, oct 2004. 13

[137] P. Xu, E. Gullotti, L. Tong, C. B. Highley, D. R. Errabelli, T. Hasan, J.-X. Cheng, D. S. Kohane, and Y. Yeo, "Intracellular Drug Delivery by Poly(lactic- co -glycolic acid) Nanoparticles, Revisited," *Molecular Pharmaceutics*, vol. 6, pp. 190–201, feb 2009. 13

[138] G. Romero, I. Estrela-Lopis, J. Zhou, E. Rojas, A. Franco, C. S. Espinel, A. G. Fernandez, C. Gao, E. Donath, and S. E. Moya, "Surface Engineered Poly(lactide- co -glycolide) Nanoparticles for Intracellular Delivery: Uptake and Cytotoxicity-A Confocal Raman Microscopic Study," *Biomacromolecules*, vol. 11, pp. 2993–2999, nov 2010. 13

[139] T. Chernenko, C. Matthäus, L. Milane, L. Quintero, M. Amiji, and M. Diem, "Label-Free Raman Spectral Imaging of Intracellular Delivery and Degradation of Polymeric Nanoparticle Systems," *ACS Nano*, vol. 3, pp. 3552–3559, nov 2009. 13

[140] T. Chernenko, F. Buyukozturk, M. Miljkovic, R. Carrier, M. Diem, and M. Amiji, "Label-free Raman microspectral analysis for comparison of cellular uptake and distribution between nontargeted and EGFR-targeted biodegradable polymeric nanoparticles," *Drug Delivery and Translational Research*, vol. 3, pp. 575–586, dec 2013. 13

[141] T. Chernenko, R. R. Sawant, M. Miljkovic, L. Quintero, M. Diem, and V. Torchilin, "Raman Microscopy for Noninvasive Imaging of Pharmaceutical Nanocarriers: Intracellular Distribution of Cationic Liposomes of Different Composition," *Molecular Pharmaceutics*, vol. 9, pp. 930–936, apr 2012. 13

[142] N. L. Garrett, A. Lalatsa, D. Begley, L. Mihoreanu, I. F. Uchegbu, A. G. Schätzlein, and J. Moger, "Label-free imaging of polymeric nanomedicines using coherent anti-stokes Raman scattering microscopy," *Journal of Raman Spectroscopy*, vol. 43, pp. 681–688, may 2012. 14

[143] A. Kawczyk-Krupka, A. M. Bugaj, W. Latos, K. Zaremba, K. Wawrzyniec, M. Kucharzewski, and A. Sieroń, "Photodynamic therapy in colorectal cancer treatment-The state of the art in preclinical research," *Photodiagnosis and Photodynamic Therapy*, vol. 13, pp. 158–174, 2016. 14, 16

[144] K. Wawrzyniec, A. Kawczyk-Krupka, Z. P. Czuba, W. Król, and A. Sieroń, "The influence of ALA-mediated photodynamic therapy on secretion of selected growth factors by colon cancer cells in hypoxia-like environment in vitro," *Photodiagnosis and Photodynamic Therapy*, vol. 12, no. 4, pp. 598–611, 2015. 14, 16

[145] N. Hodgkinson, C. A. Kruger, and H. Abrahamse, "Targeted photodynamic therapy as potential treatment modality for the eradication of colon cancer and colon cancer stem cells," *Tumor Biology*, vol. 39, no. 10, pp. 1–17, 2017. 14

[146] I. Pence and A. Mahadevan-Jansen, "Clinical instrumentation and applications of Raman spectroscopy," *Chemical Society Reviews*, vol. 45, no. 7, pp. 1958–1979, 2016. 14

[147] M. Sachar, K. E. Anderson, and X. Ma, "Protoporphyrin IX: the Good, the Bad, and the Ugly," *Journal of Pharmacology and Experimental Therapeutics*, vol. 356, pp. 267–275, jan 2016. 15, 16, 111, 112

[148] K. T. Moesta, B. Ebert, T. Handke, D. Nolte, C. Nowak, W. E. Haensch, R. K. Pandey, T. J. Dougherty, H. Rinneberg, and P. M. Schlag, "Protoporphyrin IX occurs naturally in colorectal cancers and their metastases.," *Cancer research*, vol. 61, pp. 991–9, feb 2001. 15, 16

[149] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, jan 2000. 15

[150] D. Chiabrando, S. Mercurio, and E. Tolosano, "Heme and erythropoieis: More than a structural role," *Haematologica*, vol. 99, no. 6, pp. 973–983, 2014. 15, 16, 112

[151] Y. Kamada, Y. Murayama, U. Ota, K. Takahashi, T. Arita, T. Kosuga, H. Konishi, R. Morimura, S. Komatsu, A. Shiozaki, Y. Kuriu, H. Ikoma, M. Nakanishi, D. Ichikawa, H. Fujiwara, K. Okamoto, T. Tanaka, and E. Otsuji, "Urinary 5-Aminolevulinic Acid Concentrations as a Potential Tumor Marker for Colorectal Cancer Screening and Recurrence.," *Anticancer research*, vol. 36, pp. 2445–50, may 2016. 15

[152] H. Fukuda, S. Paredes, and A. M. Batlle, "Tumor-localizing properties of porphyrins. In vitro studies using the porphyrin precursor, aminolevulinic acid, in free and liposome encapsulated forms.," *Drug design and delivery*, vol. 5, pp. 133–9, dec 1989. 15, 16, 112

[153] R. C. Krieg, H. Messmann, J. Rauch, S. Seeger, and R. Knuechel, "Metabolic Characterization of Tumor Cell-specific Protoporphyrin IX Accumulation After Exposure to 5-Aminolevulinic Acid in Human Colonic Cells," *Photochemistry and Photobiology*, vol. 76, pp. 518–525, may 2007. 16, 113, 127, 193

[154] O. A. Gederaas, S. A. Schønberg, S. Ramstad, K. Berg, A. Johnsson, and H. E. Krokan, "Cell specific effects of polyunsaturated fatty acids on 5-aminolevulinic acid based photosensitization," *Photochemical and Photobiological Sciences*, vol. 4, no. 4, pp. 383–389, 2005. 16, 113

[155] A. Kawczyk-Krupka, K. Sieroń-Stołtny, W. Latos, Z. P. Czuba, B. Kwiatek, M. Potempa, K. Wasilewska, W. Król, and A. Stanek, "ALA-induced photodynamic effect on vitality, apoptosis, and secretion of vascular endothelial growth factor (VEGF) by colon cancer cells in normoxic environment in vitro," *Photodiagnosis and Photodynamic Therapy*, vol. 13, pp. 308–315, 2016. 16

[156] H. Brunner, F. Hausmann, and R. Knuechel, "New 5-aminolevulinic acid esters - efficient protoporphyrin precursors for photodetection and photodynamic therapy," *Photochem. Photobiol.*, vol. 78, no. 5, pp. 481–486, 2003. 16

[157] A. Sznarkowska, K. Maleńczyk, L. Kadziński, K. P. Bielawski, B. Banecki, and J. Zawacka-Pankau, "Targeting of p53 and its homolog p73 by protoporphyrin IX," *FEBS Letters*, vol. 585, pp. 255–260, jan 2011. 16

[158] C. G. Hadjipanayis and W. Stummer, "5-ALA and FDA approval for glioma surgery," *Journal of Neuro-Oncology*, vol. 141, pp. 479–486, feb 2019. 16

[159] C. Ambler, D. R. Chisholm, and A. Whiting, "Synthetic retinoid (in cell killing) UK Patent Application No. GB1613712.7," 2016. 17, 129

[160] D. R. Chisholm, R. Lamb, T. Pallett, V. Affleck, C. Holden, J. Marrison, P. O'Toole, P. D. Ashton, K. Newling, A. Steffen, A. K. Nelson, C. Mahler, R. Valentine, T. S. Blacker, A. J. Bain, J. Girkin, T. B. Marder, A. Whiting, and C. A. Ambler, "Photoactivated cell-killing involving a low molecular weight, donor–acceptor diphenylacetylene," *Chemical Science*, vol. 10, no. 17, pp. 4673–4683, 2019. 17, 129, 133, 145

[161] D. R. Chisholm, *Synthesis and structures of fluorescent retinoid receptor modulators*. PhD thesis, University of Durham, 2017. 17, 129

[162] T. L. Freeman, S. E. Cope, M. R. Stringer, J. E. Cruse-Sawyer, D. N. Batchelder, and S. B. Brown, "Raman spectroscopy for the determination of photosensitizer localization in cells," *Journal of Raman Spectroscopy*, vol. 28, no. 8, pp. 641–643, 1997. 17

[163] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Musial, and R. Kordek, "Oncologic photodynamic diagnosis and therapy: Confocal Raman/fluorescence imaging of metal phthalocyanines in human breast cancer tissue in vitro," *Analyst*, vol. 139, pp. 5547–5559, aug 2014. 17

[164] H. Abramczyk, B. Brozek-Pluska, M. Tondusson, and E. Freysz, "Ultrafast dynamics of metal complexes of tetrasulfonated phthalocyanines at biological interfaces: Comparison between photochemistry in solutions, films, and noncancerous and cancerous human breast tissues," *Journal of Physical Chemistry C*, vol. 117, no. 10, pp. 4999–5013, 2013. 17

[165] B. Brozek-Pluska and M. Kopec, "Raman microspectroscopy of Hematoporphyrins. Imaging of the noncancerous and the cancerous human breast tissues with photosensitizers," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 169, pp. 182–191, 2016. 17

[166] T. Bhattacharjee, L. C. Fontana, L. Raniero, and J. Ferreira-Strixino, "In vivo Raman spectroscopy of breast tumors prephotodynamic and postphotodynamic therapy," *Journal of Raman Spectroscopy*, vol. 49, pp. 786–791, may 2018. 18

[167] A. B. Veloso, J. P. Longo, L. A. Muehlmann, B. F. Tollstadius, P. E. Souza, R. B. Azevedo, P. C. Morais, and S. W. Da Silva, "SERS Investigation of Cancer Cells Treated with PDT: Quantification of Cell Survival and Follow-up," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017. 18

[168] A. M. Fales, B. M. Crawford, and T. Vo-Dinh, "Folate Receptor-Targeted Theranostic Nanoconstruct for Surface-Enhanced Raman Scattering Imaging and Photodynamic Therapy," *ACS Omega*, vol. 1, no. 4, pp. 730–735, 2016. 18

[169] A. M. Fales, H. Yuan, and T. Vo-Dinh, "Cell-penetrating peptide enhanced intracellular Raman imaging and photodynamic therapy," *Molecular Pharmaceutics*, vol. 10, no. 6, pp. 2291–2298, 2013. 18

[170] A. Farhadi, Á. Roxin, B. C. Wilson, and G. Zheng, "Nano-enabled SERS reporting photosensitizers," *Theranostics*, vol. 5, no. 5, pp. 469–476, 2015. 18

[171] G. Lajos, D. Jancura, P. Miskovsky, J. V. García-Ramos, and S. Sanchez-Cortes, "Interaction of the Photosensitizer Hypericin with Low-Density Lipoproteins and Phosphatidylcholine: A Surface-Enhanced Raman Scattering and Surface-Enhanced Fluorescence Study," *The Journal of Physical Chemistry C*, vol. 113, pp. 7147–7154, apr 2009. 18

[172] H. S. Kim, T. P. Devarenne, and A. Han, "A high-throughput microfluidic single-cell screening platform capable of selective cell extraction," *Lab on a Chip*, vol. 15, no. 11, pp. 2467–2475, 2015.

[173] T. Nagy-Simon, M. Potara, A. M. Craciun, E. Licarete, and S. Astilean, "IR780-dye loaded gold nanoparticles as new near infrared activatable nanotheranostic agents for simultaneous photodynamic and photothermal therapy and intracellular tracking by surface enhanced resonant Raman scattering imaging," *Journal of Colloid and Interface Science*, vol. 517, pp. 239–250, 2018.

[174] N. Nwahara, O. J. Achadu, and T. Nyokong, "In-situ synthesis of gold nanoparticles on graphene quantum dots-phthalocyanine nanoplatforms: First description of the photophysical and surface enhanced Raman scattering behaviour," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 359, pp. 131–144, 2018.

[175] T. Simon, M. Potara, A. M. Gabudean, E. Licarete, M. Banciu, and S. Astilean, "Designing Theranostic Agents Based on Pluronic Stabilized Gold Nanoaggregates Loaded with Methylene Blue for Multimodal Cell Imaging and Enhanced Photodynamic Therapy," *ACS Applied Materials and Interfaces*, vol. 7, no. 30, pp. 16191–16201, 2015.

[176] L. Zhao, T. H. Kim, H. W. Kim, J. C. Ahn, and S. Y. Kim, "Surface-enhanced Raman scattering (SERS)-active gold nanochains for multiplex detection and photodynamic therapy of cancer," *Acta Biomaterialia*, vol. 20, pp. 155–164, 2015.

[177] Z. Zhou, S. Peng, M. Sui, S. Chen, L. Huang, H. Xu, and T. Jiang, "Multifunctional nanocomplex for surface-enhanced Raman scattering imaging and near-infrared photodynamic antimicrobial therapy of vancomycin-resistant bacteria," *Colloids and Surfaces B: Biointerfaces*, vol. 161, pp. 394–402, 2018. 18

[178] X. Yuan, Y. Song, Y. Song, J. Xu, Y. Wu, A. Glidle, M. Cusack, U. Z. Ijaz, J. M. Cooper, W. E. Huang, and H. Yin, "Effect of Laser Irradiation on Cell Function and Its Implications in Raman Spectroscopy," *Applied and Environmental Microbiology*, vol. 84, pp. e02508–17, feb 2018. 19

[179] C. Huang, Q. Wang, H. Yao, G. Wang, and Y. Li, "Determination of red cell by Raman spectroscopy based on microfluidic chip and optical tweezers," *Chinese Journal of Analytical Chemistry*, vol. 35, no. 10, pp. 1410–1414, 2007. 19

[180] A. Y. Lau, L. P. Lee, and J. W. Chan, "An integrated optofluidic platform for Raman-activated cell sorting," *Lab on a Chip*, vol. 8, no. 7, p. 1116, 2008. 20

[181] C. Bolwien, G. Sulz, S. Becker, H. Thielecke, H. Mertsching, and S. Koch, "Rapid detection of bacterial contamination in cell or tissue cultures based on Raman spectroscopy," in *Biomedical Optical Spectroscopy* (A. Mahadevan-Jansen, W. Petrich, R. R. Alfano, and A. Katz, eds.), no. 6853, p. 68530F, feb 2008. 20

[182] J. W. Chan, "Recent advances in laser tweezers Raman spectroscopy (LTRS) for label-free analysis of single cells," *Journal of Biophotonics*, vol. 6, pp. 36–48, jan 2013. 20

[183] S. Dochow, C. Krafft, U. Neugebauer, T. Bocklitz, T. Henkel, G. Mayer, J. Albert, and J. Popp, "Tumour cell identification by means of Raman spectroscopy in combination with optical traps and microfluidic environments," *Lab on a Chip*, vol. 11, no. 8, p. 1484, 2011. 20

[184] U. Neugebauer, S. Dochow, C. Krafft, T. Bocklitz, J. H. Clement, and J. Popp, "Diagnostics of tumor cells by combination of Raman spectroscopy and microfluidics," in *Clinical and Biomedical Spectroscopy and Imaging II* (N. Ramanujam and J. Popp, eds.), vol. 8087, (Washington, D.C.), p. 80870J, OSA, jun 2011. 20

[185] S. Dochow, C. Beleites, T. Henkel, G. Mayer, J. Albert, J. Clement, C. Krafft, and J. Popp, "Quartz microfluidic chip for tumour cell identification by Raman spectroscopy in combination with optical traps," *Analytical and Bioanalytical Chemistry*, vol. 405, pp. 2743–2746, mar 2013. 20

[186] S. Dochow, M. Becker, R. Spittel, C. Beleites, S. Stanca, I. Latka, K. Schuster, J. Kobelke, S. Unger, T. Henkel, G. Mayer, J. Albert, M. Rothhardt, C. Krafft, and J. Popp, "Raman-on-chip device and detection fibres with fibre Bragg grating for analysis of solutions and particles," *Lab on a Chip*, vol. 13, no. 6, p. 1109, 2013. 20

[187] C. Krafft, C. Beleites, I. W. Schie, J. H. Clement, and J. Popp, "Raman-based identification of circulating tumor cells for cancer diagnosis," in *Biomedical Vibrational Spectroscopy VI: Advances in Research and Industry* (A. Mahadevan-Jansen and W. Petrich, eds.), (Rotterdam), p. 970408, SensePublishers, mar 2016. 20

[188] F. Knoepp, J. Wahl, A. Andersson, J. Borg, N. Weissmann, and K. Ramser, "Development of a Gas-Tight Microfluidic System for Raman Sensing of Single Pulmonary Arterial Smooth Muscle Cells Under Normoxic/Hypoxic Conditions," *Sensors*, vol. 18, p. 3238, sep 2018. 20

[189] P. Zhang, L. Ren, X. Zhang, Y. Shan, Y. Wang, Y. Ji, H. Yin, W. E. Huang, J. Xu, and B. Ma, "Raman-activated cell sorting based on dielectrophoretic single-cell trap and release," *Analytical Chemistry*, vol. 87, no. 4, pp. 2282–2289, 2015. 21

[190] P. Zhang, Y. Su, Y. Ji, L. Ren, Y. Gong, and B. Ma, "Screening of High-Yield-Astaxanthing Haematococcus pluvialis strain by Raman Activated Cell Sorter," *19 510 th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, pp. 510–512, 2015. 21

[191] X. Wang, L. Ren, Y. Su, Y. Ji, Y. Liu, C. Li, X. Li, Y. Zhang, W. Wang, Q. Hu, D. Han, J. Xu, and B. Ma, "Raman-Activated Droplet Sorting (RADS) for Label-Free High-Throughput Screening of Microalgal Single-Cells," *Analytical Chemistry*, vol. 89, no. 22, pp. 12569–12577, 2017. 21

[192] P. Zhang, Y. Xiao, Z. Li, J. Guo, and L. Lu, "Microalgae in Microwell Arrays Exhibit Differences with Those in Flasks: Evidence from Growth Rate, Cellular Carotenoid, and Oxygen Production," *Frontiers in Plant Science*, vol. 8, pp. 1–10, jan 2018. 21

[193] E. Eriksson, J. Scrimgeour, A. Granéli, K. Ramser, R. Wellander, J. Enger, D. Hanstorp, and M. Goksör, "Optical manipulation and microfluidics for studies of single cell dynamics," *Journal of Optics A: Pure and Applied Optics*, vol. 9, pp. S113–S121, aug 2007. 21

[194] K. Ramser, J. Enger, M. Goksör, D. Hanstorp, K. Logg, and M. Käll, "A microfluidic system enabling Raman measurements of the oxygenation cycle in single optically trapped red blood cells," *Lab on a Chip*, vol. 5, no. 4, pp. 431–436, 2005.

[195] K. Ramser, J. Enger, M. Goksör, D. Hanstorp, K. Logg, and M. Käll, "A microfluidic system enabling Raman measurements of the oxygenation cycle in single optically trapped red blood cells," *Lab Chip*, vol. 5, pp. 431–436, oct 2005. 21

[196] M. Göllner, A. C. Toma, N. Strelnikova, S. Deshpande, and T. Pfohl, "A self-filling microfluidic device for noninvasive and time-resolved single red blood cell experiments," *Biomicrofluidics*, vol. 10, p. 054121, sep 2016. 21

[197] K. Ramser, W. Wenseleers, S. Dewilde, S. Van Doorslaer, and L. Moens, "The combination of resonance Raman spectroscopy, optical tweezers and microfluidic systems applied to the study of various heme-containing single cells," *Spectroscopy*, vol. 22, no. 4, pp. 287–295, 2008. 21

[198] R. S. Verma, S. Ahlawat, and A. Uppal, "Optical guiding-based cell focusing for Raman flow cell cytometer," *The Analyst*, vol. 143, no. 11, pp. 2648–2655, 2018. 21

[199] M. Li, P. C. Ashok, K. Dholakia, and W. E. Huang, "Raman-activated cell counting for profiling carbon dioxide fixing microorganisms," *Journal of Physical Chemistry A*, vol. 116, no. 25, pp. 6560–6563, 2012. 21

[200] D. McIlvenna, W. E. Huang, P. Davison, A. Glidle, J. Cooper, and H. Yin, "Continuous cell sorting in a flow based on single cell resonance Raman spectra," *Lab on a Chip*, vol. 16, no. 8, pp. 1420–1429, 2016. 21

[201] H. S. Kim, S. C. Waqued, D. T. Nodurft, T. P. Devarenne, V. V. Yakovlev, and A. Han, "Raman spectroscopy compatible PDMS droplet microfluidic culture and analysis platform towards on-chip lipidomics," *The Analyst*, vol. 142, no. 7, pp. 1054–1060, 2017. 21

[202] J. Feng, C. De La Fuente-Núñez, M. J. Trimble, J. Xu, R. E. W. Hancock, and X. Lu, "An in situ Raman spectroscopy-based microfluidic "lab-on-a-chip" platform for non-destructive and continuous characterization of Pseudomonas aeruginosa biofilms," *Chem. Commun.*, vol. 51, no. 43, pp. 8966–8969, 2015. 21

[203] B. M. Liszka, H. S. Rho, Y. Yang, A. T. M. Lenferink, L. W. M. M. Terstappen, and C. Otto, "A microfluidic chip for high resolution Raman imaging of biological cells," *RSC Adv.*, vol. 5, no. 61, pp. 49350–49355, 2015. 21

[204] C. Cao, D. Zhou, T. Chen, A. M. Streets, and Y. Huang, "Label-Free Digital Quantification of Lipid Droplets in Single Cells by Stimulated Raman Microscopy on a Microfluidic Platform," *Analytical Chemistry*, vol. 88, pp. 4931–4939, may 2016. 22

[205] T. Ideguchi, T. Nakamura, S. Takizawa, M. Tamamitsu, S. Lee, K. Hiramatsu, V. Ramaiah-Badarla, J.-w. Park, Y. Kasai, T. Hayakawa, S. Sakuma, F. Arai, and K. Goda, "Microfluidic single-particle chemical analyzer with dual-comb coherent Raman spectroscopy," *Optics Letters*, vol. 43, p. 4057, aug 2018. 22, 197

[206] K. Hiramatsu, T. Ideguchi, Y. Yonamine, S. Lee, Y. Luo, K. Hashimoto, T. Ito, M. Hase, J.-W. Park, Y. Kasai, S. Sakuma, T. Hayakawa, F. Arai, Y. Hoshino, and K. Goda, "High-throughput label-free molecular fingerprinting flow cytometry," *Science Advances*, vol. 5, p. eaau0241, jan 2019. 22, 197

[207] G. R. Sukalal, *Raman spectral analysis technique for the chemical composition of aerosols in the environment of sugar factory and agricultural university campus at Rahuri in Ahmednagar district*. PhD thesis, Savitribai Phule Pune University, 2012. 24, 26

[208] I. R. Lewis and H. G. M. Edwards, *Handbook of Raman spectroscopy : from the research laboratory to the process line*. CRC Press, 2001. 25, 28

[209] K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds*. Hoboken, NJ, USA: John Wiley & Sons, Inc., dec 2008. 27

[210] S. A. Borman, "Nonlinear Raman Spectroscopy," *Analytical Chemistry*, vol. 54, pp. 1021A–1026A, aug 1982. 27, 28

[211] J. Chalmers and P. Griffiths, "Handbook of Vibrational Spectroscopy. Theory and Instrumentation.," *Journal of the American Chemical Society*, vol. 124, pp. 9958–9958, aug 2002. 29, 30, 32

[212] W. Demtröder, *Laser Spectroscopy*. Advanced Texts in Physics, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. 30, 31

[213] V. Jain, M. C. Biesinger, and M. R. Linford, "The Gaussian-Lorentzian Sum, Product, and Convolution (Voigt) functions in the context of peak fitting X-ray photoelectron spectroscopy (XPS) narrow scans," *Applied Surface Science*, vol. 447, pp. 548–553, jul 2018. 31, 73

[214] R. Hesse, P. Streubel, and R. Szargan, "Product or sum: comparative tests of Voigt, and product or sum of Gaussian and Lorentzian functions in the fitting of synthetic Voigt-based X-ray photoelectron spectra," *Surface and Interface Analysis*, vol. 39, pp. 381–391, may 2007. 31

[215] G. Bonaccorso, *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Birmingham: Packt Publishing Ltd, 2 ed., 2018. 33, 38

[216] W. Schumacher, S. Stöckel, P. Rösch, and J. Popp, "Self-defining tree-like classifiers for interpretation of Raman spectroscopic experiments," *Journal of Chemometrics*, vol. 30, no. 5, pp. 268–283, 2016. 33, 104

[217] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philosophical Magazine Series 6*, vol. 2, pp. 559–572, nov 1901. 34

[218] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, no. 7, pp. 498–520, 1933. 34

[219] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, 1987. 34

[220] I. T. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, New York: Springer-Verlag, 2002. 34, 35, 36

[221] X. Li, T. Yang, and S. Li, "Discrimination of serum Raman spectroscopy between normal and colorectal cancer using selected parameters and regression-discriminant analysis," *Applied Optics*, vol. 51, p. 5038, jul 2012. 37

[222] H.-Y. Kim, K.-M. Lee, S.-H. Kim, Y.-J. Kwon, Y.-J. Chun, and H.-K. Choi, "Comparative metabolic and lipidomic profiling of human breast cancer cells with different metastatic potentials.," *Oncotarget*, vol. 7, no. 41, pp. 67111–67128, 2016. 37

[223] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY: Springer New York, 2009. 38, 41, 42

[224] T. J. Harvey, C. Hughes, A. D. Ward, E. C. Faria, A. Henderson, N. W. Clarke, M. D. Brown, R. D. Snook, and P. Gardner, "Classification of fixed urological cells using Raman tweezers," *Journal of Biophotonics*, vol. 2, no. 1-2, pp. 47–69, 2009. 40, 42

[225] R. Cangelosi and A. Goriely, "Component retention in principal component analysis with application to cDNA microarray data," *Biology Direct*, vol. 2, pp. 1–21, 2007. 41

[226] Y. Chen, Y. Su, L. Ou, C. Zou, and Z. Chen, "Classification of nasopharyngeal cell lines (C666-1, CNE2, NP69) via Raman spectroscopy and decision tree," *Vibrational Spectroscopy*, vol. 80, pp. 24–29, 2015. 41

[227] X. Li, T. Yang, S. Li, D. Wang, Y. Song, and K. Yu, "Different classification algorithms and serum surface enhanced Raman spectroscopy for noninvasive discrimination of gastric diseases," *Journal of Raman Spectroscopy*, vol. 47, no. 8, pp. 917–925, 2016. 41

[228] R. Quinlan, "Data Mining Tools See5 and C5.0." 42, 100, 103

[229] A. Kyriakides, E. Kastanos, K. Hadjigeorgiou, and C. Pitris, "Classification of Raman spectra using the correlation kernel," *Journal of Raman Spectroscopy*, vol. 42, no. 5, pp. 904–909, 2011. 42

[230] E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines.," *International journal of oncology*, vol. 32, pp. 653–62, mar 2008. 42

[231] Ł. Ożog and D. Aebisher, "Singlet oxygen lifetime and diffusion measurements," *European Journal of Clinical and Experimental Medicine*, vol. 16, no. 2, pp. 123–126, 2018. 46

[232] P. Tabeling, *Introduction to Microfluidics*. Oxford University Press., 2005. 46

[233] D. J. Beebe, G. a. Mensing, and G. M. Walker, "Physics and applications of microfluidics in biology.," *Annual review of biomedical engineering*, vol. 4, pp. 261–286, 2002. 46, 47

[234] M. J. Fuerstman, A. Lai, M. E. Thurlow, S. S. Shevkoplyas, H. a. Stone, and G. M. Whitesides, "The pressure drop along rectangular microchannels containing bubbles," *Lab on a Chip*, vol. 7, no. 11, p. 1479, 2007. 47

[235] X. Xu, P. Sarder, Z. Li, and A. Nehorai, "Optimization of microfluidic microsphere-trap arrays," *Biomicrofluidics*, vol. 7, no. 1, p. 014112, 2013. 47

[236] H. Bach and N. Neuroth, eds., *The Properties of Optical Glass*. Schott Series on Glass and Glass Ceramics, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. 54

[237] J. Icha, M. Weber, J. C. Waters, and C. Norden, "Phototoxicity in live fluorescence microscopy, and how to avoid it," *BioEssays*, vol. 39, p. 1700003, aug 2017. 56

[238] M. Koch, C. Suhr, B. Roth, and M. Meinhardt-Wollweber, "Iterative morphological and mollifier-based baseline correction for Raman spectra," *Journal of Raman Spectroscopy*, vol. 48, pp. 336–342, feb 2017. 66, 69, 70, 71

[239] B. Elfsson, I. Wallin, S. Eksborg, K. Rudaeus, a. M. Ros, and H. Ehrsson, "Stability of 5-aminolevulinic acid in aqueous solution.," *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences*, vol. 7, pp. 87–91, jan 1999. 76

[240] D. D. Carlo, L. Y. Wu, and L. P. Lee, "Dynamic single cell culture array," *Lab on a Chip*, vol. 6, no. 11, p. 1445, 2006. 81, 82

[241] W.-H. Tan and S. Takeuchi, "A trap-and-release integrated microfluidic system for dynamic microarray applications.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 4, pp. 1146–51, 2007. 83, 84

[242] A. H. Churchman, V. Mico, J. G. de Pablo, S. A. Peyman, S. Freear, S. D. Evans, J. Gala de Pablo, S. A. Peyman, S. Freear, S. D. Evans, J. G. de Pablo, S. A. Peyman, S. Freear, and S. D. Evans, "Combined flow-focus and self-assembly routes for the formation of lipid stabilized oil-shelled microbubbles," *Microsystems & Nanoengineering*, vol. 4, p. 17087, feb 2018. 88

[243] A. S. Klymchenko, "Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications," *Accounts of Chemical Research*, vol. 50, no. 2, pp. 366–375, 2017. 91, 127

[244] N. Ridgway and R. McLeod, *Biochemistry of Lipids, Lipoproteins and Membranes*. Elsevier Science, 6 ed., 2015. 93

[245] I. R. Hill and I. W. Levin, "Vibrational spectra and carbon-hydrogen stretching mode assignments for a series of n-alkyl carboxylic acids," *The Journal of Chemical Physics*, vol. 70, no. 2, p. 842, 1979. 94, 95

[246] R. G. Nuzzo, E. M. Korenic, and L. H. Dubois, "Studies of the temperature-dependent phase behavior of long chain n -alkyl thiol monolayers on gold," *The Journal of Chemical Physics*, vol. 93, pp. 767–773, jul 1990. 94, 95

[247] B. Duranton, V. Holl, Y. Schneider, S. Carnesecchi, F. Gossé, F. Raul, and N. Seiler, "Polyamine metabolism in primary human colon adenocarcinoma cells (SW480) and their lymph node metastatic derivatives (SW620).," *Amino acids*, vol. 24, no. 1-2, pp. 63–72, 2003. 96

[248] D. Damania, H. Subramanian, V. Backman, E. C. Anderson, M. H. Wong, O. J. T. McCarty, and K. G. Phillips, "Network signatures of nuclear and cytoplasmic density alterations in a model of pre and postmetastatic colorectal cancer," *Journal of Biomedical Optics*, vol. 19, p. 016016, jan 2014. 96

[249] S. Maddula and J. I. Baumbach, "Heterogeneity in tumor cell energetic metabolome at different cell cycle phases of human colon cancer cell lines," *Metabolomics*, vol. 7, no. 4, pp. 509–523, 2011. 96, 105

[250] C. Laurich, M. A. Wheeler, J. Iida, C. L. Neudauer, J. B. McCarthy, and K. M. Bullard, "Hyaluronan mediates adhesion of metastatic colon carcinoma cells1," *Journal of Surgical Research*, vol. 122, pp. 70–74, nov 2004. 96

[251] C. Johannessen, P. C. White, and S. Abdali, "Resonance Raman optical activity and surface enhanced resonance Raman optical activity analysis of cytochrome c," *Journal of Physical Chemistry A*, vol. 111, no. 32, pp. 7771–7776, 2007. 98

[252] T. Knutsen, H. M. Padilla-Nash, D. Wangsa, L. Barenboim-Stapleton, J. Camps, N. McNeil, M. J. Difilippantonio, and T. Ried, "Definitive Molecular Cytogenetic Characterization of 15 Colorectal Cancer Cell Lines," *Genes Chromosomes Cancer*, vol. 49, no. 3, pp. 204–223, 2010. 105

[253] ECACC, "European Collection of Authenticated Cell Cultures (ECACC)." 105

[254] ATCC, "ATCC: The Global Bioresource Center." 105

[255] T. Seierstad, K. Røe, B. Sitter, J. Halgunset, K. Flatmark, A. H. Ree, D. R. Olsen, I. S. Gribbestad, and T. F. Bathen, "Principal component analysis for the comparison of metabolic profiles from human rectal cancer biopsies and colorectal xenografts using high-resolution magic angle spinning 1H magnetic resonance spectroscopy.," *Molecular cancer*, vol. 7, p. 33, 2008. 105, 106

[256] "Colorectal Cancer Atlas," 2017. 105

[257] P. Foa, a. T. Maiolo, L. Lombardi, H. Toivonen, T. Rytömaa, and E. E. Polli, "Growth pattern of the human promyelocytic leukaemia cell line HL60.," *Cell and tissue kinetics*, vol. 15, no. 4, pp. 399–404, 1982.

[258] R. A. Fleck, S. Romero-Steiner, and M. H. Nahm, "Use of HL-60 cell line to measure opsonic capacity of pneumococcal antibodies.," *Clinical and diagnostic laboratory immunology*, vol. 12, no. 1, pp. 19–27, 2005. 105

[259] H. Brunner, F. Hausmann, R. Krieg, E. Endlicher, J. Scholmerich, R. Knuechel, and H. Messmann, "The effects of 5-aminolevulinic acid esters on protoporphyrin IX production in human adenocarcinoma cell lines Brunner H.," *Photochemistry and Photobiology*, vol. 74, no. 5, pp. 721–725, 2001. 113, 127, 193

[260] J. Y. Chen, Q. Peng, and H. J. Jodl, "Infrared spectral comparison of 5-aminolevulinic acid and its hexyl ester," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. 59, no. 11, pp. 2571–2576, 2003. 114

[261] M. Abe, T. Kitagawa, and Y. Kyogoku, "Resonance Raman spectra of octaethylporphyrinato-Ni(II) and meso-deuterated and 15N substituted derivatives. II. A normal coordinate analysis," *The Journal of Chemical Physics*, vol. 69, no. 10, pp. 4526–4534, 1978. 113, 114, 116

[262] S. Choi, T. G. Spiro, K. C. Langry, and K. M. Smith, "Vinyl influences on protoheme resonance Raman spectra: nickel(II) protoporphyrin IX with deuterated vinyl groups," *Journal of the American Chemical Society*, vol. 104, pp. 4337–4344, aug 1982. 113, 116

[263] H. Brunner and H. Sussner, "Resonance Raman scattering on haemoglobin," *Biochimica et Biophysica Acta (BBA) - Protein Structure*, vol. 310, no. 1, pp. 20–31, 1973.

[264] M. Sarkar and A. L. Verma, "Effect of vinyl groups on resonance Raman spectra of protoporphyrin-IX," *Journal of Raman Spectroscopy*, vol. 17, pp. 407–414, oct 1986. 113, 114

[265] S. Hu, I. K. Morris, J. P. Singh, K. M. Smith, and T. G. Spiro, "Complete assignment of cytochrome c resonance Raman spectra via enzymic reconstitution with isotopically labeled hemes," *Journal of the American Chemical Society*, vol. 115, pp. 12446–12458, dec 1993. 113, 115, 116

[266] S. Hu, K. M. Smith, and T. G. Spiro, "Assignment of Protoheme Resonance Raman Spectrum by Heme Labeling in Myoglobin," *Journal of the American Chemical Society*, vol. 118, pp. 12638–12646, jan 1996. 116

[267] A. Juzeniene, K. P. Nielsen, and J. Moan, "Biophysical Aspects of Photodynamic Therapy," *Journal of Environmental Pathology, Toxicology and Oncology*, vol. 25, no. 1-2, pp. 7–28, 2006. 123

[268] S. Shashkova and M. C. Leake, "Single-molecule fluorescence microscopy review: shedding new light on old problems," *Bioscience Reports*, vol. 37, p. BSR20170031, aug 2017. 127

[269] Z. Huang, H. Lui, D. I. McLean, M. Korbelik, and H. Zeng, "Raman Spectroscopy in Combination with Background Near-infrared Autofluorescence Enhances the In Vivo Assessment of Malignant Tissues," *Photochemistry and Photobiology*, vol. 81, no. 5, p. 1219, 2005. 127

[270] N. Uzunbajakava, A. Lenferink, Y. Kraan, E. Volokhina, G. Vrensen, J. Greve, and C. Otto, "Nonresonant Confocal Raman Imaging of DNA and Protein Distribution in Apoptotic Cells," *Biophysical Journal*, vol. 84, no. 6, pp. 3968–3981, 2003. 127, 128

[271] R. P. Carney, S. Hazari, M. Colquhoun, D. Tran, B. Hwang, M. S. Mulligan, J. D. Bryers, E. Girda, G. S. Leiserowitz, Z. J. Smith, and K. S. Lam, "Multispectral Optical Tweezers for Biochemical Fingerprinting of CD9-Positive Exosome Subpopulations," *Analytical Chemistry*, vol. 89, no. 10, pp. 5357–5363, 2017. 128

[272] B. Kang, M. M. Afifi, L. A. Austin, and M. A. El-Sayed, "Exploiting the nanoparticle plasmon effect: Observing drug delivery dynamics in single cells via Raman/fluorescence imaging spectroscopy," *ACS Nano*, vol. 7, no. 8, pp. 7420–7427, 2013. 128

[273] E.-O. Ganbold, J. Yoon, D. Kim, and S.-W. Joo, "Nonidentical intracellular drug release rates in Raman and fluorescence spectroscopic determination," *Physical Chemistry Chemical Physics*, vol. 17, no. 5, pp. 3019–3023, 2015. 128

[274] Y. Zeng, S. Zhao, H. Wei, M. P. Tucker, M. E. Himmel, N. S. Mosier, R. Meilan, and S.-Y. Ding, "In situ micro-spectroscopic investigation of lignin in poplar cell walls pretreated by maleic acid," *Biotechnology for Biofuels*, vol. 8, p. 126, dec 2015. 128

[275] H.-J. van Manen and C. Otto, "Hybrid Confocal Raman Fluorescence Microscopy on Single Cells Using Semiconductor Quantum Dots," *Nano Letters*, vol. 7, pp. 1631–1636, jun 2007. 128

[276] X. Li, M. Jiang, J. W. Y. Lam, B. Z. Tang, and J. Y. Qu, "Mitochondrial Imaging with Combined Fluorescence and Stimulated Raman Scattering Microscopy Using a Probe of the Aggregation-Induced Emission Characteristic," *Journal of the American Chemical Society*, vol. 139, pp. 17022–17030, nov 2017. 128

[277] Z. Chen, D. W. Paley, L. Wei, A. L. Weisman, R. A. Friesner, C. Nuckolls, and W. Min, "Multicolor live-cell chemical imaging by isotopically edited alkyne vibrational palette," *Journal of the American Chemical Society*, vol. 136, no. 22, pp. 8027–8033, 2014. 128, 129

[278] D. E. Dolmans, D. Fukumura, and R. K. Jain, "Photodynamic therapy for cancer," *Nature Reviews Cancer*, vol. 3, pp. 380–387, may 2003. 128

[279] W. Fan, P. Huang, and X. Chen, "Overcoming the Achilles' heel of photodynamic therapy.," *Chem. Soc. Rev.*, vol. 45, no. 23, pp. 6488–6519, 2016.

[280] N. Shirasu, S. O. Nam, and M. Kuroki, "Tumor-targeted Photodynamic Therapy," *Anticancer Res.*, vol. 33, pp. 2823–2831, 2013. 128

[281] J. H. Barnard, J. C. Collings, A. Whiting, S. A. Przyborski, and T. B. Marder, "Synthetic retinoids: Structure-activity relationships," *Chemistry - A European Journal*, vol. 15, no. 43, pp. 11430–11442, 2009. 129, 132

[282] V. B. Christie, J. H. Barnard, A. S. Batsanov, C. E. Bridgens, E. B. Cartmell, J. C. Collings, D. J. Maltman, C. P. F. Redfern, T. B. Marder, S. Przyborski, and A. Whiting, "Synthesis and evaluation of synthetic retinoid derivatives as inducers of stem cell differentiation," *Organic & Biomolecular Chemistry*, vol. 6, no. 19, p. 3497, 2008. 129

[283] G. Clemens, K. R. Flower, P. Gardner, A. P. Henderson, J. P. Knowles, T. B. Marder, A. Whiting, and S. Przyborski, "Design and biological evaluation of synthetic retinoids: probing length vs. stability vs. activity," *Molecular BioSystems*, vol. 9, no. 12, p. 3124, 2013. 129, 132

[284] W. N. Martens, R. L. Frost, J. Kristof, and J. Theo Kloprogge, "Raman spectroscopy of dimethyl sulphoxide and deuterated dimethyl sulphoxide at 298 and 77 K," *Journal of Raman Spectroscopy*, vol. 33, pp. 84–91, feb 2002. 129, 130

[285] M. Krämer, U. H. F. Bunz, and A. Dreuw, "Comprehensive Look at the Photochemistry of Tolane," *The Journal of Physical Chemistry A*, vol. 121, pp. 946–953, feb 2017. 132

[286] C. Ferrante, U. Kensy, and B. Dick, "Does diphenylacetylene (tolan) fluoresce from its second excited singlet state? Semiempirical MO calculations and fluorescence quantum yield measurements," *The Journal of Physical Chemistry*, vol. 97, pp. 13457–13463, dec 1993.

[287] Y. Hirata, T. Okada, and T. Nomoto, "Photoinduced intramolecular charge separation of p-N , N-dimethylamino-p ' -cyano-diphenylacetylene in polar solvents," *Chem. Phys. Lett.*, vol. 278, pp. 133–138, 1997.

[288] M. Szyszkowska, I. Bylińska, and W. Wiczk, "Influence of an electron-acceptor substituent type on the photophysical properties of unsymmetrically substituted diphenylacetylene," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 326, pp. 76–88, jul 2016.

[289] M. Biswas, P. Nguyen, T. B. Marder, and L. R. Khundkar, "Unusual size dependence of nonradiative charge recombination rates in acetylene-bridged compounds," *Journal of Physical Chemistry A*, vol. 101, no. 9, pp. 1689–1695, 1997.

[290] J. C. Collings, A. C. Parsons, L. Porrès, A. Beeby, A. S. Batsanov, J. A. Howard, D. P. Lydon, P. J. Low, I. J. Fairlamb, and T. B. Marder, "Optical properties of donor-acceptor phenyleneethynylene systems containing the 6-methylpyran-2-one group as an acceptor," *Chemical Communications*, no. 21, pp. 2666–2668, 2005.

[291] P. Nguyen, Z. Yuan, L. Agocs, G. Lesley, and T. B. Marder, "Synthesis of Symmetrical and Unsymmetric 1,4-Bis(P-R-Phenylethynyl)Benzenes Via Palladium Copper-Catalyzed Cross-Coupling and Comments on the Coupling of Aryl Halides with Terminal Alkynes," *Inorganica Chimica Acta*, vol. 220, no. 1-2, pp. 289–296, 1994. 132

[292] C. Dehu, F. Meyers, and J. L. Bredas, "Donor-acceptor diphenylacetylenes: geometric structure, electronic structure, and second-order nonlinear optical properties," *Journal of the American Chemical Society*, vol. 115, pp. 6198–6206, jul 1993. 132

[293] P. Nguyen, G. Lesley, T. B. Marder, I. Ledoux, and J. Zyss, "Second-Order Nonlinear Optical Properties of Push - Pull Bis(phenylethynyl)benzenes and Unsymmetric Platinum Bis(phenylacetylide) Complexes," *Chem. Mater.*, vol. 9, no. 13, pp. 406–408, 1997. 132

[294] S. A. Schønberg, A. G. Lundemo, T. Fladvad, K. Holmgren, H. Bremseth, A. Nilsen, O. Gederaas, K. E. Tvedt, K. W. Egeberg, and H. E. Krokan, "Closely related colon cancer cell lines display different sensitivity to polyunsaturated fatty acids, accumulate different lipid classes and downregulate sterol regulatory element-binding protein 1," *FEBS Journal*, vol. 273, pp. 2749–2765, jun 2006. 134

[295] L. Ashton and E. W. Blanch, "PH-induced conformational transitions in α-lactalbumin investigated with two-dimensional Raman correlation variance plots and moving windows," *Journal of Molecular Structure*, vol. 974, no. 1-3, pp. 132–138, 2010. 136

[296] A. Barth and C. Zscherp, "What vibrations tell about proteins," *Quarterly Reviews of Biophysics*, vol. 35, no. 4, p. S0033583502003815, 2002. 136

[297] Z. Movasaghi, S. Rehman, and D. I. ur Rehman, "Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues," *Applied Spectroscopy Reviews*, vol. 43, no. 2, pp. 134–179, 2008. 136

[298] D. F. Wallach, S. P. Verma, and J. Fookson, "Application of laser Raman and infrared spectroscopy to the analysis of membrane structure," *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, vol. 559, no. 2-3, pp. 153–208, 1979. 136

[299] Dow Corning, "Dow Corning Corporation Sylgard 184 Silicone Elastomer Kit," tech. rep., Dow Corning Corporation, 2010. 162

[300] R. N. Palchesko, L. Zhang, Y. Sun, and A. W. Feinberg, "Development of Polydimethylsiloxane Substrates with Tunable Elastic Modulus to Study Cell Mechanobiology in Muscle and Nerve," *PLoS ONE*, vol. 7, p. e51499, dec 2012. 162

[301] L. Jayes, A. P. Hard, C. Séné, S. F. Parker, and U. A. Jayasooriya, "Vibrational Spectroscopic Analysis of Silicones: A Fourier Transform-Raman and Inelastic Neutron Scattering Investigation," *Analytical Chemistry*, vol. 75, pp. 742–746, feb 2003. 163

[302] D. Cai, A. Neyer, R. Kuckuk, and H. M. Heise, "Raman, mid-infrared, near-infrared and ultraviolet-visible spectroscopy of PDMS silicone rubber for characterization of polymer optical waveguide materials," *Journal of Molecular Structure*, vol. 976, pp. 274–281, jul 2010. 162, 163

[303] S. C. Bae, H. Lee, Z. Lin, and S. Granick, "Chemical imaging in a surface forces apparatus: Confocal Raman spectroscopy of confined poly(dimethylsiloxane)," *Langmuir*, vol. 21, no. 13, pp. 5685–5688, 2005. 163

[304] Y. Jin, G. A. Guirgis, and J. R. Durig, "Conformational Studies of Methyl Vinyl Silane from Temperature-Dependent FT-IR Spectra of Xenon and Krypton Solutions," *Structural Chemistry*, vol. 11, no. 4, pp. 229–240, 2000. 162

[305] A. F. Goncharov and V. V. Struzhkin, "Raman spectroscopy of metals, high-temperature superconductors and related materials under high pressure," *Journal of Raman Spectroscopy*, vol. 34, no. 7-8, pp. 532–548, 2003. 164

[306] J. Y. Baek, G. H. Kwon, J. Y. Kim, J. H. Cho, S. H. Lee, K. Sun, and S. H. Lee, "Stable Deposition and Patterning of Metal Layers on the PDMS Substrate and Characterization for the Development of the Flexible and Implantable Micro Electrode," *Solid State Phenomena*, vol. 124-126, pp. 165–168, 2007. 166

[307] T. W. Lee, S. Jeon, J. Maria, J. Zaumseil, J. W. P. Hsu, and J. A. Rogers, "Soft-contact optical lithography using transparent elastomeric stamps: Application to nanopatterned organic light-emitting devices," *Advanced Functional Materials*, vol. 15, no. 9, pp. 1435–1439, 2005. 166

[308] M. K. Hedayati, F. Faupel, and M. Elbahri, "Review of plasmonic nanocomposite metamaterial absorber," *Materials*, vol. 7, no. 2, pp. 1221–1248, 2014. 167

[309] M. K. Hedayati, M. Javaherirahim, B. Mozooni, R. Abdelaziz, A. Tavassolizadeh, V. S. K. Chakravadhanula, V. Zaporojtchenko, T. Strunkus, F. Faupel, and M. Elbahri, "Design of a Perfect Black Absorber at Visible Frequencies Using Plasmonic Metamaterials," *Advanced Materials*, vol. 23, pp. 5410–5414, dec 2011. 167

[310] W.-W. Chen, C.-H. Chien, C.-L. Wang, H.-H. Wang, Y.-L. Wang, S.-T. Ding, T.-S. Lee, and T.-C. Chang, "Automated quantitative analysis of lipid accumulation and hydrolysis in living macrophages with label-free imaging," *Analytical and Bioanalytical Chemistry*, vol. 405, pp. 8549–8559, oct 2013. 177

[311] H. Abramczyk, J. Surmacki, M. Kopeć, A. K. Olejnik, K. Lubecka-Pietruszewska, and K. Fabianowska-Majewska, "The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue," *The Analyst*, vol. 140, no. 7, pp. 2224–2235, 2015. 177

[312] O. Ibrahim, A. Maguire, A. D. Meade, S. Flint, M. Toner, H. J. Byrne, and F. M. Lyng, "Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections," *Analytical Methods*, vol. 9, no. 32, pp. 4709–4717, 2017. 177, 197

[313] S. Burikov, T. Dolenko, S. Patsaeva, Y. Starokurov, and V. Yuzhakov, "Raman and IR spectroscopy research on hydrogen bonding in water-ethanol systems," *Molecular Physics*, vol. 108, pp. 2427–2436, sep 2010. 182, 184, 186

[314] H. Mishra, A. M. Schrader, D. W. Lee, A. Gallo, S.-Y. Chen, Y. Kaufman, S. Das, and J. N. Israelachvili, "Time-Dependent Wetting Behavior of PDMS Surfaces with Bioinspired, Hierarchical Structures," *ACS Applied Materials & Interfaces*, vol. 8, pp. 8168–8174, mar 2016. 182

[315] A. H. Churchman, *Routes for the Formation of Thin Oil Shelled Microbubbles towards Hydrophobic Drug Delivery*. PhD thesis, University of Leeds, 2017. 186

[316] R. A. Bamford, A. Smith, J. Metz, G. Glover, R. W. Titball, and S. Pagliara, "Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy," *BMC Biology*, vol. 15, p. 121, dec 2017. 187

[317] K. Hamasha, Q. I. Mohaidat, R. A. Putnam, R. C. Woodman, S. Palchaudhuri, and S. J. Rehse, "Sensitive and specific discrimination of pathogenic and nonpathogenic Escherichia coli using Raman spectroscopy-a comparison of two multivariate analysis techniques," *Biomedical Optics Express*, vol. 4, no. 4, p. 481, 2013. 187

[318] L. Tirinato, C. Liberale, S. Di Franco, P. Candeloro, A. Benfante, R. La Rocca, L. Potze, R. Marotta, R. Ruffilli, V. P. Rajamanickam, M. Malerba, F. De Angelis, A. Falqui, E. Carbone, M. Todaro, J. P. Medema, G. Stassi, and E. Di Fabrizio, "Lipid Droplets: A New Player in Colorectal Cancer Stem Cells Unveiled by Spectroscopic Imaging," *STEM CELLS*, vol. 33, pp. 35–44, jan 2015. 193